

Year 1 phonics screening check

Pilot 2011: Technical report

**Standards and
Testing Agency**

An executive agency of the
Department for Education

About this document

What is this document about?

This document provides a technical evaluation of the Year 1 phonics screening check, including information relating to Ofqual's common assessment criteria of validity, reliability, minimising bias, comparability and manageability as set out in its *Regulatory Framework for National Assessment arrangements* (Ofqual, 2011).

Who is this document for?

- This document is primarily aimed at a technical audience, but contains information that will be of interest to all stakeholders involved in the Year 1 phonics screening check, including schools.
- The document may be of particular interest to schools involved in the pilot in June 2011 and other education professionals.

Contents

1.	Introduction and policy background.....	5
1.1	The purpose of the Year 1 phonics screening check.....	5
1.2	Executive summary.....	6
2.	The Assessment framework.....	11
2.1	Development of the pilot framework.....	11
2.2	Independent review of the assessment framework.....	18
2.3	The Assessment framework for roll-out.....	21
3.	The pilot.....	23
3.1	Instrument development.....	23
3.2	School selection.....	29
3.3	The Administration guide.....	30
3.4	Training.....	31
3.5	Evaluation.....	33
3.6	Pilot logistics.....	35
4.	Standard setting.....	38
4.1	Standard setting participants.....	38
4.2	Performance level label and description.....	40
4.3	Data analysis.....	42
4.4	Standard setting meeting.....	42
4.5	Final evaluation.....	45
4.6	Final outcomes.....	46
5.	Live data analysis.....	47
5.1	Classical Test Theory.....	47
5.2	Item response theory.....	50
5.3	Sub-group analysis.....	52
5.4	Differential item functioning.....	53
5.5	Analysis by phonics teaching.....	54
5.6	Analysis by term of birth.....	55
6.	Check re-check study.....	56
7.	Inter-rater reliability study.....	58
8.	Children with SEN.....	60
8.1	Identifying groups of children.....	60
8.2	Meetings with stakeholders.....	60
8.3	Administration guide for the pilot and modified versions.....	61
8.4	Findings from the independent evaluation of the pilot in relation to children with SEN.....	62
8.5	Meetings with stakeholder organisations and selected pilot schools.....	63
9.	Arrangements for roll-out.....	67
10.	Conclusion.....	68
10.1	Validity.....	68
10.2	Reliability.....	69
10.3	Comparability.....	70
10.4	Minimising bias.....	71
10.5	Manageability.....	71
10.6	Overall statement in relation to common criteria.....	72
	Annex A – Decision log for the assessment framework.....	73
	Annex B – Specification for item review.....	78
	Annex C – Invitation to schools to participate in the pilot.....	89

Annex D – Administration guide 91
Annex E – Agenda for training day 103
Annex F – ADSR 104
Annex G – Agenda for standard setting meeting..... 127
Annex H – Check-re-check guide 128
References 130

1. Introduction and policy background

The Government has established a check of phonic decoding at the end of Year 1 and to make the results of this check available to parents.

In response to the public consultation on the introduction of the phonics screening check, the Government confirmed its intention to develop a check, which was piloted in summer 2011, with a view to full national roll-out in 2012. The check focuses solely on decoding using phonics and confirms whether children have reached an appropriate standard by the end of Year 1, identifying children who need additional support from their school to catch up.

The purpose of this technical report is to provide evidence for the phonics screening check in relation to the Ofqual common assessment criteria of validity, reliability, comparability, minimising bias and manageability.

1.1 The purpose of the Year 1 phonics screening check

The purpose of the phonics screening check will be to confirm whether or not children have learned phonic decoding to an age-appropriate standard.

Children who have not reached this level should receive extra support from their school to ensure they can improve their decoding skills, and will then have the opportunity to retake the phonics screening check to confirm that they have reached this standard.

Use of data

After full national roll-out it is intended that the following uses will be made of the data:

- Individual children's results will be made available to parents, so that parents are kept informed about their child's progress in developing phonic decoding skills.
- School-level results will be recorded on RaiseOnline, and made available to Ofsted for use in inspections. (School-level results will not be published in performance tables.)
- Report on national results to track standards over time.
- Report national and local authority (LA) results to allow schools to benchmark the performance of their children.

The impact of the Year 1 phonics screening check

- It is hoped that the Year 1 phonics screening check will encourage schools to pursue a rigorous phonics programme for all children at the start of primary school.
- The Year 1 phonics screening check should identify children who have not learned to decode using phonics to an age-appropriate standard by the end of Year 1. These children should then receive additional support to ensure they can improve their decoding skills.
- By promoting the teaching of systematic synthetic phonics and identifying children who need extra support, it is hoped that introducing the Year 1 phonics screening check will

lead to an increase in the number of children able to read competently by the time they reach the end of Key Stages 1 and 2.

Accurate evaluation of the potential impacts of the phonics screening check was not possible during the pilot because the pilot data will not be made available in the ways specified above. However, the impact on children and schools will be monitored over the early years of roll-out.

1.2 Executive summary

The pilot of the Year 1 phonics screening check had three main purposes:

- To develop and trial the phonics screening check materials;
- To seek the views of pilot schools on implementation; and
- To generate evidence in relation to the Ofqual common assessment criteria of validity, reliability, comparability, minimising bias and manageability.

The pilot was conducted by the Department for Education with some support from the Qualifications and Curriculum Development Agency (QCDA). Following the launch of the Standards and Testing Agency (STA) in October 2011, responsibility for the development and delivery of the Year 1 phonics screening check transferred to the STA.

Development

The phonics screening check materials were developed using standard test development process, involving technical assessment specialists, teachers and experts in phonics. The structure of the phonics screening check was subject to consultation, trialling and independent review and the items were each reviewed and trialled with approximately 1000 children. A full standard setting process was also carried out to determine the number of marks required by children to demonstrate that they had achieved the expected standard.

The Department believes that these processes are demonstrably robust and in line with international best practice such that there can be confidence in the outcomes of this process.

Evaluation

An independent evaluation of the pilot was carried out by Sheffield Hallam University, with the majority of schools involved in the pilot providing positive feedback.

Most teachers and children taking part in case studies understood the purpose of the phonics screening check correctly and monitoring visits showed that the administration of the phonics screening check was conducted consistently and appropriately by most schools visited. On average the phonics screening check took four to nine minutes to complete per child and the majority (65 per cent) of pilot schools reported that the time commitment required to administer the phonics screening check was generally 'straightforward' or 'very straightforward' to manage (on average schools spent 3 hours preparing for the phonics screening check and 12.5 hours administering the phonics screening check). Schools reported that the administration guidance

was clear and straightforward and the practice marking workshop at the regional training events was felt to be particularly useful¹.

However, a fifth of schools did report concerns that the time required to administer the phonics screening check was difficult to manage and 54 per cent felt that a longer administration window should be allowed.

The vast majority of schools (90 per cent or more) said that the content of the phonics screening check was suitable for children working at the expected standard of phonics across most aspects of the phonics screening check's content. Lower proportions of pilot schools felt the phonics screening check was suitable in relation to the number of words (83 per cent), the type of vocabulary used in real words (80 per cent) and the use of pseudo-words (74 per cent).

Three quarters of schools felt that the phonics screening check assessed phonic decoding ability accurately overall for their children². However, less than half of teachers were confident in the accuracy of the phonics screening check's assessment for lower ability children and children with special educational needs (SEN), English as an additional language (EAL) and language difficulties (one third of schools were neutral in their response for these groups). In relation to lower ability pupils, it should be noted, though, that the purpose of the phonics screening check is not to assess by how far children are below the standard, only whether they have achieved it or not. Therefore, it is not surprising that pupils who are far below the expected standard may struggle with the phonics screening check.

The phonics screening check helped almost half (43 per cent) of pilot schools to identify children with phonic decoding issues that they were not already aware of, although over half felt it did not help in this way.

The experience of the phonics screening check was perceived to be positive for most children. Sixty-two per cent of pilot schools felt the experience had been positive for all children, and just under a third (31 per cent) said it was neither positive nor negative. Between 23 per cent and 29 per cent of surveyed schools felt the experience was negative for children with speech and language difficulties, other SEN and weak phonics skills.

Common assessment criteria

Validity

The development of a validity argument must start with an understanding of the purpose of the assessment. As stated above, the purpose of the phonics screening check is to confirm whether or not children have learned phonic decoding to an age-appropriate standard such that those children who have not met the standard are provided with additional support to catch-up. As a result there are several questions that need to be answered to ensure that the assessment is sufficiently valid:

- Is the phonics screening check an appropriate assessment of phonic decoding?

¹ Face-to-face training will not take place during roll-out; however, the lessons from the face-to-face training will be implemented to ensure guidance is sufficient to enable consistent administration and scoring of the phonics screening check, including the development of video training materials to replicate elements of the face-to-face training.

² It should be noted that the standard on the phonics screening check had not been set and no results had been provided to schools when they were asked to complete the evaluation.

- Is the expected standard on the phonics screening check age-appropriate?
- Are children who have not met the standard on the phonics screening check in need of additional support?

In relation to the first question, the pilot has collected a great deal of evidence relating to the content of the phonics screening check. The experts involved in the development of the phonics screening check have a wealth of expertise and experience in the field and have validated the specification for the phonics screening check throughout the process. Although the independent experts who reviewed the phonics screening check materials raised some concerns with the specification, these were relatively minor and not consistent. On balance, the evidence from the independent experts provides the Department with sufficient confidence that the phonics screening check is assessing phonic decoding appropriately.

In the evaluation the vast majority of teachers (in most cases over 90 per cent) felt that all elements of the content of the phonics screening check were suitable for children at the end of Year 1. Some concerns were regarding the use of pseudo-words (21 per cent of respondents) and unfamiliar real words (20 per cent of respondents). However, there are strong arguments in favour of including such words in the phonics screening check to ensure children are not relying on visual memory.

The review of the phonics screening check against several phonics programmes also lends weight to the argument that the screening check is appropriate in terms of the phonemes and word structures covered.

Therefore, the Department believes that the phonics screening check is an appropriate assessment of phonic decoding for children at the end of Year 1.

In relation to the second question, the expected standard was developed in conjunction with a group of leading experts in the field. The descriptor was reviewed and independently verified by two groups of practicing teachers involved in the standard setting process. They believed that it was an appropriate expectation to have despite the fact that, based on the pilot data³, a minority of children at the end of Year 1 were currently achieving this standard.

The Department, therefore, believes that the expected standard on the phonics screening check is age appropriate.

The answer to the final question cannot be answered until live results are returned to schools. However, feedback will be requested from schools and will be reported in the final technical report on the pilot and initial roll-out in Autumn 2012.

Reliability

To demonstrate sufficient reliability for the phonics screening check, the following aspects must be considered:

- The internal consistency
- The classification consistency
- The classification accuracy

³ For the reasons stated in section 4.6, it is not possible to generalise national results in roll-out from results in the pilot

- The consistency of scoring

The analysis of the evidence from the pilot has demonstrated high levels of internal consistency for the phonics screening check, with values of Cronbach's alpha of around 0.95. A reasonable standard error of measurement has been identified such that one can be 95 per cent confident that a child's true score lies within two standard errors of measurement of their observed score i.e. we are 95 per cent confident that a child's true score lies within the region of the observed score plus or minus 5 marks.

Classification consistency refers to the extent to which children are classified the same way in repeated applications of a procedure. Evidence from the check-re-check study indicates that approximately 90 per cent of children have been consistently classified.

Classification accuracy refers to how precisely children have been classified. Reasonable estimates of classification accuracy will only be valid once the phonics screening check has been administered in all schools. Therefore, further work on reliability will be analysed and reported in autumn 2012.

Consistency of scoring relates to the extent to which children are classified the same way when scored by different teachers. Evidence from the inter-rater reliability study indicates that even with the limitations of the study, 92 per cent of children have been consistently classified.

At present, the Department is satisfied that the phonics screening check is a sufficiently reliable assessment.

Comparability

When introducing a new assessment such as the phonics screening check, there are often no existing assessments with which to be comparable. However, the pilot has trialled sufficient items to develop up to nine different phonics screening checks. Since all items have been trialled together in a cross-over design, with items appearing in different forms, it has been possible to link all items together on a single scale. This ensures a number of different assessment instruments can be developed to be comparable.

Minimising bias

The qualitative evidence shows that although children with weaker phonics skills, often those children with SEN, find the phonics screening check difficult, the access arrangements that have been put in place make the assessment accessible to children. This has been verified by groups representing all major special educational needs.

It is true that certain groups of children do perform less well on the screening check, e.g. boys, children eligible for free school meals and children with SEN. However, these groups have traditionally performed less well on National Curriculum assessments so this is not unexpected. This also does not mean that the phonics screening check is biased against them since it is possible that the phonics screening check is simply correctly identifying their lower phonics skills.

Analysis has shown a difference in overall scores when comparing the term in which children were born, with younger children likely to do less well on the phonics screening check. This outcome was anticipated based on experience from other National Curriculum tests. However, the Government has high expectations for all children and therefore the policy does not take into account age for the purpose of setting standards on the phonics screening check. Therefore no attempt has been made to account for the age of the child. Schools, however,

should consider this when deciding the appropriate interventions to put in place for children who do not meet the expected standard in the screening check.

The evidence presented from the analysis of differential item functioning for gender, EAL and SEN provides the Department with confidence that there is no strong evidence that the Year 1 phonics screening check is discriminating between children on the basis of anything other than ability to decode using phonics.

Manageability

Most of the evidence relating to the Ofqual common criteria of manageability comes from the evaluation survey discussed previously.

The Department therefore believes that the administration of the screening check is a valuable use of teachers' time as part of their on-going assessment of children and that the evidence from the evaluation of the pilot indicates that the Year 1 phonics screening check is manageable for schools in this context.

Overall

Having examined all of the evidence gathered so far through the pilot, the Department is satisfied that the Year 1 phonics screening check is sufficiently valid for the defined purpose, with acceptable levels of reliability, which is fair for children and manageable for schools. However, as stated previously, additional analysis will be carried out to ensure that the Department can be more confident in this assertion

2. The Assessment framework

The assessment framework sets out the structure of the phonics screening check and provides details of the administrative arrangements. The framework for the pilot was developed in a number of stages and was published on the Department website, although it is no longer available. The framework for the roll-out has been developed from the pilot framework and contains only minor amendments reflecting learning from the pilot. The Assessment framework for roll-out is available on the Department website at www.education.gov.uk/assessment.

2.1 Development of the pilot framework

The pilot framework was initially developed in conjunction with four leading phonics experts:

- Jenny Chew;
- Ruth Miskin;
- Rhona Stainthorp; and
- Morag Stuart.

A series of meetings were held between Department officials, technical assessment specialists, teachers and these phonics experts in September and October 2010. During these meetings, the basic specification for the phonics screening check was developed. This draft was used to gather stakeholder views in the consultation on the Year 1 phonics screening check and to construct a sample check that was administered in an informal trial with 17 schools across the country in November 2010.

The responses to the consultation raised a number of concerns with the framework, some of which were addressed in the subsequent draft. However, although acknowledged, some concerns were not addressed since they would have fundamentally altered the purpose of the phonics screening check. As stated in the Government response to the consultation, published in March 2011:

- 4.9 In the pilot we will continue to ask each pupil to read 40 items with a teacher, which will be a mix of real words and pseudo-words. This is the most effective way to design a reliable check of phonic decoding, which is the purpose of this check.
- 4.10 Pseudo-words used will be new to all pupils, and so there will be no unintended bias based on visual memory of words or vocabulary knowledge (for example pupils with EAL may have a smaller vocabulary and so find reading real words more difficult). Pseudo-words are an established assessment method in many schools, for example they are included in 'Letters and Sounds'. The evidence from the pre-trialling confirms pseudo-words are not confusing for pupils.
- 4.11 We acknowledge the concerns about the use of pseudo-words expressed in the consultation. Through the piloting we will check that pseudo-words perform reliably as a means of assessing phonic decoding for all pupil groups. We will also review whether there should be a majority of real words in the check in light of evidence from the technical piloting.
- 4.12 The pre-trialling suggests the screening check will take 2-3 minutes per pupil to administer. Given this, we think 40 items is a manageable length for pupils and

teachers. We will determine exactly how many questions are required to generate a reliable result through the piloting this summer; if fewer questions are necessary, then we will consider reducing the number of words in the check.

- 4.13 The majority of consultation respondents thought that teachers should administer the check, which will help to inform their teaching, and so we will stipulate that only teachers can administer the check.
- 4.14 We will make some adjustments to the proposals set out in the consultation document to ensure the check is manageable for schools to administer. Based on the responses to the public consultation, we propose to allow a window of one week for the check to take place, and to produce one check to cover the whole window rather than one check per day. We also propose to allow more than one teacher in each school to administer the assessment. This will allow large schools sufficient flexibility to organise the check without disrupting their Key Stage 1 teaching. Schools will be responsible for ensuring that their administration of the check is consistent and fair for all pupils and that it produces accurate and reliable results which can be compared to national outcomes.
- 4.15 We will support teachers to score the check consistently, but we have adjusted some of the arrangements to allow teachers greater scope to use their professional judgement. There will not be a time limit for each response. This will ensure children do not feel pressurised taking the check, and teachers can use their judgement to decide when to move a child onto the next word in the check. We agree with the majority of respondents that self-correction should be encouraged because it demonstrates good reading processes. Where a child immediately self-corrects their response answer they will receive credit for reading the word.
- 4.16 These rules mean the check will not assess fluency of decoding. The feedback from the pre-trialling suggests it is not essential to assess fluency as part of this check, although teachers will continue to be interested in fluency as part of children's wider reading development.

The assessment framework for the pilot was also reviewed against five phonics programmes to determine whether the assignment of grapheme-phoneme correspondences (GPCs) to particular sections in the phonics screening check was consistent with their coverage in the programmes. The five phonics programmes were identified as being suitable to be used in the study based on the revised set of core criteria produced by the Department that define the key features of an effective systematic synthetic phonics teaching programme. The criteria are that the product:

- presents high quality systematic, synthetic phonic work as the prime approach to decoding print, i.e. a phonics 'first and fast' approach;
- enables children to start learning phonic knowledge and skills using a systematic, synthetic programme by the age of five, with the expectation that they will be fluent readers having secured word recognition skills by the end of Key Stage 1;
- is designed for the teaching of discrete, daily sessions progressing from simple to more complex phonic knowledge and skills and covering the major grapheme/phoneme correspondences;
- enables children's progress to be assessed;

- uses a multi-sensory approach so that children learn from simultaneous visual, auditory and kinaesthetic activities which are designed to secure essential phonic knowledge and skills;
- demonstrates that phonemes should be blended in order from left to right ‘all through the word’ for reading;
- demonstrates how words can be segmented into their constituent phonemes for spelling and that this is the reverse of blending phonemes to read words;
- ensures that children apply phonic knowledge and skills as their first approach to reading and spelling even if a word is not completely phonically regular;
- ensures that children are taught high frequency words that do not conform completely to grapheme/phoneme correspondence rules;
- provides fidelity to the teaching framework for the duration of the programme, to ensure that these irregular words are fully learnt; and
- ensures that, as children move through the early stages of acquiring phonics, they are invited to practise by reading texts which are entirely decodable for them, so that they experience success and learn to rely on phonemic strategies.

The programmes⁴ used in this validity study were:

- Sound Discovery
- Speed Sounds (Read Write Inc)
- Letters and Sounds
- Jolly Phonics
- Phonics International

The programme documentation for each programme was used to identify the order in which GPCs were introduced. The results are provided in tables 1 to 5 below:

Table 1 – Sound Discovery

Programme	Sound Discovery
Evidence from	Manual, pages 17 and 24
Step 1	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, qu, r, s, t, u, v, w, x, y, z
Step 2	ai, ar, ch, ee, er (her), ie (tie), ng, oa, oi, oo (book and spoon), or, ou (out), sh, th, ue (cue)
Step 3	a-e, al, au, aw, ay, ck, dge, ea, e-e, ew, ey, ff, gh (f), i-e, igh, ir, nk, o-e, oul, ow, oy, ph, ss, tch, u-e, ur, wh

⁴ The full list of programmes that have met the criteria is available at <http://www.education.gov.uk/schools/teachingandlearning/pedagogy/phonics>. At the time of conducting the study, the process of verifying programmes had not been completed and therefore not all programmes that met the criteria had been identified for inclusion in the study.

Table 2 – Speed Sounds (Read Write Inc)

Programme	Speed Sounds
Evidence from	Lesson plans, pages 3 and 173
Set 1	a, b, c, ch, d, e, f, g, h, i, j, k, l, m, n, ng, nk, o, p, q, r, s, sh, t, th, u, v, w, x, y, z
Set 2	air, ar, ay, ee, igh, ir, oo, or, ou, ow (blow), oy
Set 3 and complex speed sound chart	a-e, ai, are, aw, cious, ea, ear, er, ew, i-e, ire, oa, o-e, oi, ow (cow), tion, tious, u-e, ur, ure

Table 3 – Letter and Sounds

Programme	Letters and Sounds
Evidence from	Phase booklets
Phase 2	a, b, c, ck, d, e, f, ff, g, h, i, k, l, ll, m, n, o, p, r, s, ss, t, u
Phase 3	ai, air, ar, ch, ear, ee, er (her), igh, j, ng, oa, oi, oo, or, ow (cow), qu, sh, th, ur, ure, v, w, x, y, z, zz
Phase 5	a-e, au, aw, ay, ea, e-e, ew, i-e, ie, ir, o-e, oe, ou, ow (blow), oy, ph, u-e, ue, wh

Table 4 – Jolly Phonics

Programme	Jolly Phonics
Evidence from	Teacher's book pages 6 and 11
Page 6	a, ai, ar, b, c, ch, ck, d, e, ee, er (her), f, ff, g, h, i, ie (tie), j, k, l, ll, m, n, ng, o, oa, oi, oo, or, ou (out), p, qu, r, s, sh, t, th, u, ue, v, w, x, y, z
Page 11	a-e, al, au, aw, ay, ea, e-e, ew, i-e, igh, ir, o-e, ow, oy, u-e, ur

Table 5 – Phonics International

Programme	Phonics International
Evidence from	Website
Unit 1	a, c, ck, e, h, l, k, n, p, r, s, t
Unit 2	ai, ay, b, d, f, ff, g, ie, igh, j, l, le, ll, m, o, oa, ow, ss, u, w, y
Unit 3	ea, ee, or, se, wh, z, ze, zz
Unit 4	ch, ng, nk, oo, sh, th, v, ve, x
Unit 5	ar, ce, er, ge, oi, ou, ow, oy, qu, ue
Unit 6	a-e, ae, air, are, ci, cy, e-e, ear, eer, ere, gi, gy, i-e, ier, ir, o-e, oe, oes, our, re, u-e, ur, wor
Unit 7	al, au, aw, al, bu, dge, el, gu, il, kn, mb, oar, oor, ore, our, rh, sc, tch, wr
Unit 8	ci, gh, gg, gn, ous, ph, qua, si, ssi, st, ti, wa, war
Unit 9	aigh, ie, eigh, ey
Unit 10	eu, ew, iew, ui
Unit 11	augh, ough, quar, que
Unit 12	alm, alt, eau, gue, ine, mn, ps, ture

Using these lists, all of the GPCs that appeared in at least two programmes were identified. These were then agreed to be the possible list of GPCs that could appear in the phonics screening check. They were as follows:

a a-e ai air al ar are au aw ay b c ch
ck d dge e ea ear ee e-e eer er ere ew ey
f ff g gh h l i-e ie igh ir j k l
ll m n ng nk o oa o-e oe oi oo or ou
ow oy p ph qu (q) r s sh ss t tch th u
u-e ue ur ure v w wh x y z zz

For each programme, the stage was identified when at which at least one GPC had been introduced for each of the 42 phonemes of English. These were as follows:

- Sound Discovery – step 2
- Speed Sounds – set 2
- Letters and Sounds – phase 3
- Jolly Phonics – page 6
- Phonics International – unit 6

GPCs that appeared in each programme up to the identified point were considered suitable for section 1 in the phonics screening check according to that programme. Those which appeared after were considered appropriate for section 2.

Each GPC was then placed into one of the following categories as shown in table 6 below.

Table 6 – Initial placement of GPCs in categories

Category	GPCs
All programmes place grapheme in section 1	a, ar, b, c, ch, d, e, ee, f, g, h, i, j, k, l, m, n, ng, o, oo, or, p, qu(q), r, s, sh, t, th, u, v, w, x, y, z
All programmes place grapheme in section 2	au, aw, ew
Grapheme does not appear in one programme	e-e, ph, wh
Grapheme does not appear in two programmes	air, al, are, ear, ey, nk
Grapheme does not appear in three programmes	ere, gh, oe, ure
Grapheme placed in section 1 by four programmes and section 2 by the other	ai, oa, oi, u-e
Grapheme placed in section 2 by four programmes and section 1 by the other	a-e, ea, i-e, o-e
Grapheme placed in section 1 by three programmes and section 2 by the other two	igh, ue
Grapheme placed in section 2 by three programmes and section 1 by the other two	ay, ck, ir, oy, ur
Different pronunciations of the grapheme are placed in different sections	ie, ou, ow
Double letters	ff, ll, ss, zz
Consonant trigraphs	dge, tch

For some categories, the decision of whether to assign the GPCs to section 1 or section 2 of the phonics screening check was straightforward. However, for some categories it was decided to

speak to the authors of the programmes to determine the most appropriate assignment to sections.

In consultation with the authors of the programmes changes were agreed to be acceptable for that programme. These changes are shown in table 7 below.

Table 7 – Agreed reassignments of GPCs

Programme	GPCs acceptable in section 1	GPCs acceptable in section 2
Sound Discovery	ck, ff, ll, ss, zz	air
Speed Sounds	ck, ff, ll, oi, ss, zz	e-e
Jolly Phonics		ai, ph, wh

This changed the assignment to categories as shown in table 8.

Table 8 – Final placement of GPCs in categories

Category	GPCs
All programmes place or are content to place grapheme in section 1	a, ar, b, c, ch, ck, d, e, ee, f, g, h, i, j, k, l, m, n, ng, o, oi, oo, or, p, qu(q), r, s, sh, t, th, u, v, w, x, y, z
All programmes place or are content to place grapheme in section 2	air, au, aw, e-e, ew, ph, wh
Grapheme does not appear in one programme	None
Grapheme does not appear in two programmes	air, al, are, ear, ey, nk
Grapheme does not appear in three programmes	ere, gh, oe, ure
Grapheme placed in section 1 by four programmes and section 2 by the other	ai, oa, u-e
Grapheme placed in section 2 by four programmes and section 1 by the other	a-e, ea, i-e, o-e
Grapheme placed in section 1 by three programmes and section 2 by the other two	igh, ue
Grapheme placed in section 2 by three programmes and section 1 by the other two	ay, ir, oy, ur
Different pronunciations of the grapheme are placed in different sections	ie, ou, ow
Double letters	ff, ll, ss, zz
Consonant trigraphs	dge, tch

Each category was then assigned to an appropriate section as shown in table 9 below.

Table 9 – Assignment of categories to section

Category	Section assignment
All programmes place or are content to place grapheme in section 1	Section 1
All programmes place or are content to place grapheme in section 2	Section 2
Grapheme does not appear in one programme	Do not include in check
Grapheme does not appear in two programmes	Do not include in check
Grapheme does not appear in three programmes	Do not include in check
Grapheme placed in section 1 by four programmes and section 2 by the other	Section 2
Grapheme placed in section 2 by four programmes and section 1 by the other	Section 2
Grapheme placed in section 1 by three programmes and section 2 by the other two	Section 2
Grapheme placed in section 2 by three programmes and section 1 by the other two	Section 2
Different pronunciations of the grapheme are placed in different sections	Section 2
Double letters	Section 1
Consonant trigraphs	Do not include in check

This meant the GPCs were assigned to sections as follows:

Section 1 a ar b c ch ck d e ee f ff
 g h l j k l ll m n ng o
 oi oo or p qu(q) r s ss sh t th
 u v w x y z zz

Section 2 a-e ai air au aw ay e-e ea ew i-e ie
 igh ir o-e oa ou ow oy ph u-e ue ur wh

As a result of the review, the following changes were made to the section assignment compared to the consultation document:

er (/ ɜ:/) moved from section 1 to section 2

ey (/ i:/ & / eɪ /) removed from check

oe (/ əʊ /) removed from check

oo (/ u:/ & / ʊ /) moved from section 2 to section 1

th (/θ/ & /ð/) moved from section 2 to section 1

ur (/ 3:/) moved from section 1 to section 2

As a result of the evidence, the pilot framework was amended and used to develop the pilot materials. The decision log for the assessment framework, which sets out the evidence for each decision relating to the content of the framework, can be found in Annex A. The framework was published on the Department website on 3 June 2011.

2.2 Independent review of the assessment framework

Once the framework for the pilot had been finalised, it was sent for external review by national and international independent phonics experts, i.e. people who had not been involved in the development of the phonics screening check. The experts involved were Alison Bailey, Brian Byrne, Rhona Johnston, Maggie Snowling and Janet Vousden.

The experts have a variety of experience including:

- teaching;
- literacy consultancy;
- trialling teaching material for introducing early decoding skills and phonemic awareness; and
- research on topics such as:
 - early literacy and language development
 - impact of systematic phonics teaching on vulnerable groups
 - reading development
 - reading disorders
 - nonword reading
 - the teaching of analytic and synthetic phonics.

Three of the experts have 30+ years in the field, and the other two have 11 and 19 years' experience.

Content

There was general agreement regarding the majority of grapheme-phoneme correspondences (GPCs) that are included in the phonics screening check. There were a few examples given by some of the experts of GPCs that they did not believe should be included and some which should, although there was also acceptance of the rationale for not selecting certain GPCs. However, the lack of consensus amongst the experts does not provide convincing evidence that the original selection of GPCs was inappropriate.

All of the experts agreed that there weren't any orthographical word structures included that shouldn't be. Three of the experts felt there were other orthographical word structures that could be included, although their suggestions differed. One of the experts felt that the balance of different orthographical structures in section one was dominated by words containing digraphs, although the concern was linked to a desire to make the phonics screening check more diagnostic, which is not part of the purpose. In general, however, this does seem to provide evidence that the orthographical word structures are suitable for the phonics screening check.

Cognitive domain

There were some concerns raised with some of the terminology used in the cognitive domain, but again these concerns were not consistent amongst experts.

One expert felt that there were some elements of the cognitive domain statement that are not appropriate for the phonics screening check, specifically the reference to phoneme awareness in 3.3.1 and the term 'segmenting' in 3.3.2.

Three of the experts believed there were elements missing from the cognitive domain statement from what they would consider an accurate reflection of the required cognitive skills needed to decode using phonics successfully, although one of these was questioning the wording/phrasing rather than suggesting additional elements. The other two suggest different things – that alternative pronunciations of real words should be allowed and that secure knowledge of multi-letter graphemes was necessary, including the ability to segment into appropriate sized graphemes. Another expert highlighted that phonemic awareness should be assessed in the follow-up testing of children who did not achieve the expected standard before intervention programmes are developed.

As a result of all these comments, the cognitive domain statement has been revised (see section 2.3)

Structure

All of the experts agreed or strongly agreed that 40 words are sufficient to provide a broad assessment of whether the child has demonstrated an ability to decode using phonics and that this would be manageable for the majority of children at the end of Year 1. One expert was concerned about children with significant learning difficulties and therefore wanted to ensure that schools were made aware of the discontinuation rule that has been developed to help teachers to know when to stop the phonics screening check.

All of the experts strongly agreed that the phonics screening check should contain pseudo-words. One expert disagreed with the inclusion of real words. Another agreed it was useful to include them but felt this highlighted a problem with nomenclature, and that with the inclusion of real words it is no longer a check purely of phonics. They thought the proportion of pseudo-words to real words was appropriate but noted that only low-frequency words would require decoding using phonics. One of the experts that agreed with the inclusion of real words felt that more words in the phonics screening check should be pseudo-words. The others who agreed with the inclusion of real words thought that the proportions were appropriate.

Two experts disagreed that the phonics screening check should contain a mixture of high and low frequency words but they had opposing views on the proportion of real words that should be low frequency. Overall two experts felt that 0 per cent – 20 per cent of real words should be low frequency (one because the non-words do the same job), two felt the proportion should be 41 per cent to 60 per cent and one felt it should be between 81 per cent and 100 per cent (to prevent recognition of words which is not a test of decoding).

All experts agreed that the descriptions on pages 10 and 13 of the assessment framework provide for sufficient sampling from the content and cognitive domains.

Two of the experts disagreed with the decision not to include two-syllable pseudo-words in the phonics screening check because of the difficulty in ensuring consistent scoring, although one

of these acknowledged that including them would lead to the omission of something else and therefore wasn't particularly concerned.

All of the experts agreed that the constrained unigrams and trigrams (Appendix B of the assessment framework) were appropriate and sufficient for the phonics screening check. Two experts disagreed with the list of constrained bigrams, although one felt some were missing and the other that some shouldn't be in the set.

There was agreement amongst the experts on the appropriateness of the item specification from page 16–22 of the Assessment framework for each of the orthographical structures. One of the experts did make some comments/suggestions.

In general, the evidence on the structure of the phonics screening check seems to support the conclusion that the phonics screening check is a sufficiently valid assessment of decoding using phonics.

Administration

Four relatively minor suggestions were made about the administration guide to ensure consistency of administration. Two experts suggested that all children should be given the practice items and one that a guide to acceptable pronunciation of pseudo-words was needed and guidance on acceptable encouragement. Although the use of practice items will remain optional in the phonics screening check, the other suggestions will be addressed in guidance for the roll-out.

Two of the experts disagreed that the inclusion of a context for pseudo-words was necessary, though they were not concerned that it would invalidate the outcomes. All of the experts agreed that the context selected was appropriate.

All of the experts agreed that the standard version will be accessible to the majority of children at the end of Year 1. When asked if the school-based modifications would impact on the interpretation of the outcome for the children who use the modified version the experts did not anticipate any significant impact, with the possible exception of BSL. One expert felt that the manipulation of font size might affect word reading (but not non-word reading). They all agreed that the scoring of alternative pronunciations of graphemes used in pseudo-words. Four agreed that the alternative pronunciations of graphemes used in real words (explained on page 26 of the Assessment framework) were appropriate, with the caveat that a list should be provided of the acceptable alternatives for the pseudo words and that some leniency should also be given to low frequency real words. One expert felt that the scoring for real words could disadvantage children with more limited English vocabularies.

All of the experts agreed that allowing children to self-correct is appropriate. Four of the experts agreed that allowing children as long as necessary to respond is appropriate but one felt that a guide should be given to time allowed per item as allowing too much time could be demoralising.

Phonics screening check construction

All of the experts agreed that the list of real words that could be included in the phonics screening check described in the trial word generation document was sufficient. Four of the experts agreed that the process for developing the list of pseudo-words that could be included in the phonics screening check described in the trial word generation document was sufficient.

The other expert thought that the word-likeness was important and that the neighbourhood size alone may not be sufficient to do this.

The experts all agreed that the item review process described in the Item review specification was sufficient.

Overall

In general, the experts believed that the phonics screening check would have a positive effect on classroom practice, leading to fewer children finishing Key Stage 1 without the necessary grasp of phonic decoding.

Three of the experts agreed or agreed strongly with the statement 'In my professional opinion the Year 1 phonics screening check is a sufficiently valid assessment of the construct of phonics to meet the stated purpose'. Of the two who did not agree, one did not answer the question because they wanted to know the threshold from standard setting, which had not taken place at this time. The other felt that the phonics screening check was not, and should not, be solely concerned with phonic decoding and that a time limit should be imposed for children to respond.

2.3 The Assessment framework for roll-out

The Assessment framework for the roll-out was published in January 2012. The following changes were made to the pilot framework as a result of the findings from the independent review and the evaluation report on the pilot (see section 3.5):

- Inclusion of a picture of an imaginary creature next to each pseudo-word (in response to evaluation findings)
- Provision of a greater number of words on the practice sheet (in response to feedback from pilot schools and SEN organisations)
- Amendment to the cognitive domain statement as shown below in red (in response to independent review findings)

Cognitive domain

To respond correctly to the items in the Year 1 phonics screening check, children need to be familiar with the content domain being assessed, but they also need to draw on a range of cognitive skills. Describing these skills plays a crucial role in the development of any assessment, since they are vital in ensuring that the phonics screening check covers the appropriate range of cognitive skills across the content domains already outlined.

The first domain, 'knowing', covers the facts, concepts, and procedures children need to know, while the second, 'applying', focuses on the ability of children to apply knowledge and conceptual understanding to read words.

Knowing

The ability to decode phonically depends on knowledge of and familiarity with grapheme-phoneme correspondences (including multi-letter graphemes) and knowing how to blend phonemes into words. This knowledge is dependent on the children having secure knowledge of the letters of the alphabet; having an ability to parse letter strings into

appropriately sized graphemes; and being phonemically aware. Knowledge of the alphabet includes recognising each letter as a discrete visual identity and its own sound value. Phonemic awareness is defined as explicit ability to reflect upon and manipulate the sounds in words. Specific knowledge of grapheme-phoneme correspondences (and phoneme-grapheme correspondences) is dependent on being able to map. The more relevant knowledge a child is able to recall and the wider the range of decoding rules he or she has understood, the greater the potential for reading a wider range of phonically decodable words. Children need to be able easily to recall the basic facts and conventions of phonic decoding in order to read unfamiliar words.

Applying

The applying domain involves the application of knowledge to a range of phonically decodable words in order to be able to read fluently. Children should have confidence in blending using appropriate pronunciations of phonemes for the given context. In relation to the phonics screening check, the context is the letter string of the word or pseudo-word to be read. They should be able to parse the sequence of letters to generate the correct sequence of phonemes which they then blend into the correct word or pseudo-word.

In addition, sections on administration arrangement, access arrangements and reporting were removed from the Assessment framework because they repeated guidance that is now located in other documents such as the *Assessment and reporting arrangements* booklet (ARA) which is available on the Department website at www.education.gov.uk/assessment.

3. The pilot

The Year 1 phonics screening check was piloted in 296 schools in June 2011. Pilot schools were selected in one of the following categories:

- A random sample of schools that were invited to participate;
- Schools that were invited to take part in the pre-trialling;
- Schools that requested to take part in the pilot; and
- Schools that were invited to take part as part of the SEN studies that were undertaken.

Full details of the pilot schools are given in section 3.2. Pilot schools attended face-to-face training to prepare them for administering the phonics screening check, which was administered during the week of 13–17 June 2011.

3.1 Instrument development

In the pilot, 18 different versions, or forms, of the phonics screening check were trialled in pilot schools. As set out in the framework, 40 words were included in each form and a total of 360 words were trialled, meaning that each word appeared in two forms of the phonics screening check. The words used in the pilot forms were selected in the following way.

Real words

The online word database www.neuro.mcw.edu/mcword/ was used to generate all real words with three to eight letters (36,117 words returned) and determine their neighbourhood size⁵. This database was suggested by Rhona Stainthorp, one of the experts involved in the development of the assessment framework. A number of possible databases could have been used for this part of process, but this database was selected because it produced the most words.

The database used by MCWord is based on the CELEX *efw.cd* file. This file includes all the English word forms from a COBUILD corpus of both written and spoken text, which contains approximately 17,900,000 instances of word use. There are approximately 16,600,000 written examples, and 1,300,000 spoken examples. This was felt to be sufficient to generate all possible real words.

The words returned by the MCWord database were input into the Children's Printed Word Database (CPWD) www.essex.ac.uk/psychology/cpwd to determine whether they appear in children's books and were therefore eligible to be used in the phonics screening check (7,753 words returned). The CPWD was also used to determine each word's frequency in the database and its number of syllables. The aim of the CPWD project was to construct a database of words that appear in books for children aged five- to nine-years-old, making the words appropriate for children in the age range of the phonics screening check. Table 10 shows the number of words generated at each stage of the process.

⁵ The neighbourhood size for a word is the number of real words that can be created by changing just one letter. For example, the word 'hat', has the neighbours 'cat', 'hit', 'has', etc.

Table 10 – Number of words generated through process

Number of letters	Initial number of words	Number of words in CPWD
3	708	377
4	2481	1156
5	4703	1632
6	7398	1745
7	10134	1694
8	10693	1149
Total	36117	7753

The orthographical structure of all the real words that could be included in the phonics screening check was then determined and words were grouped by word length, the number of syllables and orthographical structure. The following code was used to classify the words orthographically:

C = consonant

V = vowel

CC = consonant digraph (to enable words to be grouped appropriately when sorted in Excel for selection in the phonics screening check, the consonant digraph 'qu' was also categorised this way rather than the more conventional CV)

VV = vowel digraph (to enable words to be grouped appropriately when sorted in Excel for selection in the phonics screening check, the r-, w- and y-controlled vowel digraphs such as 'ar', 'aw' and 'ay' were also categorised this way rather than the more conventional VC)

VCV = split digraph

VVV = vowel trigraph (to enable words to be grouped appropriately when sorted in Excel for selection in the phonics screening check both vowel trigraphs 'air' and 'igh' were categorised in this way rather than the more conventional VVC and VCC respectively)

The reason that non-conventional labelling was used was to facilitate easier grouping of words with particular features in various IT packages.

Words were then selected to ensure the following coverage:

- Use of all letters of the alphabet;
- Use of all appropriate digraphs for the relevant section;
- Use of all appropriate consonant clusters;
- Range of word frequencies; and
- Range of neighbourhood sizes.

Table 11 shows number of words that were selected for further review in each orthographical structure.

Table 11 – Number of real words selected for further review

Section	Structure	Number of words
1	<u>CCVC</u>	10
1	<u>CVCC</u>	10
1	<u>CVVC</u>	12
1	<u>CCVCC</u>	10
1	<u>CCVVC</u>	8
1	<u>CCVC</u>	10
1	<u>CCVVC</u>	10
1	<u>CCVCC</u>	10
1	<u>CVCC</u>	10
1	<u>CVVCC</u>	10
1	<u>CCVCC</u>	10
2	<u>CVVC</u>	10
2	<u>CVCV</u>	10
2	<u>CCVCV</u>	10
2	<u>CCVCV</u>	5
2	<u>CCVVC</u>	6
2	<u>CCCVCC</u>	3
2	<u>CVVCC</u>	6
2	<u>CVV</u>	5
2	<u>CVVV</u>	5
2	<u>CCVVV</u>	5
2	<u>CCVCC</u>	14
2	<u>CCVVCC</u>	14
2	<u>CCCVCC</u>	2
2	<u>CCCVC</u>	8
2	<u>CCCVCV</u>	5
2	<u>CCCVCC</u>	2
2	<u>CCCVV</u>	4
2	2 syllable 5 letter	15
2	2 syllable 6 letter	15
2	2 syllable 7 letter	15
2	2 syllable 8 letter	15

These were then reviewed against the following criteria by a consortium led by the University of Reading Institute for Education and involving the Institute for Education at the University of London to ensure the words:

- met the requirement of the assessment framework;
- were phonically decodable, taking into account regional accents; and
- did not have an inappropriate meaning colloquially or in a regional dialect.

The specification for the review work is included in Annex B.

Once the final list of appropriate words for inclusion in the Year 1 phonics screening check was agreed, words were selected for inclusion in word blocks. A word block was a group of four words that would appear on one page of the phonics screening check. The orthographical structures of words in each word block are given in the assessment framework. Within each word block, words were checked to ensure:

- as far as possible, words did not start with the same letter (this was not possible on words containing three consonant clusters which all begin with the letter s);
- the words did not form a sentence when read in sequence;
- there was not an over-use of a letter with low frequency in common usage (for example j or x) in a word block;
- when two or more consonant or vowel digraphs appeared in a word block, different digraphs were selected; and
- each word block contained a mixture of high and low frequency words.

See below for further detail on the creation of forms for use in the pilot.

Pseudo-words

A list of all possible words with the orthographical structures included in the phonics screening check was generated using the rules defined in the Assessment framework. For example, CCVC words were all generated using appropriate consonant clusters at the start of words (bl, br, cl, cr, dr, dw, fl, fr, gl, gr, kr, pl, pr, sc, sk, sl, sm, sn, sp, st, sw, tr, tw) all vowels (a, e, i, o, u) and all consonants that are allowed to appear at the end of words (b, d, f, g, h, k, l, m, n, p, r, s, t, w, x, y, z).

All these words, which included both real and pseudo-words were then checked within www.neuro.mcw.edu/mcword to determine their neighbourhood size.

To determine whether the words could be used as pseudo-words, they were also checked, using Google as appropriate, to ensure they met the following criteria:

- They were not real words.
- They were not forenames.
- They were not a common abbreviation.
- They were not a homophone for a real word.
- They were not a common acronym.
- They were not a common brand.
- They were not inappropriate for some other reason.

Pseudo-words were then selected to ensure the following coverage:

- Use of all letters of the alphabet.
- Use of all appropriate digraphs for the relevant section.
- Use of all appropriate consonant clusters.
- Range of neighbourhood sizes.

Table 12 shows number of pseudo-words that were selected for further review in each orthographical structure.

Table 12 – Number of pseudo-words selected for further review

Section	Structure	Number of words
1	CVC	40
1	VCC	15
1	<u>CCVC</u>	10
1	<u>CVCC</u>	12
1	<u>CVVC</u>	15
1	<u>CCVCC</u>	3
1	<u>CCVVC</u>	10
1	CCVC	10
1	<u>CCVVC</u>	10
1	<u>CCVCC</u>	10
1	CVCC	10
1	<u>CVVCC</u>	10
1	<u>CCVCC</u>	10
2	<u>CVVC</u>	10
2	<u>CVCV</u>	10
2	<u>CCVCV</u>	10
2	<u>CCVCV</u>	5
2	<u>CCVVC</u>	6
2	<u>CCCVCC</u>	3
2	<u>CVVCC</u>	6
2	<u>CVV</u>	5
2	<u>CVVV</u>	5
2	<u>CCVVV</u>	5
2	CCVCC	13
2	<u>CCVVCC</u>	13
2	<u>CCCVCC</u>	4
2	CCCVC	10
2	<u>CCCVCV</u>	10
2	<u>CCCVCC</u>	5
2	<u>CCCVV</u>	5

These were then reviewed against the following criteria by a consortium led by the University of Reading Institute for Education and involving the Institute for Education at the University of London to ensure the words:

- met the requirement of the specification;
- was phonically decodable, taking into account regional accents;
- was not a homophone in English;
- was not a homophone for an inappropriate word in another language covered by the review (common words in other languages should also be flagged although it may still be deemed appropriate to include these words); and
- was not used colloquially or in a regional dialect.

Once the final list of appropriate words for inclusion in the Year 1 phonics screening check was agreed, words were selected for inclusion in word blocks. The orthographical structures of words in each word block are given in the assessment framework. Within each word block, words were checked to ensure:

- as far as possible, words did not start with the same letter (this was not possible on words containing three consonant clusters which all begin with the letter s);
- there was not an over-use of a letter with low frequency in common usage (for example j or x) in a word block; and
- when two or more consonant or vowel digraphs appeared in a word block, different digraphs were selected.

Creation of pilot forms

Word blocks were named using the following convention – section (1 or 2), word type (R = real or P = pseudo), page number within the phonics screening check (i-x), form letter (A-J, excluding I) – for example 1RivB.

Word blocks were then organised into the 18 forms to be used in the pilot using the tables 13 to 15.

Table 13 – Assignment of word blocks to forms 1 to 6

	Form 1	Form 2	Form 3	Form 4	Form 5	Form 6
Page 1	1PiA	1PiB	1PiC	1PiD	1PiE	1PiF
Page 2	1PiiJ	1PiiA	1PiiB	1PiiC	1PiiD	1PiiE
Page 3	1PiiiH	1PiiiJ	1PiiiA	1PiiiB	1PiiiC	1PiiiD
Page 4	1RivG	1RivH	1RivJ	1RivA	1RivB	1RivC
Page 5	1RvF	1RvG	1RvH	1RvJ	1RvA	1RvB
Page 6	2PviE	2PviF	2PviG	2PviH	2PviJ	2PviA
Page 7	2PviiD	2PviiE	2PviiF	2PviiG	2PviiH	2PviiJ
Page 8	2RviiiC	2RviiiD	2RviiiE	2RviiiF	2RviiiG	2RviiiH
Page 9	2RixB	2RixC	2RixD	2RixE	2RixF	2RixG
Page 10	2RxA	2RxB	2RxC	2RxD	2RxE	2RxF

Table 14 – Assignment of word blocks to forms 7 to 12

	Form 7	Form 8	Form 9	Form 10	Form 11	Form 12
Page 1	1PiG	1PiH	1PiJ	1PiA	1PiB	1PiC
Page 2	1PiiF	1PiiG	1PiiH	1PiiH	1PiiJ	1PiiA
Page 3	1PiiiE	1PiiiF	1PiiiG	1PiiiF	1PiiiG	1PiiiH
Page 4	1RivD	1RivE	1RivF	1RivD	1RivE	1RivF
Page 5	1RvC	1RvD	1RvE	1RvB	1RvC	1RvD
Page 6	2PviB	2PviC	2PviD	2PviJ	2PviA	2PviB
Page 7	2PviiA	2PviiB	2PviiC	2PviiG	2PviiH	2PviiJ
Page 8	2RviiiJ	2RviiiA	2RviiiB	2RviiiE	2RviiiF	2RviiiG
Page 9	2RixH	2RixJ	2RixA	2RixC	2RixD	2RixE
Page 10	2RxB	2RxC	2RxD	2RxE	2RxF	2RxF

Table 15 – Assignment of word blocks to forms 13 to 18

	Form 13	Form 14	Form 15	Form 16	Form 17	Form 18
Page 1	1PiD	1PiE	1PiF	1PiG	1PiH	1PiJ
Page 2	1PiiB	1PiiC	1PiiD	1PiiE	1PiiF	1PiiG
Page 3	1PiiiJ	1PiiiA	1PiiiB	1PiiiC	1PiiiD	1PiiiE
Page 4	1RivG	1RivH	1RivJ	1RivA	1RivB	1RivC
Page 5	1RvE	1RvF	1RvG	1RvH	1RvJ	1RvA
Page 6	2PviC	2PviD	2PviE	2PviF	2PviG	2PviH
Page 7	2PviiA	2PviiB	2PviiC	2PviiD	2PviiE	2PviiF
Page 8	2RviiiH	2RviiiJ	2RviiiA	2RviiiB	2RviiiC	2RviiiD
Page 9	2RixF	2RixG	2RixH	2RixJ	2RixA	2RixB
Page 10	2RxC	2RxD	2RxE	2RxF	2RxG	2RxH

3.2 School selection

Sample and invitation

In order to trial items with 10,000 children, a sample of approximately 300 schools was judged necessary and would be stratified by geographical region and school reading attainment at Key Stage 1. Based on response rates for National Curriculum test (NCT) test development, a sample of 450 schools was invited to express an interest in participating in the pilot on 14 January 2011 using the Department's schools' email. A copy of the invitation is attached at Annex D.

As a representative sample was essential, schools were asked to express their interest in participating, rather than being guaranteed participation if they wanted to be involved. Using the screeningcheck.phonics@education.gsi.gov.uk email, responses were tracked and a list of expressions of interest maintained ahead of analysis to determine who would be included in the pilot. Participation in the pilot was optional for schools.

Response rate

After an initial flurry of interest, a week after the invitation was issued the overall response rate was still low. Chasing calls were made to schools as well as making use of local authority contacts and exploring if other communication activities where possible. The chasing calls revealed that a number of headteachers had not received the invite email, either due to incorrect details on Edubase, the Department's database of schools, or problems with junk mail filters. It's also possible that some saw the invite but did not register that it was for action.

Chasing activity helped to boost the response rate, however it remained low at just over 100 schools after almost two weeks. A second, representative sample of 459 schools was selected and email invitations issued via the Department's schools' email on 4 February 2011, with an amended closing date of 18 February 2011. Responses were taken for a couple of weeks beyond the date, due to the low response rate and problems with email receipt.

Due to the second sample and chasing activity, an increased response rate was achieved with the response number of schools reaching 200. To achieve a representative sample, responses were analysed and focused chasing phone calls were undertaken to around 45 schools to boost response rates in two regions, along with lower and middle attaining schools.

On 11 February 2011, contact details for LAs with high numbers of schools in the sample were obtained from QCDA and approximately 15 LAs were sent details of their schools in the sample and asked to chase their responses where possible. This again helped to boost the response rate.

Additional schools

In addition to the sample, invitations were also sent to some additional schools that expressed an interest in the pilot, e.g. Primary Heads Reference Group members, schools that had indicated a desire to be involved in their consultation response, schools that participated in the pre-trial, schools identified by LA contacts or other schools. These were self-selecting, rather than random, and so could potentially have skewed the sample. A total of 48 additional responses were received, and it was decided that these schools should be used as a top-up and advocates, relying on the sample schools for data analysis purposes. All 48 of these schools were asked to take part in the pilot on 28 March 2011.

Final sample

A total of 254 schools from the two samples expressed an interest in participating by 3 March 2011. These responses were analysed by the Department, in order to ensure pilot schools were representative by geographical region and reading attainment at Key Stage 1. These 254 schools then received confirmation of participating in the pilot on 25 March 2011.

The final number of schools participating in the pilot was 298. This included 229 schools from the two samples, 15 pre-trial schools, 50 additional schools and four SEN schools.

3.3 The Administration guide

The Administration guide was initially developed by someone with several years' experience of developing the Assessment and reporting arrangements (ARA) and Test administration guide (TAG) for national curriculum tests (NCTs) and single level tests (SLTs).

The current TAG was used as a template for the guide, although not all sections of the current TAG were relevant given the different nature of the administration for the Year 1 phonics screening check. The sections of the Administration guide were chosen to reflect the process that administrators would need to go through to effectively administer the phonics screening check.

The Administration guide was reviewed by a number of people to refine the content including:

- Department policy staff with responsibility for reviewing NCT documentation;
- a former primary school headteacher and marking programme leader for NCTs; and
- the organisation responsible for the evaluation and monitoring of the pilot⁶.

During the evaluation (see section 3.5), schools reported that the administration guidance was clear and straightforward. The vast majority (89 per cent) of pilot schools felt the guidance was 'useful' or 'very useful'. A copy of the administration guide is available in Annex D.

⁶ The organisation responsible for the evaluation and monitoring of the pilot did not carry out a formal review. However, its questions regarding the evidence that monitors need to collect during monitoring visits was used to improve the wording of the guide.

3.4 Training

With around 300 schools participating in the pilot, more than one event was needed to enable teachers to attend around school holidays and other school commitments. In addition, pilot schools were located across the country, so a geographical spread of venues was needed. To keep the events to a manageable size for breakout groups to discuss scoring, it was decided to hold four events with approximately 75 attendees at each. Therefore, three events were held at Department buildings (two in London and one in Sheffield) one event was held in the Midlands, at the Department for Work and pensions (DWP) offices in Coventry.

The programme for the day was designed to cover the core elements of administering and scoring the phonics screening check, but was also intended to provide additional professional development related to the teaching of phonics. To ensure the latter, the Department's Effective Practice Team (EPT) developed and delivered elements of the training events. The agenda for the training day can be found at Annex E.

For the test administration and marking and scoring sessions, Marian Shepard, a former marking programme leader for NCTs, was appointed as a scoring consultant to provide expert advice.

The following report was prepared by the scoring consultant following the completion of training.

Development of training materials

The materials were developed over a very short time – approximately two weeks. At the first meeting with DfE officials it was agreed that I would prepare the marking, scoring and reporting parts of the presentation on the administration of the check and the training script for use by facilitators in the practice of scoring in breakout groups. DfE were to provide recordings of children attempting the practice words (which had been selected to demonstrate various judgement issues) and slides showing the scoring materials.

The recording of children's efforts proved to be insufficiently clear/loud for use in the scoring training sessions. The resultant 'contrived' recording also presented some difficulties as a twelve-year-old struggled to blend slowly/lisp/self-correct as required to demonstrate training points. Furthermore, even when we thought the clips were clear, we later found that some teachers had different perceptions of some words.

Once the recordings had been selected, I worked up a script for the facilitators to use in training. The draft of this and of the presentation slides showing the materials and explaining the marking and scoring of the check and the reporting thereof were passed to DfE for consideration. Following discussion, some adjustments were made.

The presentation and script were then trialled with the group of facilitators and this led to one or two further amendments for clarification of training points.

This process was valuable and resulted in materials that were fit for the purpose of the pilot but I would make the following recommendations for consideration when preparing online training for the roll-out of the national screening check:

Recommendations:

- Considerably more time would be needed to allow for the preparation of training that does not include face-to-face discussion.

- Much more time would need to be devoted to the recording process to ensure that a number of these very young children produced enough examples of different responses to words and that there are sufficient clear examples to use.
- Video recordings should be used, enabling teachers to use visual cues and providing an experience closer to the actual check with their own children.
- More than one marking/training expert to develop the materials as a team with the opportunity to consult a phonics expert.
- Trialling of the online training with groups of teachers from different types of schools early enough in the academic year to allow for a second trial should it be necessary and with at least some of the teachers involved doing a run through of the training words with some children to see if any problems ensue. Feedback from all participants should be used to refine materials as necessary.

The training days

The first day at Coventry had the smallest number of teachers with possibly the most manageable venue for the size of group. Delegates had opportunities throughout the day to raise questions. This was invaluable for the resolution of general and individual concerns. The team was able to incorporate the answers to some of these questions into the dialogue for the following meetings.

The breakout group sessions threw up some problems with the recording of words, especially 'plastic'. This led to the inclusion of an initial explanation of the rationale and process of the recordings at the subsequent meetings to ensure that all delegates had a clear and consistent understanding of the scoring process.

All four meetings were well-received by almost all delegates. They did highlight some inconsistencies between some teachers' understanding of some phonemes – particularly the long 'o'. There was also some lively discussion about pronunciation issues; these will need to be addressed in the online training when live discussion will not be possible. Similar questions were raised at all meetings and, although many of these are covered in the administration guidance, they might usefully be addressed in the online training.

The next steps

The consideration of the success of the training materials at face-to-face meetings lead me to make the following additional suggestions/recommendations for the preparation of online training:

- Screen Materials – unpack the advantage of using practice sheets with the less able/confident children to help them with their delivery and ensure that they understand about pseudo words.
- Refine and explain the use of the grapheme/phoneme handout.
- Scoring – integrate slides' explanations of principles with practice examples in groups.
- Recording – some teachers would have liked more examples of responses to each word and this would be even more helpful when discussion is not possible. Include: different pronunciations of pseudo words; examples of different regional and EAL pronunciations of real words; different examples of speech impediments.
- Include a video recording of a group of teachers listening to the above recordings and discussing the scoring (resulting in the correct decision!)

- The frequently asked questions that arose at the meetings and the answers to these could possibly be put together in some way for inclusion in the online training.
- One of the most useful features of the training days was the opportunity for teachers to discuss issues of administration and scoring of the check. In order to avoid Year 1 teachers (particularly those in single class entry schools) feeling isolated, they could be encouraged to involve colleagues in Reception and Year 2 and literacy co-ordinators to do the training as well and to discuss the implications for the other year groups, perhaps including consideration of how the Year 1 teacher could be supported during the administration time. Obviously, schools in cluster groups or with good support from local authorities would have opportunities for discussion with teachers in other schools.

All of the recommendations in the scoring consultant's report are being implemented for the roll-out.

Feedback from the evaluation was highly positive about training, with the report noting that 'the practice marking workshop at the regional training events was felt to be particularly useful' and that 'case study teacher comments about the training events were highly positive. In particular, the practice marking workshop was seen as essential in giving teachers confidence in administering the check.'

3.5 Evaluation

This report pulls together evidence from a number of strands of work through the pilot. One major strand of evidence is taken from the independent evaluation of the pilot, undertaken by Sheffield Hallam University.

The aims of the evaluation were to:

- assess how the phonics screening check pilot was perceived by schools, parents/carers and children;
- evaluate the effectiveness of its administration; and
- carry out a series of monitoring visits to schools to assess the extent to which the administration of the phonics screening check pilot was standardised.

The objectives of the evaluation included:

- to gather school, parent/carer and child perceptions of the phonics screening check pilot;
- to identify what (if any) information parents/carers would like on the phonics screening check pilot and how they would like this communicated;
- to monitor and gather perceptions of the phonics screening check pilot administration process and corresponding guidance; and
- to identify which phonics programmes are currently taught in schools participating in the pilot and how these are delivered.

Methodology

The following research methods were used to address the evaluation objectives:

- two surveys (using combined online and postal methods) conducted with lead teachers for the phonics screening check in all 300 pilot schools, with response rates of 97 per cent (first survey) and 90 per cent (second survey). The first survey focussed on how phonics teaching is currently delivered in pilot schools, and took place a few weeks before the phonics screening check took place. The second focussed on the administration and content of the phonics screening check, and was administered shortly after the phonics screening check took place;
- case studies carried out in 20 schools, which included interviews with a senior leader, the phonics screening check lead teacher (where the two were different) and small groups of children, addressing similar issues to the second survey but asking for more detailed explanations from a wider group of respondents;
- monitoring visits to a further 20 schools; and
- a survey of parents/carers, with a response rate of 26 per cent from participating schools.

Key findings

Purpose and administration of the phonics screening check

- Most teachers and children taking part in case studies understood the purpose of the phonics screening check correctly.
- Monitoring visits showed that the administration of the phonics screening check was conducted consistently and appropriately by most schools visited.
- On average the phonics screening check took four to nine minutes to complete per child, although this varied depending on child knowledge and educational need. Just over half (54 per cent) of schools felt that a longer window of time was needed to carry out the phonics screening check.
- The majority (65 per cent) of pilot schools reported that the time commitment required to administer the phonics screening check was generally 'straightforward' or 'very straightforward' to manage, while a fifth found it difficult. The average time spent preparing for the phonics screening check was approximately three hours and it took an average of 12.5 hours to administer, although it took more time in larger schools.
- Schools reported that the Department administration guidance was clear and straightforward. The vast majority (89 per cent) of pilot schools felt the guidance was 'useful' or 'very useful'. The practice marking workshop at the regional training events was felt to be particularly useful. Almost two-thirds of case study schools suggested that something similar be provided in audio/visual format for national roll-out.

The content, suitability and impact of the phonics screening check

- The vast majority of schools (90 per cent or more) said that the content of the phonics screening check was suitable for children working at the expected standard of phonics across most aspects of the phonics screening check's content. Lower proportions of pilot schools felt the phonics screening check was suitable in relation to the number of words (83 per cent), the type of vocabulary used in real words (80 per cent) and the use of pseudo-words (74 per cent).

- Most children felt that the use of pseudo-words on the phonics screening check was a 'fun', novel aspect. However the majority (60 per cent) of pilot schools said that pseudo-words caused confusion for some children, while 12 per cent said they caused confusion for most children.
- Three quarters of schools felt that the phonics screening check assessed phonic decoding ability accurately overall for their children⁷. However, less than half of teachers were confident in the accuracy of the phonics screening check's assessment for lower ability children and children with special educational needs (SEN), English as an additional language (EAL) and language difficulties.
- The phonics screening check helped almost half (43 per cent) of pilot schools to identify children with phonic decoding issues that they were not already aware of, although over half felt it did not help in this way. Case study schools stated that they intended to use the outcomes to plan phonics delivery, support children and inform teaching. Schools supported the Department's position that results should not be made publicly available.
- Pilot schools would like detailed results of the phonics screening check. Almost all (97 per cent) schools would like child-level results, nine tenths would like benchmarking data, and 88 per cent would like commentary on national-level results.
- The experience of the phonics screening check was perceived to be positive for most children. Sixty-two per cent of pilot schools felt the experience had been positive for all children, and just under a third (31 per cent) said it was neither positive nor negative. Between 23 per cent and 29 per cent of surveyed schools felt the experience was negative for children with speech and language difficulties, other SEN and weak phonics skills.

Communication relating to the phonics screening check

- Most schools communicated the purpose of the phonics screening check to children in a low key way, commonly as a game or a one-to-one reading session. Most had not informed parents/carers of their child's involvement in the phonics screening check. Parents/carers responding to the parent/carer survey (in 17 schools) most frequently wanted to receive information on their child's performance on the phonics screening check (99 per cent), how the school intends to respond to their child's performance (97 per cent) and information about what they could do to support their child's phonic ability (96 per cent). Most case study schools wished to report findings to parents/carers themselves, in a form that could enable them to support their child's learning, and in a sensitive, appropriate way.

3.6 Pilot logistics

An existing contract with Publishing Delivery Service (PDS) was used for printing, distribution and scanning throughout the pilot.

Decisions

A log of decisions in each area is set out in table 16 below.

⁷ It should be noted that the standard on the phonics screening check had not been set and no results had been provided to schools when they were asked to complete the evaluation.

Table 16 – Logistics decision log

Area	Decision	Rationale
Mark sheets	Use optical mark read (OMR) sheets, completed by teacher	<p>Item level data required for psychometric analysis, in order to develop tests for future years.</p> <p>Sheets need to be completed by teacher listening to child, and OMR is the simplest way of capturing this information.</p>
No of copies of mark sheets	16,000	<p>Allows for 10,000 pre-identified children to sit the phonics check, plus five spare sheets per school. Mark sheets also required for children participating in the validity and reliability studies.</p>
Check materials	One double sided practice sheet, common to all, along with ten-page single sided gloss laminated and spiral bound check materials specific to each form.	<p>All children would be undertaking the same practice words, so separate sheets not needed. For children unable to access the test beyond this, they are also unaware that they haven't completed the main phonics screening check.</p> <p>Materials need to be long lasting, and ideally in the same form this year as future years. Both gloss and matt laminate dummy copies were examined, and although the gloss had some glare, the matt was more abrasive and more likely to mark.</p> <p>As there are ten sheets to the phonics screening check, it was important to have a method of keeping them in the right order. Spiral bound permits sheets to be kept together, and turned over so they are more like a book.</p>
No. of copies of check materials	80 copies each of the 18 forms.	<p>Each school will sit five different forms of the phonics screening check, and so need copies of each set of materials.</p> <p>Additional copies are also needed for work with SEN schools, and copies for the office.</p>
Distribution	All materials to be double enveloped	<p>Delivery and collection needs to be guaranteed, with the packages contained child level information.</p> <p>Due to the impact level of the child data, double enveloping is required. Materials for returning completed mark sheets and check materials in the same manner will be provided to schools.</p>
Design	PDS to design mark sheets	<p>Mark sheets need to be compatible with PDS scanners and correct format is needed.</p>

Area	Decision	Rationale
Data	Requirements to be details by the Department	All data requirements to be provided in the Assessment Delivery Service Requirement (ADSR). See Annex F.

Deliveries and collections

A total of 301 packages were delivered to 299 schools during the process (packs for two schools had to be delivered twice). A number of issues were identified during this process, some of which caused additional work for schools. However, all schools received materials in time to undertake the phonics screening check.

Although all 301 packages were collected from schools during the process, there were more concerns raised by schools about the difficulties faced in this process. Many schools telephoned the phonics helpline for support and some schools added additional comments to their evaluation form. In addition, the mark sheets from one school were lost.

Scanning and data capture

There were a number of issues identified during scanning. Some of these related to incorrect completion of mark sheets by schools and others through inaccuracies during the scanning and data collection process. Following a number of quality assurance processes, which identified further issues, all mark sheets were scanned twice and all anomalies checked by hand.

Despite these issues, there were sufficient data that had cleared all quality assurance checks in order to carry out the analysis required to demonstrate validity and reliability.

4. Standard setting

As a new assessment, it was necessary to carry out a standard setting exercise on the phonics screening check. The standard setting procedure selected by psychometricians at QCDA was the bookmark procedure. Full details of the procedure can be found in Cizek and Bunch (2009 http://www.sagepub.com/upm-data/13067_Chapter10.pdf). This standard setting procedure was selected for some of the reasons provided in this chapter, namely that:

- 'from the perspective of those who will be asked to make judgments via this method, it presents a relatively simple task to participants, and one with which, at a conceptual level, they may already be familiar';
- 'in addition to being relatively easy for participants, the Bookmark method is also comparatively easy for those who must implement the procedure'; and
- 'from a psychometric perspective, the method has certain advantages because of its basis in item response theory (IRT) analyses, and because of the fidelity of the method to the test construction techniques that spawned the assessment'.

This procedure is widely used internationally to set standards and there has been some experience of using the bookmark procedure in England in National Curriculum tests. It is true to say that the use of the bookmark procedure was discontinued in National Curriculum tests because of concerns over its fitness for purpose. This concern stemmed from the use of the procedure in order to attempt to maintain standards and not to set standards, a purpose for which the procedure is not designed. Therefore, QCDA was content that this method was appropriate for setting the standards on the phonics screening check.

Following the first standard setting exercise, and because of the importance of ensuring an appropriate standard was set, it was decided to carry out a validation exercise by repeating the exercise with a second independent group of teachers. The process followed on both days was identical and reflected the standard procedure described by Cizek and Bunch (2009).

4.1 Standard setting participants

There were two sets of individuals involved in the standard setting process:

- Phonics experts who had been involved in the development of the Assessment framework – they were involved in the development of the performance descriptor and attended the standard setting meeting as observers
- Teachers involved in the administration of the phonics screening check pilot – they were participants in the standard setting meetings

The teacher participants were selected from those schools involved in the pre-trial and the main pilot. For both standard setting meetings, teachers were asked to express an interest in taking part in the exercise and as a result 25 teachers took part in the process on the first day and 26 on the second.

Information was collected on each participant to ensure a range of experience was represented. Tables 17 and 18 indicate the range of experience of participants on each day.

Table 17 – Standard setting participants experience day 1

	Minimum	Maximum	Average (Mean)
No. of years teaching	1 year	32 years	10.3 years
No. of years teaching phonics	1 year	32 years	7.1 years
No. of years teaching year 1	0 years	15 years	4.5 years
No. of children to whom check was administered	4 children	90 children	43.0 children

Table 18 – Standard setting participants experience day 2

	Minimum	Maximum	Average (Mean)
No. of years teaching	1 year	30 years	10.7 years
No. of years teaching phonics	1 year	30 years	8.5 years
No. of years teaching year 1	0 years	20 years	4.0 years
No. of children to whom check was administered	0 children	90 children	46.6 children

Tables 19 and 20 provide details on the group characteristic profiles for which national figures are available.

Table 19 – Standard setting participants characteristics day 1

Gender	23	Female (92%)
	2	Male (8%)
Age	6	Under 30 (24%)
	5	30-39 (20%)
	12	40-49 (48%)
	1	50+ (4%)
	1	Prefer not to say (4%)
Ethnicity	23	White British (92%)
	1	White Irish (4%)
	1	White other (4%)

Table 20 – Standard setting participants characteristics day 2

Gender	23	Female (88%)
	3	Male (11%)
Age	3	Under 30 (11%)
	9	30-39 (35%)
	10	40-49 (38%)
	4	50+ (15%)
	0	Prefer not to say (0%)
Ethnicity	26	White British (100%)
	0	White Irish (0%)
	0	White other (0%)

Although the group experiences were wide ranging and relatively representative of the teaching profession, with a slight over sample of experienced teachers given the nature of the exercise, the profile of the group characteristics shows that they came from a relatively homogenous background. However, the latest figures on the school workforce released by the Department (November 2010, provisional) indicates that the number of males in the group is in line with the

number of males in the primary workforce (11 per cent of primary teachers are male) and given the size of the group, the ethnic background is not unexpected from a random sample of teachers (94.7 per cent of primary teachers are white).

4.2 Performance level label and description

The first task in any standard setting procedure is to define the performance level label and description. For the phonics screening check, the performance level labels were agreed as:

- Met the expected phonic decoding standard for a child at the end of Year 1.
- Not met the expected phonic decoding standard for a child at the end of Year 1.

Given the nature of this assessment, therefore, it is only necessary to define the performance level description for the first of these performance level labels.

The performance level descriptor was drafted at a meeting with the phonics experts who had been involved in the creation of the assessment framework. This descriptor was shared with teachers involved in the standard setting process as part of their preparation activity. Participants were asked to undertake a number of activities to determine whether they believed the descriptor was appropriate.

The descriptor states that:

Children who have achieved the expected standard at the end of Year 1 will have experience of decoding all of the types of words that appear in the Year 1 phonics screening check. They will know the grapheme-phoneme correspondences and be able to blend phonemes in words with the orthographical structures that have been included in the phonics screening check. However, children at the minimum expected standard will **not** necessarily score full marks.

In particular this means that in the phonics screening check, a child working at the minimum expected standard should be able to decode:

- **all** items with simple structures containing single letters and consonant digraphs;
- **most** items containing frequent and consistent vowel digraphs;
 - frequent means that the vowel digraph appears often in words read by children in Year 1
 - consistent means the digraph has a single or predominant phoneme correspondence
- **all** items containing a single two-consonant string with other single letters (i.e. CCVC or CVCC);
- **most** items containing two two-consonant strings and a vowel (i.e. CCVCC);
- **some** items containing less frequent and less consistent vowel digraphs, including split digraphs;
- **some** items containing a single three-consonant string; and
- **some** items containing two syllables.

It should be noted that items containing a number of the different features listed above will become more difficult. It will become less likely that a child working at the minimum expected standard will be able to decode such items appropriately. For example, a child will be less likely to decode an item containing both a consonant string and a less frequent vowel digraph, than an item with a consonant string but a frequent, consistent vowel digraph.

In order to validate the categorisation of words in this way and to examine whether the performance descriptor was sufficient to explain the difficulty of the items in the phonics screening check an additional piece of work was carried out using regression analysis.

In order to do this, a meeting was held with the phonics experts to determine the factors which could explain the difference in item difficulty, as determined by item response theory (see section 5.2 below). The baseline model for the regression analysis used the features listed in the performance descriptor to categorise all 360 words in the pilot. The adjusted R-squared value was 0.696, which means that the categories used to construct the performance descriptor explain almost 70 per cent of the variation in item difficulty.

Other potential factors were then added to the model, one by one, starting with the factors that the phonics experts thought most important in determining item difficulty. The mean frequency of neighbours (words that are formed by changing only one letter of the item, e.g. ban is a neighbour of bad), the number of neighbours, bigram frequency (the average frequency of the bigrams (two consecutive letters) in each item) and complexity (number of letters in an item minus the number of phonemes) for each word in the trial was calculated using the CPWD and N-Watch (www.pc.rhul.ac.uk/staff/c.davis/Utilities). The mean frequency of neighbours was not significant when added to the model. Adding the number of neighbours increased the adjusted R-squared value to 0.729. Bigram frequency was also not significant. Adding complexity to the model, along with the number of neighbours increased adjusted R-squared further to 0.744. The baseline model was also run separately for real words and pseudo-words, to account for any potential differences between them. The adjusted R-squared values were 0.695 and 0.712 respectively. The results from the regression analysis give us confidence that the performance descriptor is adequate. It describes aspects of words that can explain the varying item difficulty of those words and does so for both real words and pseudo-words.

As preparation for the standard setting meeting, teachers were asked to undertake a number of activities to support their understanding of the performance descriptor. These activities included:

- reviewing their teaching of phonics at the end of Year 1 to determine whether this was in line with the expectations in the descriptor;
- reflecting on when particular elements of phonics had been introduced to children who had reached the expected standard by the end of Year 1; and
- producing their own performance descriptor to demonstrate their expectations for children at other stages throughout key stage 1

In their preparation activities, 78 per cent of participants on the first day and 84 per cent on the second day indicated that all the features of the descriptor had been covered in their teaching with children by the end of Year 1. Those that indicated that they had not covered all features specified a small number of graphemes or words structures that they had not covered. 70 per cent of participants on the first day and 84 per cent on the second day agreed with the qualifiers related to how often children working at the appropriate standard would be able to decode words with particular features. Of those who did not agree, some felt that children would be able to respond correctly to more words and other felt it should be fewer. At both meetings, there was some discussion of what was meant by words such as 'most' and 'some'. At the first

standard setting meeting, a high level agreement was reached indicating that 'some' would mean up to 50 per cent of the time children would be able to decode the words correctly and 'most' meant that this would happen more than 50 per cent of the time. However, at the second meeting, although these figures were acknowledged, it was felt that 'most' and 'some' would mean different things for different words and that for words containing three-consonant strings, but not other feature, 'some' could mean over 50 per cent.

At both standard setting meetings, there was unanimous agreement that the descriptor was an appropriate expectation for children at the end of Year 1 and it was therefore used as the basis for all discussions at the standard setting meeting.

4.3 Data analysis

A two-parameter item response theory (IRT) model using a response probability of two-thirds (see Cizek & Bunch, 2009, p167) was used to assemble the ordered item booklets. In addition, a one-parameter logistic model was used to confirm the item order. Items from the eighteen forms were included in the ordered item booklet to ensure that standard setting participants could place their bookmarks with sufficient precision. As might be expected when comparing results from one and two parameter models, some item order differed substantially. To avoid confusion and because there were sufficient numbers of items to do so, these items were removed from the booklet with no impact on the precision of the process.

4.4 Standard setting meeting

The first standard setting meeting took place on 21 September 2011 and the second on 2 December 2011. An agenda for both days is provided in Annex G which details the activities that took place in the standard setting meeting. The agenda for the standard setting meeting provides less time for activities than that suggested by Cizek and Bunch (2009) because the nature of the phonics screening check is less complex than the sorts of tests for which the suggested agenda was developed. There was sufficient time at the meeting for all teachers to complete the exercises.

In addition to undertaking the bookmark procedure, participants were asked to complete evaluation forms in order to provide a level of confidence in the final outcomes.

Both meetings started with a review of the preparation activities. The purpose of this session was to ensure that all participants were comfortable with the performance level descriptor. Although there was some discussion, particularly at the first meeting, of the descriptor being 'aspirational' for some schools, the participants agreed that the standard was achievable for children by the end of Year 1. In the baseline evaluation for both meetings, all participants agreed with the statement 'I am comfortable with the performance level descriptor', with 14 out of 25 agreeing strongly on the first day and 6 out of 26 on the second.

The next stage of the meeting involved an explanation of the bookmarking procedure. Participants were given multiple opportunities to ask questions to ensure that they had fully understood what was expected of them. Again, the baseline evaluation indicated that all participants were comfortable that they understood the procedure, with 17 out of 25 agreeing strongly on the first day and 12 out of 26 on the second.

At the end of the introductory sessions, participants were asked in the baseline evaluation whether 'the discussion of the standard setting procedure was sufficient to allow me to feel confident that my colleagues and I will be able to set a pass mark'. Again, all participants agreed with this statement, with ten out of 25 agreeing strongly on the first day and ten out of 26 on the second.

Round 1

In round 1, participants were asked to work independently to identify the last item in the ordered item booklet where children who are **just** working at the appropriate standard would have a two-thirds chance of success on that item.

On the first day, the decisions for each participant ranged from item 177 to item 340, with an average of 273. However, the majority of decisions were clustered in two regions – between 200 and 250 and between 300 and 340. Although participants had been asked to work independently, there was often agreement on each table.

In the group discussion that followed, those participants who had indicated in the higher region explained that this was because they had wanted to see evidence of children achieving on all of the different elements listed in the performance descriptor. Those in the lower region had felt that there were items here that they felt children working just at the standard would struggle with, although there were some words following that point which children might be able to decode properly.

On the second day the decisions for each participant ranged from item 188 to item 320, with an average of 263, with the majority of decisions spread over the range 227 to 320. The range of responses on each table indicated that the group had worked independently.

In the group discussion that followed, participants commented on the difficulty of making a decision because the order of the items in the booklet, although based on evidence from the pilot, was contrary to their expectations of difficulty.

At this point, participants were organised onto different tables to ensure a range of opinions were considered on each table.

Round 2

In round 2, participants were asked to work in groups of five or six to agree a single item where the bookmark should be placed.

On the first day in this round the bookmarks ranged from item 244 to item 310, with three groups between item 296 and item 310, one group at item 276 and one group at item 244.

At this point, some participants raised concerns that they may be setting expectations too high. A further discussion was then had to determine what we meant by a child who is just at the expected standard. It was agreed that, although they would demonstrate the features indicated in the descriptor, these would not be secure. Therefore, where the qualifiers 'some' or 'most' were used, these should be interpreted on the low side. Also, it was agreed that not all features included in the descriptor need to have occurred before the bookmark could be placed in the booklet.

There followed a further discussion about the performance descriptor to confirm that participants were still content that it reflected the expected standard for children at the end of Year 1. Participants confirmed that this was the case but agreed that, although this might be the

standard expected at the end of Year 1, children who were just at the standard might not necessarily succeed with all features in the phonics screening check due to a number of factors.

Participants were organised onto different tables again and allowed to review their decision from round 2 in an extra round. In this additional round on the first day, participants worked in their new groups of 5 to reach consensus, which ranged from 243 to 252.

On the second day in this round the bookmarks ranged from item 237 to item 285, with one group unable to reach consensus.

In the discussion that followed, the group focused on two sets of words that were influencing their decisions. The first were a set around 240, which participants felt were difficult and had led several to place their bookmarks in this region and the second were a set around 260 and 270 which participants felt were relatively easy and had led some to ensure that the bookmark was placed after this point. Over the course of the discussion, the consensus moved to the lower end of the range identified in round 2, between 240 and 250.

It was at this point that participants were shown impact data. This data indicated the threshold on each of the forms related to each of the items selected for the bookmark and the percentage of children who would achieve the standard given that threshold.

Using the consensus from the first day, with a bookmark of 243, the threshold marks on the 18 forms ranged from 32 to 35 out of 40, with 29.6 per cent of children who took the phonics screening check in the pilot achieving the standard. With a bookmark of 252, the thresholds on the 18 forms ranged from 33 to 35 out of 40, with 28.2 per cent of children who took the phonics screening check in the pilot achieving the standard.

Using the consensus from the second day, with a bookmark of 240, the threshold marks on the 18 forms ranged from 32 to 35 out of 40, with 29.6 per cent of children who took the phonics screening check in the pilot achieving the standard. With a bookmark of 250, the thresholds on the 18 forms ranged from 33 to 35 out of 40, with 28.2 per cent of children who took the phonics screening check in the pilot achieving the standard.

Participants were reminded that, even though they had now seen the impact data, this had not amended the agreed standard. Indeed, on both days participants confirmed that they still believed that the standard was appropriate. However, there was significant concern that so few children were achieving the standard. Although some indicated that this may be because there was insufficient focus on teaching phonics in some schools, a small number of participants on the first day, and one on the second day felt that the translation of the standard into a bookmark had expected too much of children.

Participants were asked to focus again on children who had achieved a sufficient standard by the end of Year 1 such that teachers were confident that they would go on to become confident readers without specialist intervention.

Participants were reorganised into groups for the final round of the process.

Round 3

On the first day, following small group discussions, each group was asked to explain their rationale for their groups' decision.

One group that stayed at item 244 said that they hadn't wanted to move just because of the impact data they had seen. They felt that the group had put a lot of work in earlier in the day

and made careful judgements based on the expected standard and they didn't want to negate that work.

The group that moved to item 237 felt that this was true but had still felt able to move their bookmark down slightly on reflection.

One teacher in the two groups that moved to item 190 said that she had looked at the performance descriptor and thought about the children that would meet that and her feeling was that that they would achieve a score of about 28 out of 40. Others said that they felt that after item 177 more complex word structures appeared and it was at this point that the performance descriptor was met. They felt that their expectations earlier in the day may have been too high (when looking at the booklet).

The group which moved to item 145 felt that this was the earliest point at which the performance descriptor could be met. Items in the descriptor that stated that children should achieve on all such words appeared before this point and after that, more complex words appeared.

In the final discussion, some commented that they should not be swayed by other pressures (media, pressures on teachers etc) to move away from the agreed expectation (in line with Phase 5 of letters and sounds) otherwise they would be saying this (phase 5/the agreed expected standard) is not actually what should be expected in Year 1. This made some feel that they should stick in the region they had all agreed on after round 2a.

After some discussion members of the group were asked whether they still agreed with item 244 or would want the bookmark set lower. The group was split 50/50. The group was instructed to go back to the booklets and look through the words with a view to coming to a consensus. After looking at the booklet the group converged on items 219 to 223, with a slight majority preference for 223.

On the second day, the groups were also asked to provide a rationale for their decisions. Again, the main issue seemed to be the items between 240 and 245 and whether these should be before or after the bookmark. There was some discussion of ensuring that appropriately high expectations were set, although this was countered by the argument that these should also be achievable.

Consensus was reached more easily on the second day with a majority vote to select item 240. Those who had voted for 250 were content to go with the majority decision.

4.5 Final evaluation

Having completed the exercise, participants were asked to complete a final evaluation. In the evaluation, participants were asked a variety of questions to determine how they felt the meeting had gone. On both days, all participants were mostly positive about the experience, indicating that they understood the task, had been given clear instructions and that the discussions had been useful or very useful.

The final question in the evaluation asked them to select one of four responses to the question 'How comfortable are you with your final recommended pass mark?' – very uncomfortable, somewhat uncomfortable, fairly comfortable and very comfortable.

On the first day, three participants said they were somewhat uncomfortable; indicating on their form that their reservations related to the number of children achieving the standard rather than

they felt the standard was incorrectly placed. Fifteen participants indicated that they were fairly comfortable⁸ and five that they were very comfortable. Two participants did not answer the question.

On the second day, one participant said they were somewhat uncomfortable; indicating on their form that they felt the standard was too high considering the age of the children. Fourteen participants said that they were fairly comfortable and 11 that they were very comfortable.

Overall, the outcomes from the evaluation and the fact that both groups independently reached similar outcomes indicate that there can be a significant amount of confidence in the outcomes of the standard setting process.

Since the outcomes from the two groups were within one standard error of each other, the Department is confident that the results of standard setting have been validated. Therefore, to calculate the final standard for the phonics screening check, an average of the outcomes from the two groups was taken.

4.6 Final outcomes

Selecting the averaged outcome of the two days as defined by the average item from the two days, item 231, as the final bookmark, the thresholds on the 18 forms ranges from 31 to 34 out of 40 (see section 5.2). This means that in the pilot, 31.8 per cent of children met the expected phonic decoding standard for a child at the end of Year 1.

These outcomes were published by the Department at www.education.gov.uk/inthenews/inthenews/a00200672/a-third-of-children-reach-expected-level-in-pilot-of-phonics-check in December 2011.

It should be noted here that there are a number of factors that mean that we cannot be certain that these outcomes will be reflective of national results following roll-out.

- Pilot schools were only informed of their participation in the pilot by the start of the term in which the phonics screening check took place. In roll-out, schools have been made aware at the start of the academic year.
- Pilot schools were only made aware of the content of the phonics screening check at the training meetings, which took place up to two weeks before the administration week. In roll-out, this information will be provided to schools earlier in the academic year.
- Pilot schools were not made aware of the expected standard before they administered the phonics screening check. The expected standard has been included in the assessment framework, which was published in January 2012 and further communications will be provided to schools during the Spring and Summer terms.

These factors are likely to impact on national results during roll-out. In addition, the increased focus on the teaching of phonics, through the availability of matched funding for approved phonics products is likely to impact school results in the future.

⁸ One of these participants contacted the Department the day after the event to state that on reflection they would like to change their response to very uncomfortable. They were concerned that although they understood the standard and felt it should be the aspiration and that if too many pupils failed to achieve the standard the phonics screening check would not be taken seriously by schools and would be demoralising for pupils.

5. Live data analysis

The live data from the pilot was analysed using classical test theory (CTT) and item response theory (IRT). CTT focuses on the performance of a whole test or assessment, in this case the different forms of the phonics screening check. IRT considers the performance of individual items within a test or assessment, in this case the words used in the phonics screening check. Evidence from CTT is dependent on the sample of children taking the test or assessment, therefore care must be taken when interpreting results. Analysis using IRT attempts to account for the different abilities of children taking different forms of the assessment in order that direct comparisons can be made. All analysis in this section is based on the representative sample of schools involved in the pilot.

5.1 Classical Test Theory

Table 21 below shows some general classical statistics for each form of the phonics screening check.

Table 21 – Classical analysis outputs

	No. Children	Mean score	Standard Deviation	Cronbach's Alpha	Standard Error of Measurement
Form 1	449	24.6	10.9	0.957	2.3
Form 2	513	24.2	10.7	0.949	2.4
Form 3	469	24.3	11.5	0.960	2.3
Form 4	510	23.8	11.2	0.957	2.3
Form 5	431	23.9	10.8	0.953	2.3
Form 6	466	25.1	11.3	0.960	2.3
Form 7	438	27.1	10.1	0.951	2.2
Form 8	454	25.2	10.6	0.951	2.3
Form 9	505	26.4	10.9	0.958	2.2
Form 10	450	25.3	11.0	0.955	2.3
Form 11	460	27.3	10.0	0.948	2.3
Form 12	470	26.2	10.9	0.957	2.3
Form 13	469	25.0	11.2	0.959	2.3
Form 14	481	24.6	11.6	0.962	2.3
Form 15	447	24.6	11.6	0.961	2.3
Form 16	431	23.7	11.4	0.960	2.3
Form 17	548	24.5	11.2	0.959	2.3
Form 18	460	24.2	10.8	0.953	2.4

As can be seen in the table, the mean score on each form ranged from 23.7 to 27.3 marks. This could indicate a difference in difficulty between the different forms of 3.6 marks. However, although children were randomly assigned to forms it is unlikely that the profile of abilities of children taking each form of the phonics screening check is identical and therefore, some of the difference is likely to be attributed to the children rather than the difficulty of different versions of the phonics screening check. Further discussion of the relative difficulties of the different forms will be included in the IRT section of the report where such factors can be accounted for. The standard deviations on all forms were mostly within 1 mark of each other.

Cronbach's alpha is a measure of the internal consistency of a test or assessment, with a maximum value of 1. High values of Cronbach's alpha, such as those obtained on every form of the phonics screening check, indicate that performances on different items within the phonics screening check are highly correlated – i.e. that a child who performs well on one item is likely to perform well on another. Internal consistency is required to ensure that interpretations of total scores on the phonics screening check are valid – i.e. that a child with a high score has consistently performed well on the construct being measured by the assessment, in this case the ability to decode using phonics. Values of Cronbach's alpha of more than 0.9 are generally considered excellent. However, it should be borne in mind that the nature of items in the phonics screening check, single words to be read by a child, are likely to lead to high values of alpha because of their similarity.

Another indication of the reliability of the phonics screening check is the standard error of measurement. The standard error of measurement is an estimate that allows the user to determine a confidence interval around an observed score. In the case of the phonics screening check the average standard error of measurement across the eighteen forms is 2.3. This means that we can be 95 per cent confident that a child's "true score" is within 5 marks of their observed score. This standard error of measurement is calculated with the standard deviation and the Cronbach's alpha and as a result is a function of the form that has been taken and is considered the same across the score range.

A similar analysis for each section is shown in table 22.

Table 22 – Classical analysis by section

	Section 1			Section 2		
	Mean score	Standard Deviation	Cronbach's Alpha	Mean score	Standard Deviation	Cronbach's Alpha
Form 1	15.3	4.8	0.903	9.3	6.6	0.942
Form 2	14.5	5.0	0.897	9.8	6.1	0.918
Form 3	14.0	5.4	0.911	10.3	6.6	0.940
Form 4	14.6	5.2	0.910	9.2	6.5	0.936
Form 5	14.7	4.9	0.895	9.2	6.6	0.937
Form 6	14.5	5.3	0.917	10.6	6.5	0.935
Form 7	16.0	4.4	0.888	11.1	6.2	0.934
Form 8	15.2	4.8	0.893	10.0	6.4	0.933
Form 9	15.4	5.1	0.913	11.1	6.3	0.937
Form 10	15.1	5.1	0.905	10.3	6.6	0.937
Form 11	16.1	4.5	0.892	11.2	6.1	0.927
Form 12	15.2	5.1	0.909	11.1	6.3	0.935
Form 13	15.1	5.2	0.915	10.0	6.6	0.940
Form 14	14.6	5.3	0.918	10.0	6.7	0.940
Form 15	14.6	5.4	0.917	10.0	6.9	0.947
Form 16	13.7	5.2	0.907	10.0	6.6	0.944
Form 17	14.9	5.2	0.915	9.6	6.6	0.940
Form 18	14.7	5.1	0.902	9.4	6.3	0.931

As is to be expected, given the specification, children performed better on section 1 than on section 2. In all forms, the Cronbach's alpha for section 1 is lower than section 2. However, since large numbers of children are scoring high marks on section 1, there is less opportunity for the section to discriminate between higher and lower performers, hence lower values of Cronbach's alpha.

Within each form, there were differences between the performance of boys and girls. Table 23 shows the mean scores for boys and girls on each form of the phonics screening check.

Table 23 – Classical analysis by gender

	Males			Females		
	No. children	Mean score	Standard Deviation	No. children	Mean score	Standard Deviation
Form 1	201	23.2	11.6	248	25.7	10.1
Form 2	240	22.5	11.0	273	25.8	10.1
Form 3	244	23.4	11.7	225	25.3	11.2
Form 4	261	23.0	11.9	249	24.6	10.3
Form 5	212	22.8	11.6	219	25.0	9.9
Form 6	236	24.6	12.1	228	25.7	10.6
Form 7	208	26.3	10.1	230	27.8	10.1
Form 8	230	25.1	10.6	224	25.3	10.5
Form 9	259	25.0	12.0	246	28.0	9.5
Form 10	227	24.6	11.4	223	26.0	10.6
Form 11	239	26.3	10.3	221	28.3	9.5
Form 12	243	24.9	11.6	226	27.7	10.0
Form 13	231	24.7	11.4	238	25.4	11.1
Form 14	259	24.5	11.7	222	24.7	11.6
Form 15	205	23.6	11.8	242	25.4	11.4
Form 16	212	23.0	12.0	218	24.3	10.8
Form 17	293	23.6	11.4	255	25.5	10.9
Form 18	229	23.0	11.2	231	25.3	10.3

As can be seen in the table, girls outperformed boys on all versions of the phonics screening check by between 0.2 and 3.2 marks. Again, care should be taken when interpreting these results since these figures are related to the particular samples taking each form.

This report will not provide detailed analysis by item because of the confidential nature of the items, which will be used in the roll-out of the phonics screening check. However, some overall analysis of items will be reported here.

Table 24 shows the average classical facilities and discriminations for items on different pages of the phonics screening check.

For one-mark items, such as those in the phonics screening check, facilities are equivalent to the percentage of children who answered each item correctly. Discrimination relates to the ability of an item to differentiate between high and low performers, specifically, the relationship between child performance on an item and their total score. Items with high discrimination will help ensure that children are appropriately classified as having met or not met the standard. Item with low discrimination will lead to increased misclassification. It should be noted that the calculated discriminations are corrected point biserial correlations, as such values greater than 0.30 are acceptable.

Table 24 – Average facilities and discriminations from classical analysis

	Average Facility	Average Discrimination
Page 1	84.36	0.42
Page 2	71.72	0.54
Page 3	67.89	0.56
Page 4	76.06	0.56
Page 5	72.19	0.59
Page 6	43.78	0.57
Page 7	51.64	0.59
Page 8	48.77	0.63
Page 9	55.68	0.64
Page 10	52.84	0.68

As expected, the average facilities on pages in section 1 (pages 1-5) are higher than those in section 2 (pages 6-10). In addition, the facilities for comparable real and pseudo-words (i.e. with the same orthographical structures and graphemes) indicate that the average facilities for pseudo-words are around 4-5 per cent lower than for real words (comparing pages 2 and 4, pages 3 and 5, pages 6 and 8 and pages 7 and 9). As ever with classical analysis, these values are sample dependent and therefore although indicative, may not be directly comparable.

The values of discrimination for all items are generally good or very good. Although acceptable, the discriminations for page 1 are lower. This is to be expected given that the facilities for these items are so high, leaving little opportunity to discriminate between high and low performers.

5.2 Item response theory

As indicated above, IRT considers the performance of individual items in the phonics screening check but including all of the items across all 18 forms in one analysis, thereby creating an item bank of phonics screening check words. Using the software package Mplus v5.2, a two-parameter model estimated item difficulty and discrimination as well as child ability.

Other IRT models are available, however, the two-parameter model is considered to be the most suitable in this context as estimating both difficulty and discrimination is meaningful. In examining the two-parameter model, it is clear that estimating discrimination is the most appropriate route because of the range of values obtained. This makes the one-parameter model less appropriate. Estimating a lower asymptote parameter in a three-parameter model is possible but meaningful interpretation of this parameter in this context is unclear.

There are two main assumptions in item response theory: unidimensionality and local independence. The assumption of unidimensionality suggests a single underlying construct in the data that we call ability. In the case of the phonics screening check it would be the ability to decode using phonics. The assumption of local independence assumes that the items are not related to each other except through child ability. It is well established that IRT is robust to minor violations of these assumptions; and that it is important to evaluate these assumptions

The assumption of local independence was tested using Yen's Q_3 statistic. For any pair of items the Q_3 statistic is calculated as the correlation between the extent to which children achieve above or below their expected score given their ability on one item and the extent to which they achieve above or below their expected score on the other item. The estimates of ability for each

child and the item parameters derived from the IRT model were used to calculate the expected score on each item for each child. From this, the difference between the expected score and actual score was calculated and the correlations between these differences. For the assumption of local independence to be upheld these correlations should be close to zero. The average Q3 statistic for all 360 items in the trial was -0.02. Although this is less than zero, the degree of violation of local independence is relatively small.

Unidimensionality was tested with confirmatory factor analysis and for all forms was found to be well within expectations of good model fit for a unitary construct. Bentler and Hu (1999) recommend that model fit be considered good if the Tucker Lewis Index (TLI) is not less than 0.95 and the root mean square error of approximation is not more than 0.05. All form exhibited TLI and RMSEA values within these recommendations, the average TLI across 18 forms was 0.98 and the average RMSEA was 0.035.

The evidence presented on the IRT assumptions clearly supports the use of IRT to analyse the phonics data. With respect to item fit, Yen (2006) advises that ‘definitive conclusions about the best way to measure item fit cannot yet be drawn’ and that large sample sizes increase the number of items misfitting. Examining item fit graphically shows that the vast majority of items fit the model.

The Department is therefore confident that the IRT model chosen fits the data and is appropriate for the analysis of the Year 1 phonics screening check data.

The scale on which item response theory operates is different from classical test theory and generally revolves around a mean ability of zero and standard deviation of one. The scale of item difficulty ranges from -2.65 to 1.14. This means that items with a difficulty less than zero are less difficult than items with a difficulty greater than zero. The discrimination scale is a bit more difficult to interpret, but the general principle is, as with classical test theory, the larger the value the better. The scale of discriminations on the phonics screening check ranges from 0.53 to 2.72. Table 25 shows the average difficulty and discrimination from the IRT model for each page of the phonics screening check. It should be noted that these figures have been calculated only to provide a general picture of each page of the phonics screening check and have no technical significance.

Table 25 – Average difficulty and discrimination from IRT analysis

	Average Difficulty	Average Discrimination
Page 1	-1.684	0.811
Page 2	-0.830	1.075
Page 3	-0.629	1.134
Page 4	-0.942	1.243
Page 5	-0.750	1.453
Page 6	0.230	1.349
Page 7	-0.024	1.313
Page 8	0.066	1.812
Page 9	-0.139	1.744
Page 10	-0.034	1.992

Although the values from the IRT model are not as intuitive to interpret as results from classical analysis, it is clear that the findings relating to the differences between real and pseudo-words that followed table 24 are supported by the values from the IRT analysis.

IRT was also used to determine the relative difficulty of the different forms that were piloted. IRT analysis links scores on each version of the phonics screening check through the common items on different forms to predict the scores a child of a given ability would achieve on each different version. It is most important to understand the score that a child working at the threshold of the appropriate standard would achieve on each form, since this cut-score will be used to determine whether children have met the expected standard or not. Table 26 shows the expected score for a child working at the ability threshold determined through the standard setting process. These thresholds have then been rounded down to the nearest integer to ensure that all children who are determined as working above the threshold are classified as meeting the appropriate standard.

The table also includes information on the percentage of children on each form who achieved the standard. It can be seen that the proportions ranged from 29 per cent to 39 per cent, indicating that the abilities of children taking each form were varied. As stated previously, this means that interpretations of the classical analysis need to be treated with care.

Table 26 – Cut-scores on each form

Form	Cut-score linked to threshold ability	Rounded cut-score	Percentage of children achieving standard
Form 1	33.9	33	29%
Form 2	32.0	31	33%
Form 3	33.8	33	32%
Form 4	33.3	33	27%
Form 5	32.8	32	29%
Form 6	35.0	34	29%
Form 7	34.3	34	33%
Form 8	33.6	33	31%
Form 9	33.9	33	38%
Form 10	33.4	33	33%
Form 11	34.1	34	32%
Form 12	33.8	33	37%
Form 13	33.7	33	33%
Form 14	33.4	33	33%
Form 15	34.5	34	30%
Form 16	32.8	32	31%
Form 17	33.9	33	29%
Form 18	32.9	32	31%

5.3 Sub-group analysis

Some analysis has been done on gender, FSM, EAL and SEN, comparing the percentage of children achieving the standard. The outcomes from this analysis are provided in table 27.

Table 27 – Sub-group analysis

		N (children)	Met the standard (%)	SD	Likelihood ratio	Sig
Gender					10.261	0.001
	Male	4258	30	46		
	Female	4241	33	47		
FSM Eligibility					92.871	0.000
	No	6903	34	47		
	Yes	1562	22	41		
SEN					418.153	0.000
	No SEN	7091	36	48		
	SEN	1374	10	30		
EAL					2.252	0.133
	English	6948	31	46		
	Other	1502	33	47		

There was a significant difference in the percentage of children achieving the standard between boys and girls, with more girls achieving the standard than boys.

As might be expected from analysis of other national curriculum tests, children with FSM performed significantly worse than children without FSM provision; and children with SEN also performed significantly worse than children without SEN. Children with EAL did not perform significantly differently in terms of the percentage achieving the standard than those children who are native speakers of English.

5.4 Differential item functioning

Differential item functioning was examined using BILOG-MG which uses a 1 parameter model to estimate item difficulty and estimate group differences, while keeping the discrimination parameter fixed. Group differences in item difficulty were calculated for gender (boy/girl), EAL (EAL/not EAL) and SEN (SEN/not SEN) categories. The sample sizes for these comparisons are given in Table 28 and show that the sample sizes for all groups were sufficient.

Table 28 – Group sample sizes

Boy	Girl	Not EAL	EAL	Not SEN	SEN
4201	4174	6883	1492	7022	1353

The number of items (out of 360) exhibiting differential item functioning is shown in Table 29. There were very few items exhibiting serious differential items functioning. There are no clear substantive explanations for the differential item functioning of these items in the trial, therefore, the items will remain in the item bank and continue to be monitored.

Table 29 – Number of items exhibiting differential item functioning

Significance	Gender	SEN	EAL
0.01%	1	0	3
1%	6	3	6
5%	10	10	15

5.5 Analysis by phonics teaching

As part of the evaluation, schools were asked to provide details of their phonics teaching practices. Due to the nature of the responses, it is not possible to use the outcomes to fully assess the quality of phonics teaching in each pilot school. However, two questions do provide some insight as to the status of phonics within each school.

The first relates to whether schools encourage children to use strategies other than phonics to decode unfamiliar words and the second to how schools organise the teaching of phonics.

Examining the percentage of children who achieved the standard by whether schools always encourage children to use phonics as a strategy to decode unfamiliar phonically regular words or schools encourage children to use a range of cueing systems suggests that schools that always encourage children to use phonics have higher child performance on the phonics screening check than those schools who encourage a range of cueing systems. The figures are shown in table 30. However, it should be noted that the number of children in schools that encourage a range of cueing systems far exceeds the number of children in schools that encourage phonics only (5964 vs 1989) so these results should be interpreted with some caution.

Table 30 – Phonics as primary strategy

	N (children)	Met the standard (%)	SD	Likelihood ratio	Sig
				32.946	0.000
We always encourage children to use phonics as the strategy to decode unfamiliar phonically regular words	1989	37	48		
We encourage children to use a range of cueing systems, such as context or picture cues, as well as phonics	5964	30	46		

Examining the percentage of children who achieved the standard by whether schools always teach phonics in discrete sessions or mostly teach phonics in discrete sessions suggests that schools that always teach phonics in discrete sessions have higher child performance on the phonics screening check than those schools that mostly teach phonics in discrete sessions. The figures are shown in table 31. Again, it should be noted that the number of children in schools that mostly teach phonics in discrete sessions far exceeds the number of children in schools

that always teach phonics in discrete sessions (4870 vs 2295) so these results should be interpreted with some caution.

Table 31 – Phonics teaching

	N (children)	Met the standard (%)	SD	Likelihood ratio	Sig
				30.010	0.000
Always teach phonics in discrete sessions	2295	36	48		
Mostly teach phonics in discrete sessions, sometimes integrate phonics into literacy sessions/other curriculum work	4870	30	46		

5.6 Analysis by term of birth

Examining the percentage of children achieving the standard by term of birth of the children suggests that children born earlier in the academic year perform better than those born later in the academic year. The figures are shown in table 32.

Table 32 – term of birth

	N (children)	Met the standard (%)	SD	Likelihood ratio	Sig
				152.649	0.000
Autumn	2876	39	49		
Spring	2716	31	46		
Summer	2872	24	43		

This outcome was anticipated based on experience from other National Curriculum tests. However, the Government has high expectations for all children and therefore the policy does not take into account age for the purpose of setting standards on the phonics screening check.

6. Check re-check study

One of the concerns raised about introducing a check for children at the end of Year 1 is that children may perform differently on different days depending on a number of factors. This study was designed to determine the extent to which this is true and the likely impact on outcomes for schools.

One hundred schools in the main sample of the pilot were invited to be involved in the study. They were provided with additional information at the training days to enable them to carry out the study, see Annex H for the guide for this study. Schools in the study were asked to administer two different forms of the phonics screening check to each child within their school. All forms were pre-assigned to ensure that phonics screening checks were delivered in a random order. Due to issues relating to the time available to administer the phonics screening check, data was provided from 84 schools and 2730 children.

As each of the forms of the phonics screening check were at different levels of difficulty, it is not possible to undertake a direct comparison of scores achieved by children on the different versions. However, an analysis has been undertaken to determine the proportion of children who were classified in the same way on both versions of the screening check as having met or not met the expected standard.

Table 33 shows the number of children in each category on the check and the re-check.

Table 33 – Outcomes from the check re-check study

		Re-check		
		Not met the standard	Met the standard	Total
Check	Not met the standard	1741	159	1900
	Met the standard	113	717	830
	Total	1854	876	2730

In the study, 90.0 per cent of children achieved the same outcomes regardless of the phonics screening check that they took. Of those that did not achieve the same outcomes, just over half performed better on the re-check and just under half performed better on the check.

Of the 272 children who achieved different outcomes on the check and re-check, 107 were within two marks of the threshold on both the check and the re-check, which is close to the standard error of measurement identified for the phonics screening check. An additional 148 children were within two marks of the threshold on either the check or the re-check.

We cannot tell from this study which classification was 'correct' for the child, however, it does indicate that of the 10.0 per cent of children that were differently classified, the score achieved by 94 per cent of these children on at least one form was within the standard error of measurement at the threshold. This would indicate that the child is likely to be on the very threshold between meeting and not meeting the expected standard. As a result, children achieving a score within the standard error or measurement of the threshold should be considered as borderline children whose classification may be different on a different form of the phonics screening check.

In the pilot, 17 per cent of children achieved scores that were within two marks of the threshold and as a result, from the evidence above, would appear to be the most at risk of being misclassified. This study suggests that the risk of misclassification appears relatively low, however, schools should consider the outcomes of children who score within two marks of the threshold, and the interventions they put in place subsequently, carefully.

7. Inter-rater reliability study

In this study, pairs of teachers scored a number of recordings of pupils taking the phonics screening check independently. The outcomes were then analysed to determine a measure of the reliability of scoring of the phonics screening check.

There were, however, a number of limitations with this study:

- Quality of the recordings – the recordings were made using a dictaphone and as a result the sound quality was not perfect. Although measures were taken to try to remove recordings where the responses were difficult to hear, this was not always possible and therefore differences in scores may be related to the quality of the recording
- Use of audio rather than video recordings – evidence from the face-to-face training sessions showed that audio recordings were not the easiest way to score a child’s responses. On many occasions, teachers felt that they wanted to see a child’s lips move in order to help them determine what the child was trying to say.
- Lack of familiarity with the children – the guidance for the phonics screening check makes it clear that the teacher administering the phonics screening check should be familiar with the child in order that they may better understand their responses, for example if a child has communication difficulties. In the study, neither teacher knew the child whose responses they were scoring.

As a result, outcomes from this study need to be treated with caution.

The first set of analysis conducted, shown in table 34 below, considered the outcomes for each pupil based on the scoring by each of the two teachers.

Table 34 – Outcome agreement from the inter-rater reliability study

		Standard Achieved – Teacher 2		
		Not met the standard	Met the standard	Total
Standard Achieved Teacher 1	Not met the standard	226	14	240
	Met the standard	12	74	86
	Total	238	88	326

This shows that 92 per cent of pupils achieved the same outcome regardless of the teacher who scored their recording. Of the 26 pupils who achieved different outcomes, 92 per cent (24 out of 26) were within one standard error of measurement of the threshold for the score recorded by at least one of the teachers. However, for the two pupils where this was not the case, the two scores were over 20 marks different, implying that there was a significant issue with the recordings. Given the limitation stated above, the percentage agreement is therefore extremely high.

Values of Cohen’s Kappa, a measure of inter-rater agreement, were also calculated for each item in the phonics screening check and a summary is shown in table 35 below.

Table 35 – Cohen’s Kappa calculated by the inter-rater reliability study

	Cohen’s Kappa
Overall average	0.80
Section 1 average	0.77
Section 2 average	0.83
Real words average	0.85
Pseudo-words average	0.75

Although there are no agreed magnitude guidelines for values of Kappa, given that factors other than agreement can influence magnitude, Fleiss (1981), states that values over 0.75 are excellent, values between 0.40 and 0.75 are fair to good, and below values below 0.40 are poor. The values of Kappa calculated for each group of words would therefore be characterised as excellent.

There is a noticeable difference in the values of Kappa between real and pseudo-words. This may be because these are more difficult for teachers to interpret since they are not used to hearing these words. However, in the roll-out, a separate sheet will be provided to all teachers to give greater support in scoring pseudo-words than was available in the pilot. For each pseudo-word, examples of equivalent onset and rimes will be provided. For example, for the pseudo-word ‘vead’, teacher will be advised that this item uses the ‘v’ from ‘vet’ and rhymes with ‘head’ or ‘bead’ (noting that these words are pronounced to rhyme in certain regional pronunciations). This additional guidance, which was requested by teachers in the pilot, should further strengthen marker reliability and ensure an appropriate score for each pupil in the phonics screening check.

8. Children with SEN

8.1 Identifying groups of children

In conjunction with the Department's team for special educational needs (SEN) five major groups of children with SEN who may require particular arrangements to access this type of assessment were identified. The groups were:

- children with hearing impairments
- children with vision impairments
- children with speech, language and communication needs
- children on the autistic spectrum
- children at risk of diagnosis with dyslexia

In addition, the public consultation on the policy asked whether there were others groups of children that should be considered in particular. Respondents suggested that particular care was taken with issues relating to EAL, but there were no other groups of children with SEN were consistently mentioned. Section 5.3 has indicated that children with EAL did not perform significantly differently in terms of child ability estimates than those children who are native speakers of English.

8.2 Meetings with stakeholders

The Department's SEN division identified key stakeholder organisations representing these groups of children. These organisations were then invited to a series of meetings between December 2010 and February 2011.

The first meeting identified the key areas which needed to be considered. Organisations were then asked to complete a template to gather their opinions about these issues more fully. All of the stakeholder groups answered the following questions:

- This is a new type of national assessment, requiring children to read words aloud working one-to-one with an adult. What particular issues might there be related to this phonics screening check?
- What adjustments do we need to consider to the test materials to promote access to the phonics screening check?
- What changes to the administration guidance for teachers, or additional guidance, would be required to promote access to this phonics screening check?
- In what circumstances, if any, should the phonics screening check be disapplied?
- Are there any signs teachers to look for during the phonics screening check which might indicate that a child should be tested for a particular special educational need?
- What short guidance or signposting (to a specialist teacher, local authority, third sector organisation, etc) should be given to teachers who think that a child might have the special educational need which your organisation works with?

At a second meeting in February it was agreed that:

- There should be a presumption that as many children as possible should have the opportunity to access the phonics screening check so that children have a comparable experience. In a small minority of cases, however, disapplication would be necessary. In the pilot, the criteria should be that children who have shown no knowledge of letter-sound correspondences should be disappplied from the check.
- The person administering the phonics screening check in the pilot should be a qualified teacher. For some children with more complex SEN a specialist teacher should be consulted in advance whenever possible. A specialist teacher could also be present during the assessment alongside the class teacher if this was practical.
- A similar range of access arrangements used for other national curriculum assessments would be suitable for the phonics screening check (although some access arrangements, such as the use of a reader would be inappropriate).
- Further evidence should be gathered for the pilot about how children responded to the phonics screening check materials. This should be through feedback from teachers, and the children's results data from the pilot, which can be matched to school census data. It was recognised that the quality of school census data on SEN and the small sample sizes for some child groups will mean that the results data can only be indicative.
- In the pilot particular attention should be paid to children with severe hearing or vision impairment, as these groups of children were most likely to be unable to access the phonics screening check.

8.3 Administration guide for the pilot and modified versions

Schools received face-to-face training on the phonics screening check, and were asked to consider the needs of each child individually. Schools were able make access arrangements at their discretion, including allowing rest breaks, use of modified versions of the phonics screening check materials, Use of British Sign Language (BSL), and rephrasing instructions and use of gestures. The full administration guide can be found in Annex D.

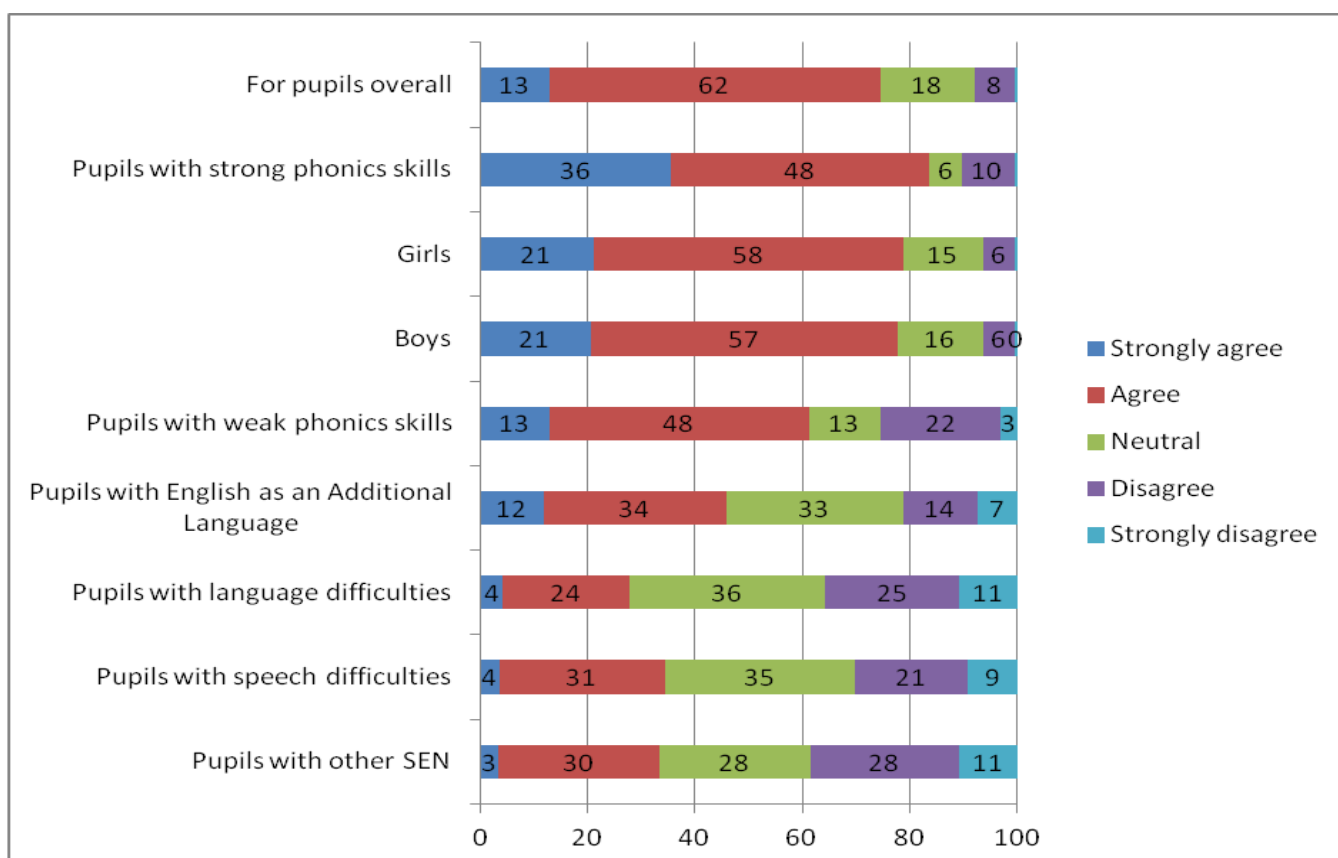
Electronic versions of the phonics screening check were made available to schools on request so they could make school based modifications. This allowed schools to tailor the materials to their children's needs, for example by choosing the right font size or removing the monster pictures for some children on the autistic spectrum. Making electronic adjustments in school was not likely to be burdensome for this assessment because the materials were simpler than for other National Curriculum assessments. The Department also provided grade 1 Braille versions of the phonics screening check on request.

In the pilot, there were requests from five schools for electronic copies of the phonics screening check and a request for a Braille version of the phonics screening check from one school.

8.4 Findings from the independent evaluation of the pilot in relation to children with SEN

As part of the main evaluation of the pilot (see section 3.5) schools were asked to provide their views on the suitability of the phonics screening check for different groups of children. Teachers felt that the phonics screening check was less suitable for children with SEN than other groups.

Figure 1: Responses to the question 'To what extent do you feel the phonics screening check accurately assessed the phonic decoding ability of your school's children?' (%)



Child Experience

71 per cent of teachers who responded to the survey found that children responded positively to the phonics screening check. However, the finding varied when teachers thought about different child groups, see figure 2.

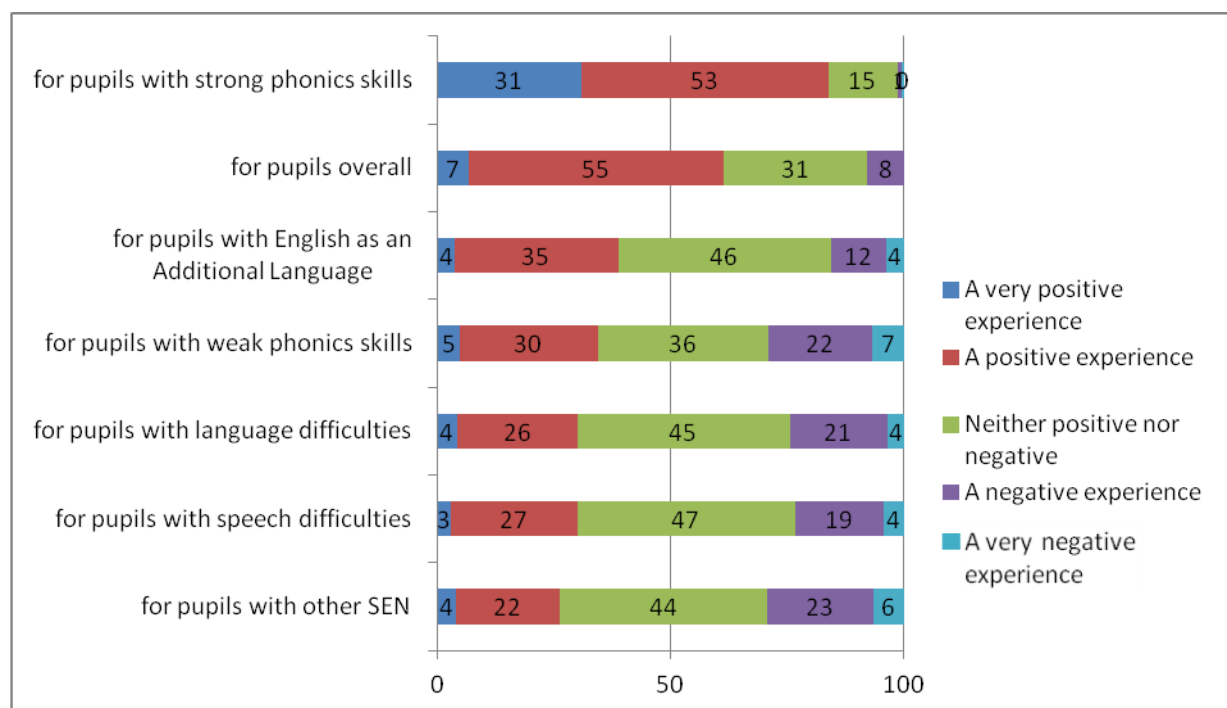
In case study schools, almost all children said that the activity was enjoyable. However, in a minority of case study schools, children became discouraged towards the end of the assessment.

Access arrangements

Access arrangements had been considered for specific children in 15 of the sampled pilot schools for children with SEN and EAL. Only four of the schools had implemented any access arrangements.

Some children were disapplied from the screening check. Schools that disapplied children generally had high levels of EAL and FSM, but not SEN.

Figure 2: School views on the experience of the children when undertaking the phonics screening check



Guidance for teachers

The majority of survey respondents (86 per cent) found the guidance on recognising and scoring responses 'very useful' or 'useful', with teachers with higher proportions of children with SEN finding it most useful.

Some teachers requested more information and guidance on which children should be disapplied, and how to deal with children with SEN, EAL, speech and language difficulties and strong regional accents.

8.5 Meetings with stakeholder organisations and selected pilot schools

A series of focussed meetings on each of the five identified categories of SEN took place. Where possible, this included the relevant main stakeholder organisations that had been involved in the development of the phonics check and teachers from a pilot school who works with children with that type of SEN. These meetings were intended to identify if there were any aspects of the phonics screening check design that meant certain children groups could not perform to their potential. The meetings were also intended to capture any detailed feedback from the pilot which could help to inform changes to the assessment guidance for roll-out.

All participants in the meetings agreed that the design of the phonics screening check did not unfairly disadvantage any child group and that it is important not to lower expectations for children with SEN. Nevertheless, some children with SEN may be on a slower trajectory of progress than their peers – for example children using Braille are less likely to be able to practice decoding at home than their peers. Consequently, some children with SEN may have struggled with some of the words in the phonics screening check, and the length of the

assessment. This is likely to account for some of the less positive findings about the suitability of the assessment in the previous section on the evaluation report.

A small number of teachers raised a concern that children struggling with the phonics screening check lose concentration towards the end of the assessment and so do not attempt to decode the items as carefully as their peers. Rest breaks can help to prevent children from becoming distracted and allow them to focus appropriately on later items in the phonics screening check.

Disengaging children from the phonics screening check

In the pilot, schools were asked to continue with the phonics screening check for all children for a reasonable amount of time where possible to ensure maximum evidence was captured for the analysis of data. It is possible that this contributed to the negative findings reported in relation to children with weak phonics skills in the evaluation.

There was consensus in all the meetings that the rules for discontinuing the phonics screening check for children should be drawn more tightly, and teachers in roll-out should be more strongly encouraged to stop the phonics screening check if a child is struggling. Guidance on disengaging children from the phonics screening check will help to prevent any children from becoming discouraged or having a negative experience. There was also agreement that there still needs to be an element of teacher judgement around when to discontinue the phonics screening check, and so a formal rule to be applied in all cases should not be developed.

Unlike in the pilot, in roll-out teachers will know the standard for the phonics screening check in advance of the assessment, and a decision on discontinuing the phonics screening check might be best linked to the point when it is clear that a child can no longer meet the standard.

Disapplication

Two per cent of children were disappplied from the phonics screening check. In some discussions teachers indicated that they entered children for the phonics screening check in the pilot because one of the aims of the work was to gather evidence about the policy, but they may be reluctant to enter those children in the future. The presumption of entering as many children for the phonics screening check where possible should still apply, although teachers should continue to consider disapplication where children have shown no understanding of grapheme-phoneme correspondences. In roll-out, parents should be informed in advance if schools to choose to disapply their child.

Guidance for schools on actions to take after the phonics screening check

Several stakeholder organisations argued strongly that a mark sheet should be designed to encourage teachers to analyse error patterns of children during the phonics screening check. This may help teachers to identify children who should be assessed for a particular type of SEN which was previously undiagnosed. For example, children with hearing impairments may consistently decode letters with high frequency sounds incorrectly, and children with speech, communication and language needs are more likely to decode a word incorrectly in relatively unpredictable ways.

The stakeholder organisations also felt that the Department should consider developing a mark sheet for roll-out which allows teachers to record the phonemes which have been read correctly and whether the blending was correct on each word. This would need to be consistent with supporting teachers to score the phonics screening check accurately. Using the mark sheet to

record more detailed information should also be optional for schools to avoid increasing burdens on teachers.

The Department will consider the stakeholder organisations' request that guidance be provided to schools on the next steps they could take if they suspect a child should be assessed for a particular type of SEN following the phonics screening check.

Specific issues for child groups

Children with vision impairment

Two children in one pilot school were Braille users. Given the age of children taking the phonics screening check, there is a risk that the phonics screening check will partially assess a child's tactile skills as well as their decoding skills. Although this can only be indicative given a very small sample, the children using Braille gained similar scores on the phonics screening check to their peers at the same phase of the school's phonics programme.

The teacher noted some errors when children were decoding letters which were mirror images of each other in Grade 1 Braille (for example 'e'/'i' and 'j'/'h'). Although this could show an issue with tactile rather than decoding skills, it was agreed that responses displaying this type of error should be scored incorrect. There is some similarity with the error non-vision impaired children can make by swapping 'b' and 'd' when decoding. Teachers administering the phonics screening check should be aware of the possibility of this type of error when considering any actions to follow-up from the phonics screening check.

Children using Braille showed no particular difficulty understanding the pseudo-word activity, despite the absence of monster pictures to accompany some of the pseudo-words.

Blending can be more difficult skill for children using Braille because it is harder to scan back to the first grapheme after sounding out all the letters using Braille as opposed to a visual check. The phonics screening check should pick-up issues with blending, and so it was agreed that no special arrangements were necessary to address this point. However, in guidance, teachers should be encouraged to remind children using Braille about the importance of blending before starting the phonics screening check.

It was agreed that Grade 1 Braille is appropriate for the phonics screening check, but care should be taken to produce high quality materials for the national roll-out.

It was agreed that other children with vision impairment should be able to show their ability in phonics using this phonics screening check, as long as schools are able to modify the materials in an appropriate way. The electronic versions of the phonics screening check allowed this to happen appropriately in the pilot.

Children with hearing impairment

Children using British Sign Language to spell out individual letters are not using phonics in the sense of linking letters and sounds, and so consideration should be given to disapplying these children from the phonics screening check. When reading a real word, children may be able to show the sign for each letter and then the sign for a whole word. However, this is a different process to sounding out and blending a word because this relies of knowledge of the word's meaning.

Children are able to access the phonics screening check using Visual Phonics or Cued Speech, which allows children to clarify their pronunciation of letters using hand signals. In this case teachers should use their normal way of working with a child to ensure they can access the assessment.

For all children, and particularly those with less severe hearing impairments it is essential to ensure the phonics screening check takes place in a room with good acoustics and lighting. This should be communicated strongly in guidance for teachers, particularly given that it was an issue in a minority of schools in the pilot.

Children on the autism spectrum

Some children on the autism spectrum were less comfortable with the monster pictures attached to the pseudo-words, which created confusion rather than clarifying the activity. The guidance for teachers needs to be clear that this is a possibility for some children with autism (although not all) and careful consideration should be given to removing the pictures before the assessment. Practice items could be used in advance to determine the best approach for each child.

Care is also required when explaining the activity of decoding pseudo-words to some children on the autism spectrum. A child's usual teacher is well-placed to explain the activity in a suitable way for each child, and the guidance should continue to encourage teachers to use their professional judgement when explaining the activity.

Rest breaks were used with some children on the autism spectrum and some children with speech and language needs. This strategy worked for some children, who were able to finish the phonics screening check as a result, although a small minority of children were reluctant to re-start the phonics screening check. Teachers should use rest breaks according to their professional judgement, but it should be made clear in guidance that this can be a useful strategy, which can be used with any children at the school's discretion.

Children with Speech, Language and Communication Needs (SLCN)

Explaining the activity carefully is particularly important for some children with SLCN. Some children in this group would benefit from some familiarisation activities before the week of the phonics screening check, particularly in relation the pseudo-words. This proved effective for some schools in the pilot.

It is particularly important that a teacher who regularly hears the child read administers the assessment for this group of children. This helps to ensure that scoring of child's responses is accurate and the assessment can focus on ability in decoding using phonics and children are not disadvantaged if they struggle to pronounce phonemes correctly having decoded the word. One teacher in the pilot made an audio recording of their children in the pilot, and listened again to one response to check the scoring was correct.

9. Arrangements for roll-out

STA is responsible for the roll-out of the Year 1 phonics screening check. In September 2011, schools were informed that the phonics screening check would be rolled out in the academic year 2011/12.

Full details of the arrangement for roll-out can be found in the Year 1 phonics *Assessment and reporting arrangements* guidance (ARA) available at www.education.gov.uk/assessment.

However, from a technical perspective, the key arrangement will be the collection of item level data from a sample of 10,000 children. This item level data captured as part of the roll-out will be used to validate the analysis carried out on data obtained through the pilot. A further technical report will be available in the Autumn term 2012.

The Department is also producing guidance for schools on how to respond when pupils do not achieve the standard on the phonics screening check. The response of schools may be different depending on a number of factors:

- how close the child's score was to the threshold;
- the age of the pupil and the amount of phonics teaching they have received; and
- further diagnostic work undertaken by the schools to determine the gaps in the child's knowledge and skills.

This guidance will be provided by April 2012 at www.education.gov.uk/schools/teachingandlearning/pedagogy/a00197709/developing-a-new-year-1-phonics-screening-check

10. Conclusion

This section of the report will focus on the Ofqual common assessment criteria (Ofqual, 2012) and will attempt to demonstrate the quality of the Year 1 phonics screening check from the evidence base on which the decision to roll-out the phonics screening check in the academic year 2011/12.

10.1 Validity

The Ofqual regulatory framework for national assessments (2011) states that an assessment should 'generate outcomes that provide a valid measure of the knowledge, skills and understanding that the learner is required to demonstrate as specified by the assessment objectives'. It states that 'Validity is the central concept in the evaluation of the quality of assessments' such that 'processes and procedures [are] expected to [ensure and generate] evidence to support the way in which the assessment outcomes are interpreted and used'. The document also states that:

The validity of an assessment refers to the extent to which evidence and theory support the interpretation that the assessment outcomes meet their intended uses.

The evaluation of validity involves the development of a clear argument to support the proposed interpretation of the outcomes and as a consequence the intended uses of the assessment. The validity argument should be built on statements of the proposed interpretation and supporting evidence collected from all stages of the assessment process.

Therefore, the development of a validity argument must start with an understanding of the purpose of the assessment. The Department has stated that the purpose of the phonics screening check will be to confirm whether or not children have learned phonic decoding to an age-appropriate standard such that those children who have not met the standard are provided with additional support to catch-up. As a result there are several questions that need to be answered to ensure that the assessment is sufficiently valid:

- Is the phonics screening check an appropriate assessment of phonic decoding?
- Is the expected standard on the phonics screening check age-appropriate?
- Are children who have not met the standard on the phonics screening check in need of additional support?

In relation to the first question, the pilot has collected a great deal of evidence relating to the content of the phonics screening check. The experts involved in the development of the phonics screening check have a wealth of expertise and experience in the field and have validated the specification for the phonics screening check throughout the process. Although the independent experts who reviewed the phonics screening check materials raised some concerns with the specification, these were relatively minor and not consistent. On balance, the evidence from the independent experts provides the Department with sufficient confidence that the phonics screening check is assessing phonic decoding appropriately.

In the evaluation the vast majority of teachers (in most cases over 90 per cent) felt that all elements of the content of the phonics screening check were suitable for children at the end of Year 1. Some concerns were regarding the use of pseudo-words (21 per cent of respondents) and unfamiliar real words (20 per cent of respondents). However, there are strong arguments in favour of including such words in the phonics screening check to ensure children are not relying on visual memory.

The review of the phonics screening check against several phonics programmes also lends weight to the argument that the phonics screening check is appropriate in terms of the phonemes and word structures covered.

Therefore, the Department believes that the phonics screening check is an appropriate assessment of phonic decoding for children at the end of Year 1.

In relation to the second question, the expected standard was developed in conjunction with a group of leading experts in the field. The descriptor was reviewed and independently verified by two groups of practicing teachers involved in the standard setting process. They believed that it was an appropriate expectation to have despite the fact that, based on the pilot data⁹, a minority of children at the end of Year 1 were currently achieving this standard.

The Department, therefore, believes that the expected standard on the phonics screening check is age appropriate.

The answer to the final question cannot be answered until live results are returned to schools. However, feedback will be requested from schools and will be reported in the final technical report on the pilot and initial roll-out in Autumn 2012.

The development of a validity argument is an on-going process and the Department will continue to collect evidence to demonstrate that the phonics screening check is a sufficiently valid for the purpose for which it is intended.

10.2 Reliability

The Ofqual regulatory framework for national assessments (2011) states that an assessment should 'generate outcomes that provide a reliable measure of a learner's performance'. The document also states that:

Reliability is about consistency and so concerns the extent to which the various stages in the assessment process generate outcomes which would be replicated were the assessment repeated. Reliability is a necessary condition of validity, as it is not possible to demonstrate the validity of an assessment process which is not reliable. The reliability of an assessment is affected by a range of factors such as the sampling of assessment tasks and inconsistency in marking by human markers.

To demonstrate sufficient reliability for the phonics screening check, the following aspects must be considered:

- The internal consistency

⁹ For the reasons stated in section 4.6, it is not possible to generalise national results in roll-out from the results in the pilot.

- The classification consistency
- The classification accuracy
- The consistency of scoring

The analysis of the evidence from the pilot has demonstrated high levels of internal consistency for the phonics screening check, with values of Cronbach's alpha of around 0.95. A reasonable standard error of measurement has been identified such that one can be 95 per cent confident that a child's true score lies within two standard errors of measurement of their observed score i.e. we are 95 per cent confident that a child's true score lies within the region of the observed score plus or minus 5 marks.

Classification consistency refers to the extent to which children are classified the same way in repeated applications of a procedure. Evidence from the check-re-check study indicates that approximately 90 per cent of children have been consistently classified.

Classification accuracy refers to how precisely children have been classified. Reasonable estimates of classification accuracy will only be valid once the phonics screening check has been administered in all schools. Therefore, further work on reliability will be analysed and reported in autumn 2012.

Consistency of scoring relates to the extent to which children are classified the same way when scored by different teachers. Evidence from the inter-rater reliability study indicates that even with the limitations of the study, 92 per cent of children have been consistently classified.

At present, the Department is satisfied that the phonics screening check is a sufficiently reliable assessment.

10.3 Comparability

The Ofqual regulatory framework for national assessments (2011) states that an assessment should 'generate outcomes that are comparable in standards over time'. The document also states that:

Comparability is about generating assessment outcomes that are comparable in standards over time and between assessment cycles. Where a test has equivalent forms – as is the case with National Curriculum assessments, where, for example, the Key Stage 2 mathematics test in each year comprises different items, but is still treated as the same test over time – then it is important to ensure comparability of outcomes.

When introducing a new assessment such as the phonics screening check, there are often no existing assessments with which to be comparable. However, the pilot has trialled sufficient items to develop up to nine different phonics screening checks. Since all items have been trialled together in a cross-over design, with items appearing in different forms, it has been possible to link all items together on a single scale. This ensures a number of different assessment instruments can be developed to be comparable.

10.4 Minimising bias

The Ofqual regulatory framework for national assessments (2011) states that an assessment should 'minimise bias, differentiating only on the basis of each learner's ability to meet National Curriculum requirements and early learning goals'. The document also states that:

Minimising bias is about ensuring that an assessment does not produce unreasonably adverse outcomes for particular groups of learners. The minimisation of bias is related to fairness to all children and is also closely related to statutory equality duties.

The qualitative evidence shows that although children with weaker phonics skills, often those children with SEN, find the phonics screening check difficult, the access arrangements that have been put in place make the assessment accessible to children. This has been verified by groups representing a full range of special educational needs.

It is true that certain groups of children do perform less well on the phonics screening check, e.g. boys, children receiving free school meals and children with SEN. However, these groups have traditionally performed less well on National Curriculum assessments so this is not unexpected. This also does not mean that the phonics screening check is biased against them since it is possible that the phonics screening check is simply correctly identifying their lower phonics skills.

Analysis has shown a difference in overall scores when comparing the term in which children were born, with younger children likely to do less well on the phonics screening check. This outcome was anticipated based on experience from other National Curriculum tests. However, the Government has high expectations for all children and therefore the policy does not take into account age for the purpose of setting standards on the phonics screening check. Therefore no attempt has been made to account for the age of the child. Schools, however, should consider this when deciding the appropriate interventions to put in place for children who do not meet the expected standard in the phonics screening check.

The evidence presented from the analysis of differential item functioning for gender, EAL and SEN provides the Department with confidence that there is no strong evidence that the Year 1 phonics screening check is discriminating between children on the basis of anything other than ability to decode using phonics.

10.5 Manageability

Most of the evidence relating to the Ofqual common criteria of manageability comes from the evaluation survey. The Ofqual regulatory framework for national assessments (2011) states that an assessment should be 'manageable so that the scale of the assessment process is balanced by the usefulness of the outcome'. The document also states that:

Manageability relates to the feasibility of carrying out particular assessment processes. A manageable assessment process is one which places reasonable demands on schools and children. The evaluation of the reasonableness of the demands will be based on the scale of the assessment process on the participants, balanced by the usefulness of the outcomes. As with the other common criteria (validity, reliability, comparability and minimising bias), judgements about manageability must be balanced with considerations around the other common criteria.

The responsible body or bodies are expected to demonstrate that there are appropriate documented procedures in place to meet the criteria.

At a high level, the evaluation report gives some positive messages on manageability. In response to the specific question ‘How manageable was the time commitment for the year 1 phonics screening check’, 65 per cent of schools said straightforward or very straightforward to manage, and only 19 per cent said difficult or very difficult to manage, with 16 per cent neutral. This was probably helped by the clarity of the guidance, which 89 per cent of schools stated was useful or very useful. 85 per cent of schools said that the phonics screening check took between four and nine minutes to administer per child with an average of three hours of preparation time required.

As could be anticipated, administration of the phonics screening check took longer in larger schools with more children, however the full report of the evaluation states that schools were already identifying ways to speed up the process and some indicated that they would be quicker next time.

There were concerns raised about the amount of time taken to administer the phonics screening check and concerns that cover for staff would be an issue during roll-out. However, 43 per cent of schools stated that the phonics screening check had helped identify issues of which they were not previously aware. Since the purpose of this assessment is to identify children that have not reached an appropriate standard in phonic decoding to enable additional support to be put in place to ensure they catch-up with their peers, it appears that the check fulfilled this purpose in these schools. Although for many schools, the phonics screening check did not identify particular issues of which the school was not aware, for these schools the phonics screening check does provide an indication of the national standard that we expect all children to achieve by the end of Year 1.

The Department therefore believe that the administration of the phonics screening check is a valuable use of teachers’ time as part of their on-going assessment of children and that the evidence from the evaluation of the pilot indicates that the Year 1 phonics screening check is manageable for schools in this context.

10.6 Overall statement in relation to common criteria

Having examined all of the evidence gathered so far through the pilot, the Department is satisfied that the Year 1 phonics screening check is sufficiently valid for the defined purpose, with acceptable levels of reliability, which is fair for children and manageable for schools. However, as stated previously, the development of a validity argument is an on-going process and additional analysis will be carried out to ensure that the Department can continue to be confident in this assertion.

Annex A – Decision log for the assessment framework

No	Page	Section	Area	Decision	Evidence
1	22	Test administration	Administration	Must be administered by a teacher	In the consultation 59% of respondents agreed that the screening check should be administered by a teacher. Ministers therefore took this decision.
2	16	Cognitive domain	Applying	Definition as given	Definition provided by phonics experts and updated following independent review.
3	9	Check structure	Bigrams	In the check, no bigrams will be classified as low frequency	Low frequency bigrams are unlikely to appear in words commonly read by children in Year 1 and in many cases will not be regular. Including them in the test would reduce the validity of the assessment.
4	9	Check structure	Bigrams	No more than 25% of bigrams will be considered medium frequency	Medium frequency bigrams will not appear regularly in words commonly read by children in Year 1, but they will be present and may be phonically regular. Including some medium frequency bigrams allows a wider range of words to be administered in the check, including words with some of the less common letters (e.g. x, j, z) which mostly appear in medium frequency bigrams.
5	29	Appendix B	Bigrams	The bigrams indicated will not be allowed in one syllable words	The bigrams shown are low frequency
6	31	Appendix B	Bigrams	The bigrams indicated will not be allowed at the start of words	The bigrams shown do not appear at the start of phonically regular words
7	31	Appendix B	Bigrams	The bigrams indicated will not be allowed at the end of words	The bigrams shown do not appear at the end of phonically regular words
8	31	Appendix B	Bigrams	The bigrams indicated will not be allowed at the end of two syllable words	The bigrams shown do not appear at the end of phonically regular words
9	32	Appendix B	Bigrams	Bigram frequencies of High, Medium and Low are taken from the paper Case-sensitive letter and bigram frequency counts from large-scale English corpora by Michaels N Jones and D J K Mewhort from Queen's University, Kingston, Ontario, Canada	A number of possible bigram frequency tables were available to determine high, medium and low frequency bigrams. The tables selected were recommended by the phonics experts and were based on the review of 183 million words in the English language, which was felt to be sufficiently representative for this purpose.

10	32	Appendix B	Bigrams	High frequency bigrams $x \geq 10$, Medium frequency bigrams $5 \leq x < 10$, Low frequency bigrams $x < 5$ where $x = \ln(\text{round}[\exp(x)])$	A review of the tables indicated that this categorisation gave 160 low frequency bigrams, 262 medium frequency bigrams and 254 high frequency bigrams. This distribution was felt to be acceptable for this purpose since there is no agreed definition.
11	7	Content standards	Content standards	Assessed content standards are as given	Agreed by phonics experts and validated by independent review.
12	7	Non-assessed content standards	Content standards	Non-assessed content standards are as given	Agreed by phonics experts and validated by independent review.
13	8	Check structure	Format	Context to be provided for the first four pseudo-words in section 1 and the first two pseudo-words in section 2 - picture of an imaginary creature such that children are asked to name the type of creature	Context used in pre-trialling where feedback was positive. Final evaluation recommended that pictures be included with all pseudo-words.
14	9	Check structure	Format	All letters will be lower case	Agreed by phonics experts.
15	9	Check structure	Format	The standard version of the check will contain four words per page	Decision following pre-trialling to ensure clarity and ensure that children moved at a steady pace - in pre-trialling there were more words per page and some children rushed as a result.
16	23	Access arrangements	Format	White paper will be used for the standard version	Decision made for cost reasons. Schools allowed to print on different coloured paper if required.
17	23	Access arrangements	Format	Gloss laminate will be used on the standard version	Decision made for cost reasons. Schools allowed to print non-laminated version if required.
18	23	Access arrangements	Format	Sassoon infant font will be used on the standard version	Decision following pre-trialling as it is the font in which letters are most clearly different. Some issues during trialling, although font is similar to KS1 font and schools allowed to print in different font if required.
19	23	Access arrangements	Format	Font size on standard version will be 55 bold	Decision following pre-trialling to ensure clarity on page. Schools allowed to print in different font size if required.
20	23	Access arrangements	Format	Braille version will be grade 1	Agreed by VI experts at SEN steering group as the most appropriate for children of this age and this type of screening check.
21	8	Check structure	Frequency of real words	Real words will include between 40% to 60% low frequency words	Agreed by phonics experts and validated by independent review.
22	17	Item structure	Frequency of real words	Low frequency words are defined as fewer than 20 occurrences per million words in the Children's Printed Word Database	A review of the database showed that this categorisation gave 9027 low frequency words and 3425 high frequency words. This distribution was felt to be appropriate for this purpose since there is no agreed definition.

23	17	Item structure	Frequency of real words	In the check between 40% and 60% of words will be low frequency	Agreed by phonics experts and validated by independent review.
24	8	Check structure	Graphemes	Each check will contain every letter of the alphabet	Agreed by phonics experts.
25	8	Check structure	Graphemes	Not all grapheme-phoneme correspondences listed in the framework will be included in every check	Agreed by phonics experts.
26	8	Check structure	Graphemes	All grapheme-phoneme correspondences listed in the framework will be included over five years	Agreed by phonics experts.
27	8	Check structure	Graphemes	Inclusion of a grapheme will not necessarily be in proportion to its frequency in words that are appropriate for children at the end of year 1	Agreed by phonics experts.
28	20	Item piloting and item difficulty	Item observations	1000 observations per item for each of the 360 items in the trial	Agreed by psychometrician in line with international best practice.
29	16	Cognitive domain	Knowing	Definition as given	Definition provided by phonics experts and updated following independent review.
30	17	Item structure	Neighbourhood size	Neighbourhood size for real and pseudo-words will be taken from Medler, D.A., & Binder, J.R. (2005). MCWord: An On-Line Orthographic Database of the English Language. http://www.neuro.mcw.edu/mcword/	A number of databases are available which provide neighbourhood sizes. This one was recommended by phonics experts. Since this database was used to create all real words used in the check and was capable of provide neighbourhood sizes for pseudo-words, this database was selected.
31	17	Item structure	Neighbourhood size	Low neighbourhood will be defined as $N < 5$	There is no agreed definition here. Once all real words had been generated, a review of neighbourhood sizes showed that the average neighbourhood size was 4.6, with 60% of words with $N < 5$ which was felt to be appropriate for this purpose.
32	17	Item structure	Neighbourhood size	In the check, between 40% and 60% of words will have a low neighbourhood size	Agreed by phonics experts.
33	8	Check structure	Number of words	20 real and 20 pseudo-words	Agreed by phonics experts and validated by independent review.

34	8	Check structure	Practice words	4 practice words on a separate sheet to be used with every form of the check - two real and two pseudo-words with orthographic structure CV and CVC	Recommended by pre-trial schools.
35	9	Check structure	Pseudo-words	Will not be homophones for real words	Agreed by phonics experts.
36	5	The purpose of the Year 1 phonics screening check	Purpose	Purpose agreed by Ministers	
37	24	Scoring	Scoring	Child may sound out phonemes before blending	Agreed by phonics experts and recommended by pre-trial schools.
38	24	Scoring	Scoring	Children may elongate phonemes as long as they are blended fully to form the word (any gaps between phonemes would be scored as incorrect)	Agreed by phonics experts.
39	24	Scoring	Scoring	For real words inappropriate grapheme-phoneme correspondences should not be marked correct	Agreed by phonics experts.
40	24	Scoring	Scoring	All plausible alternative pronunciations of graphemes in pseudo-words will be acceptable	Agreed by phonics experts.
41	24	Scoring	Scoring	Children are allowed to self-correct	In the consultation, 91% of respondents felt that self-correction should be allowed
42	24	Scoring	Scoring	If children make several attempts, the final attempt should be scored	Agreed by phonics experts.
43	24	Scoring	Scoring	There is no time limit for children to respond	Agreed by phonics experts and recommended by pre-trial schools.
44	10	Content domain	Section 1	Grapheme phoneme correspondences as shown	Agreed following programme review.
45	11	Content domain	Section 1	Orthographical representations as shown	Agreed by phonics experts and validated by independent review.
46	12	Content domain	Section 2	Grapheme phoneme correspondences as shown	Agreed following programme review.
47	14	Content domain	Section 2	Orthographical representations as shown	Agreed by phonics experts and validated by independent review.

48	31	Appendix B	Trigrams	The VCV trigrams will not be used in split digraph words	Agreed by phonics experts and validated by independent review.
49	32	Appendix B	Trigrams	The only trigrams including a consonant digraph in the check will be shr and thr	Agreed by phonics experts and validated by independent review.
50	32	Appendix B	Trigrams	The only trigrams including three consonants at the start of one syllable words will be scr, spl, spr and str	Agreed by phonics experts and validated by independent review.
51	9	Check structure	Two syllable words	Will be real words	Difficult to invent polysyllabic pseudo-words with limited alternative pronunciations that can be scored reliably - exacerbated by potential for different stresses in two syllable words
52	20	Item structure	Two syllable words	Will contain between five and eight letters	Agreed by phonics experts.
53	20	Item structure	Two syllable words	No compound words	Agreed by phonics experts.
54	28	Appendix B	Unigrams	Words will not begin with an x	Agreed by phonics experts and validated by independent review.
55	28	Appendix B	Unigrams	One syllable words will not end with c, j, q, v, y	Agreed by phonics experts and validated by independent review.
56	28	Appendix B	Unigrams	Two syllable words will not end with a, l, j, o, u, v	Agreed by phonics experts and validated by independent review.

Annex B – Specification for item review

DEPARTMENT FOR EDUCATION

RESEARCH SPECIFICATION

Word Review for the Year 1 Phonics Screening Check Pilot

1 INTRODUCTION

The Department for Education (DfE) intends to commission a review of words for inclusion in the Year 1 Phonics Screening Check Pilot. The review will include a check of the words against a number of criteria. This specification sets out the background, requirement for the review and details of the tendering process.

2 BACKGROUND

In 2010, national results showed that just over 15% of 7 year olds did not achieve the expected level in reading (level 2) at the end of Key Stage 1. At the end of Key Stage 2 this figure is higher with almost one in five 11 year olds (17%) not achieving the expected level (level 4), an increase of 3 percentage points since 2009 (DfE, 2010).

The Government is committed to raising pupils' achievement in reading. As part of this the DfE intends to establish a Phonics Screening Check for pupils in Year 1. This commitment is mentioned in the recent Schools White Paper and the DfE business plan. The Phonics Screening Check will be a short, light-touch Check designed to confirm that pupils have grasped the basics of phonic decoding. Pupils who do not reach the expected level when they first take the Phonics Screening Check will receive additional support to help them catch up. It is proposed that these pupils will re-take the Phonics Screening Check by the end of the autumn term in Year 2 to ensure they are able to decode using phonics at the appropriate level. It is expected the Phonics Screening Check will be rolled-out nationally during the 2011-12 academic year.

The results of the Phonics Screening Check will provide valuable information to parents/carers and will ultimately form part of the arrangements for the statutory assessment of pupils in respect of the first Key Stage.

Pilot of the Phonics Screening Check

Having undertaken small-scale pre-piloting in the 2010 autumn term, the DfE will pilot the Phonics Screening Check with 300 schools to approximately 10,000 pupils in Year 1, in the summer of 2011. The purpose of the Pilot is twofold. Firstly, to undertake technical pre-testing on a variety of test items, the results of which will be used to devise a number of versions of the Phonics Screening Check for national roll-out. Test items will be piloted using different versions of the Check, though each version will contain the same orthographical structures. This element of the pilot will be undertaken by the DfE in conjunction with the QCDA, and is not included in this tender. The second purpose is to assess the suitability of the Phonics Screening Check administration process in schools which is the subject of this invitation to tender. The Pilot will only report results back to schools after the standard setting process has been completed in

September. It will therefore not be possible during the Pilot to assess the impact of additional intervention and subsequent re-testing on under-achieving pupils.

It is expected the Pilot will be administered by schools week commencing 13th June 2011. Participation in the Pilot is voluntary. Invited schools will be sourced from a sample of 450 schools representative by Key Stage 1 reading results, pupil type (indicating SEN, EAL and FSM eligibility), school type and region. All 450 schools were invited in January 2011 to express their interest in participating in the Pilot. A sample of 300 will be chosen to participate from those who have registered interest. Each participating school will receive a £250 incentive to take part.

The success of the Pilot will be assessed via a range of required elements in addition to that being commissioned here. These will include:

- An assessment of whether the Check meets Ofqual manageability criteria – this will be conducted by the DfE and reviewed by Ofqual.
- Psychometric, validity and reliability testing – designed by the Qualifications and Curriculum Development Agency (QCDA) and its successor and delivered by the DfE.
- An analysis of pupil performance on the Pilot Check and validity/reliability tests, including the type of phonics being taught by participating schools – to be conducted by the QCDA.

This specification relates to the requirement to review words before they are included in the check to ensure their appropriateness.

3 RESEARCH REQUIREMENT

The successful contractor will be provided with a list of 280 real and 290 pseudo-words for review. The words will be grouped according to the requirements of the check specification (see Annex A).

Each real words should be reviewed to ensure the following:

- The word meets the requirement of the specification
- The words is phonically decodable, taking into account regional accents
- The word does not have an inappropriate meaning colloquially or in a regional dialect

Each pseudo-word should be reviewed to ensure the following:

- The word meets the requirement of the specification
- The words is phonically decodable, taking into account regional accents
- The word is not a homophone for an inappropriate word in another language (common words in other languages should also be flagged although it may still be deemed appropriate to include these words)
- The word is not used colloquially or in a regional dialect

A list of the most common languages of pupils in schools is provided in Annex B.

4 METHODOLOGY

Tenders should include details of the proposed methodology for completing the review, including an indication of the languages that can be covered.

4.1 Data security

Tenderers should outline how the security of the word list will be ensured during and after the study.

5 PROJECT MANAGEMENT

The project will be managed at the DfE by the Assessment Division, Education Standards Directorate.

6 TIMETABLE AND OUTPUTS

It is proposed that the review takes place from 29 March 2011 to 18 April 2011.

The contractor will be expected to produce:

- A completed spreadsheet containing positive affirmation that each word has been reviewed against each of the criteria by 18 April 2011.

Annex A - Specification

Words in section 1 may only contain the following grapheme-phoneme correspondences

a	/æ/	b	/b/	c	/k/	d	/d/	e	/e/
ar	/ɑ:/			ch	/tʃ/			ee	/i:/
				ck	/k/			er	/ɜ:/
f	/f/	g	/g/	h	/h/	i	/ɪ/	j	/dʒ/
ff	/f/								
k	/k/	l	/l/	m	/m/	n	/n/	o	/ɒ/
		ll	/l/			ng	/ŋ/	oi	/ɔɪ/
								or	/ɔ:/
p	/p/	qu	/k/w/	r	/r/	s	/s/	t	/t/
						sh	/ʃ/		
						ss	/s/		
u	/ʊ/ or /u/	v	/v/	w	/w/	x	/k/s/	y	/j/
ur	/ʊr/								
z	/z/								
zz	/z/								

Real words in section 1 will be grouped using the following orthographical structures (where CC means consonant digraph and VV means vowel digraph)

Group 1	<u>CCVC</u> , <u>CVCC</u> , <u>CVVC</u> , <u>CCVCC</u> and <u>CCVVC</u>
Group 2	<u>CCVC</u> , <u>CCVVC</u> and <u>CCVCC</u>
Group 3	<u>CVCC</u> , <u>CVVCC</u> and <u>CCVCC</u>

Following review there need to be 40 usable words in group 1, 20 usable words in group 2 and 20 usable words in group 3. If, during the review process, it appears that insufficient words will be available, this should be flagged with DfE and more words will be provided.

Pseudo-words in section 1 will be grouped as follows

Group 1	CVC
Group 2	VCC
Group 3	<u>CCVC</u> , <u>CVCC</u> , <u>CVVC</u> , <u>CCVCC</u> and <u>CCVVC</u>
Group 4	<u>CCVC</u> , <u>CCVVC</u> and <u>CCVCC</u>
Group 5	<u>CVCC</u> , <u>CVVC</u> and <u>CCVCC</u>

Following review there need to be 30 usable words in group 1, 10 usable words in group 2, 40 usable words in group 3, 20 usable words in group 4 and 20 usable words in group 5. If, during the review process, it appears that insufficient words will be available, this should be flagged with DfE and more words will be provided.

Words in section 2 may only contain the following grapheme-phoneme correspondences (bold indicates that it was not used in section 1)

a	/æ/ & /ɑ:/	ar	/ɑ:/	b	/b/	c	/k/ & /s/
a-e	/eɪ/	au	/ɔ:/			ch	/tʃ/ & /k/ & /ʃ/
ai	/eɪ/	aw	/ɔ:/			ck	/k/
air	/eə/	ay	/ai/				
d	/d/	e	/e/ & /i:/	er	/ɜ:/ & /ɜ/	f	/f/
/dʒ/		ea	/e/ & /i:/	ew	/u:/	ff	/f/
		ee	/i:/	ey	/eɪ/ & /i:/		
		e-e	/i:/			g	/g/ &
i	/ɪ/ & /aɪ/	igh	/aɪ/	j	/dʒ/	k	/k/
i-e	/aɪ/	ir	/ɜ:/			l	/l/
ie	/aɪ/ & /i/					ll	/l/
m	/m/	n	/n/	o	/ɒ/ & /əʊ/	oo	/u:/ & /ʊ/
		ng	/ŋ/	oa	/əʊ/	or	/ɔ:/
				oe	/əʊ/	ou	/aʊ/ & /u:/ & /ɜ/ & /əʊ/
				o-e	/əʊ/	ow	/aʊ/ & /əʊ/
				oi	/ɔɪ/	oy	/ɔɪ/

p	/p/	qu	/k/w/	r	/r/	s	/s/
ph	/f/					sh	/ʃ/
						ss	/s/
t	/t/	u	/c/ or /ʊ/ & /j/u:/	v	/v/	w	/w/
th	/θ/ & /ð/	ue	/u:/ & /j/u:/			wh	/w/
		u-e	/u:/ & /j/u:/				
		ur	/ur/				
x	/k/s/	y	/j/ & /aɪ/ & /ɪ/ & /i:/	z	/z/		
				zz	/z/		

Real words in section 2 will be grouped using the following orthographical structures (where VCV means split digraph and VVV means vowel trigraph)

Group 1	<u>CVVC</u> , <u>CVCV</u> and <u>CCVCV</u>
Group 2	<u>CCVCV</u> , <u>CCVVC</u> , <u>CCCVCC</u> , <u>CVVCC</u> , <u>CVV</u> , <u>CVVV</u> and <u>CCVVV</u>
Group 3	<u>CCVCC</u> , <u>CCVVCC</u> and <u>CCCVCC</u>
Group 4	<u>CCCVC</u> , <u>CCCVCV</u> , <u>CCCVCC</u> and <u>CCCVV</u>
Group 5	5 letter two-syllable words
Group 6	6 letter two-syllable words
Group 7	7 letter two-syllable words
Group 8	8 letter two-syllable words

Following review there need to be 20 usable words in group 1, 20 usable words in group 2, 20 usable words in group 3, 20 usable words in group 4, 10 usable words in group 5, 10 usable words in group 6, 10 usable words in group 7 and 10 usable words in group 8. If, during the review process, it appears that insufficient words will be available, this should be flagged with DfE and more words will be provided.

Pseudo-words in section 2 will be grouped as follows

Group 1	<u>CVVC</u> , <u>CVCV</u> and <u>CCVCV</u>
Group 2	<u>CCVCV</u> , <u>CCVVC</u> , <u>CCCVCC</u> , <u>CVVCC</u> , <u>CVV</u> , <u>CVVV</u> and <u>CCVVV</u>
Group 3	<u>CCVCC</u> , <u>CCVVCC</u> and <u>CCCVCC</u>
Group 4	<u>CCCVC</u> , <u>CCCVCV</u> , <u>CCCVCC</u> and <u>CCCVV</u>

Following review there need to be 20 usable words in group 1, 20 usable words in group 2, 20 usable words in group 3 and 20 usable words in group 4. If, during the review process, it appears that insufficient words will be available, this should be flagged with DfE and more words will be provided.

The following rules also apply to each word:

Constrained unigrams

The following unigrams will not be used at the start of words in the check:

- x

The following unigrams will not be used at the end of one-syllable words in the check:

- c
- j
- q
- v
- y

The following unigrams will not be used at the end of two-syllable words in the check:

- a
- i
- j
- o
- u
- v

Constrained bigrams

The following bigrams will not be used in one-syllable words in the check:

- | | | | | | | | |
|------|------|------|------|------|------|------|------|
| ▪ aa | ▪ cp | ▪ eq | ▪ gt | ▪ jh | ▪ kt | ▪ mt | ▪ pj |
| ▪ ae | ▪ cq | ▪ ey | ▪ gv | ▪ jj | ▪ kv | ▪ mv | ▪ pk |
| ▪ aj | ▪ cs | ▪ fb | ▪ gw | ▪ jk | ▪ kw | ▪ mw | ▪ pm |
| ▪ ao | ▪ cv | ▪ fc | ▪ gx | ▪ jl | ▪ kx | ▪ mx | ▪ pn |
| ▪ aq | ▪ cw | ▪ fd | ▪ gz | ▪ jm | ▪ ky | ▪ my | ▪ pp |
| ▪ bc | ▪ cx | ▪ fg | ▪ hb | ▪ jn | ▪ kz | ▪ mz | ▪ pq |
| ▪ bd | ▪ cy | ▪ fh | ▪ hc | ▪ jp | ▪ lc | ▪ nb | ▪ pv |
| ▪ bf | ▪ cz | ▪ fj | ▪ hd | ▪ jq | ▪ lg | ▪ nc | ▪ pw |
| ▪ bg | ▪ db | ▪ fk | ▪ hf | ▪ jr | ▪ lj | ▪ nf | ▪ px |
| ▪ bh | ▪ dc | ▪ fm | ▪ hg | ▪ js | ▪ ln | ▪ nh | ▪ py |
| ▪ bj | ▪ dd | ▪ fn | ▪ hh | ▪ jt | ▪ lq | ▪ nj | ▪ pz |
| ▪ bk | ▪ df | ▪ fp | ▪ hj | ▪ jv | ▪ lr | ▪ nl | ▪ qa |
| ▪ bm | ▪ dg | ▪ fq | ▪ hk | ▪ jw | ▪ lv | ▪ nm | ▪ qb |
| ▪ bn | ▪ dh | ▪ fv | ▪ hl | ▪ jx | ▪ lw | ▪ nn | ▪ qc |
| ▪ bp | ▪ dj | ▪ fw | ▪ hm | ▪ jy | ▪ lx | ▪ np | ▪ qd |
| ▪ bq | ▪ dk | ▪ fx | ▪ hn | ▪ jz | ▪ ly | ▪ nq | ▪ qe |
| ▪ bv | ▪ dl | ▪ fy | ▪ hp | ▪ kb | ▪ lz | ▪ nr | ▪ qf |
| ▪ bw | ▪ dm | ▪ fz | ▪ hq | ▪ kc | ▪ mb | ▪ nv | ▪ qg |
| ▪ bx | ▪ dn | ▪ gb | ▪ hs | ▪ kd | ▪ mc | ▪ nw | ▪ qh |
| ▪ by | ▪ dp | ▪ gc | ▪ hv | ▪ kf | ▪ md | ▪ nx | ▪ qi |
| ▪ bz | ▪ dq | ▪ gd | ▪ hw | ▪ kg | ▪ mf | ▪ ny | ▪ qj |
| ▪ cb | ▪ dt | ▪ gf | ▪ hx | ▪ kh | ▪ mg | ▪ nz | ▪ qk |
| ▪ cc | ▪ dv | ▪ gg | ▪ hy | ▪ kj | ▪ mh | ▪ oj | ▪ ql |
| ▪ cd | ▪ dx | ▪ gj | ▪ hz | ▪ kk | ▪ mj | ▪ oq | ▪ qm |
| ▪ cf | ▪ dy | ▪ gk | ▪ jb | ▪ kl | ▪ mk | ▪ pb | ▪ qn |
| ▪ cg | ▪ dz | ▪ gm | ▪ jc | ▪ km | ▪ ml | ▪ pc | ▪ qo |
| ▪ cj | ▪ ei | ▪ gn | ▪ jd | ▪ kn | ▪ mn | ▪ pd | ▪ qp |
| ▪ cm | ▪ ej | ▪ gp | ▪ jf | ▪ kp | ▪ mq | ▪ pf | ▪ qq |
| ▪ cn | ▪ eo | ▪ gq | ▪ jg | ▪ kq | ▪ mr | ▪ pg | ▪ qr |

- | | | | | | | | |
|------|------|------|------|------|------|------|------|
| ▪ qs | ▪ sg | ▪ tx | ▪ vq | ▪ wt | ▪ xo | ▪ yl | ▪ zk |
| ▪ qt | ▪ sj | ▪ ty | ▪ vr | ▪ wv | ▪ xp | ▪ ym | ▪ zl |
| ▪ qv | ▪ sr | ▪ tz | ▪ vs | ▪ ww | ▪ xq | ▪ yn | ▪ zm |
| ▪ qw | ▪ sv | ▪ uh | ▪ vt | ▪ wx | ▪ xr | ▪ yp | ▪ zn |
| ▪ qx | ▪ sx | ▪ uj | ▪ vv | ▪ wy | ▪ xs | ▪ yq | ▪ zp |
| ▪ qy | ▪ sy | ▪ uq | ▪ vw | ▪ wz | ▪ xt | ▪ yr | ▪ zq |
| ▪ qz | ▪ sz | ▪ uu | ▪ vx | ▪ xa | ▪ xu | ▪ ys | ▪ zr |
| ▪ rf | ▪ tb | ▪ uy | ▪ vy | ▪ xb | ▪ xv | ▪ yt | ▪ zs |
| ▪ rh | ▪ tc | ▪ vb | ▪ vz | ▪ xc | ▪ xw | ▪ yv | ▪ zt |
| ▪ rj | ▪ td | ▪ vc | ▪ wb | ▪ xd | ▪ xx | ▪ yw | ▪ zv |
| ▪ rl | ▪ tf | ▪ vd | ▪ wc | ▪ xe | ▪ xy | ▪ yx | ▪ zw |
| ▪ rq | ▪ tg | ▪ vf | ▪ wf | ▪ xf | ▪ xz | ▪ yy | ▪ zx |
| ▪ rv | ▪ tj | ▪ vg | ▪ wg | ▪ xg | ▪ yb | ▪ yz | ▪ zy |
| ▪ rw | ▪ tk | ▪ vh | ▪ wj | ▪ xh | ▪ yc | ▪ zb | ▪ zz |
| ▪ rx | ▪ tl | ▪ vj | ▪ wk | ▪ xi | ▪ yd | ▪ zc | |
| ▪ ry | ▪ tm | ▪ vk | ▪ wl | ▪ xj | ▪ yf | ▪ zd | |
| ▪ rz | ▪ tn | ▪ vl | ▪ wm | ▪ xk | ▪ yg | ▪ zf | |
| ▪ sb | ▪ tp | ▪ vm | ▪ wp | ▪ xl | ▪ yh | ▪ zg | |
| ▪ sd | ▪ tq | ▪ vn | ▪ wq | ▪ xm | ▪ yj | ▪ zh | |
| ▪ sf | ▪ tv | ▪ vp | ▪ wr | ▪ xn | ▪ yk | ▪ zj | |

The following bigrams will not be used at the start of words in the check:

- | | | | |
|------|------|------|------|
| ▪ bb | ▪ ks | ▪ mp | ▪ rg |
| ▪ ck | ▪ lb | ▪ ms | ▪ rk |
| ▪ cs | ▪ ld | ▪ nd | ▪ rm |
| ▪ ct | ▪ lf | ▪ ng | ▪ rn |
| ▪ ds | ▪ lk | ▪ nk | ▪ rp |
| ▪ ff | ▪ ll | ▪ ns | ▪ rr |
| ▪ fs | ▪ lm | ▪ nt | ▪ rs |
| ▪ ft | ▪ lp | ▪ pt | ▪ rt |
| ▪ gh | ▪ ls | ▪ rb | ▪ ss |
| ▪ gs | ▪ lt | ▪ rc | ▪ ts |
| ▪ ht | ▪ mm | ▪ rd | ▪ ws |

The following bigrams will not be used at the end of one-syllable words in the check:

- All vowel digraphs
- qu
- wh

The following bigrams will not be used at the end of two-syllable words in the check:

- | | | |
|------|------|------|
| ▪ ai | ▪ ie | ▪ ou |
| ▪ ar | ▪ oa | ▪ qu |
| ▪ au | ▪ oe | ▪ ur |
| ▪ ea | ▪ oi | ▪ wh |
| ▪ ew | ▪ oo | |

The following will be the only consonant clusters including a consonant digraph in the check (CCC):

- shr
- thr

The following will be the only consonant clusters including 3 consonants in the check (CCC):

- scr

- spl
- spr
- str

The following VCV trigrams will not be used in split digraph words:

- ahe
- ehe
- ihe
- ohe
- uhe
- aje
- eje
- ije
- oje
- uje
- aqe
- eqe
- iqe
- oqe
- uqe
- are
- ere
- ire
- ore
- ure
- axe
- exe
- ixе
- oxe
- uxe

Annex B - Languages

Language	% of school population
Urdu	1.5
Panjabi	1.3
Bengali	0.9
Polish	0.6
Gujarati	0.6
Somali	0.6
Arabic	0.4
Tamil	0.3
French	0.3
Portuguese	0.3
Turkish	0.3
Bengali (Sylheti)	0.3
Panjabi (Mirpuri)	0.2
Yoruba	0.2
Spanish	0.2
Albanian/Shqip	0.2
Pashto/Pakhto	0.1
Chinese	0.1
Hindi	0.1
Lithuanian	0.1
Tagalog/Filipino	0.1
Nepali	0.1
Malayalam	0.1
Shona	0.1
Italian	0.1
German	0.1
Persian/Farsi	0.1
Russian	0.1
Panjabi (Any Other)	0.1
Akan/Twi-Fante	0.1
Chinese (Cantonese)	0.1
Kurdish	0.1
Slovak	0.1
Farsi/Persian (Any Other)	0.1
Greek	0.1
Swahili/Kiswahili	0.1
Lingala	0.1
Igbo	0.0
Dutch/Flemish	0.0
Vietnamese	0.0
Akan (Twi/Asante)	0.0
Czech	0.0
Romanian	0.0
Bengali (Any Other)	0.0
Panjabi (Gurmukhi)	0.0

Tagalog	0.0
Thai	0.0
Caribbean Creole English	0.0
Filipino	0.0
Bulgarian	0.0
Pahari (Pakistan)	0.0
Hungarian	0.0
Japanese	0.0
Tigrinya	0.0
Latvian	0.0
Luganda	0.0
Arabic (Any Other)	0.0
Sinhala	0.0
Korean	0.0
Swahili (Any Other)	0.0
Serbian/Croatian/Bosnian	0.0
Dari Persian	0.0
Amharic	0.0
Chinese (Mandarin/Putonghua)	0.0
Katchi	0.0
Ebira	0.0
Swedish	0.0
Caribbean Creole French	0.0
Konkani	0.0
Panjabi (Pothwari)	0.0
Afrikaans	0.0

Annex C – Invitation to schools to participate in the pilot

Department for
Education
14 January 2011

This email is an official communication to a sample of primary schools from the Department for Education.

Please do not reply to this message by email as this mailbox is not monitored.

Invitation to Year 1 phonics screening check pilot

Dear colleague

The Government is introducing a Year 1 phonics screening check as part of its commitment to raise reading standards. A public consultation is open until 14 February asking for responses to the detailed policy proposals. We welcome responses to the consultation which can be found on the [Department's consultations website](#).

Ministers intend to introduce the screening check nationally from 2012 and we will be trialling it with a representative sample of 300 primary schools in June 2011. The trial will help to make sure that the screening check operates effectively.

This is an invitation to your school to participate in the June 2011 trial. The screening check will be a short, light touch assessment focused on children's decoding through phonics skills. It will be administered by staff known to the child.

Being part of the trial would involve:

Trialling test materials: one teacher from each pilot school will be asked to attend a training day early in the summer term for which we will pay £140 to cover supply needs. The teacher will then administer a short draft test to all Year 1 pupils in the school over a week-long period in summer 2011, likely to be 13-17 June. The test should take 5-10 minutes per pupil to administer, including the time taken to record the results.

Participation in monitoring and evaluation: the teacher will ensure the completion of surveys on the phonics schemes used and on the experience of the check, including feedback on how the guidance and check could be improved. The teacher would also arrange for pupils and parents to complete short surveys about their experiences of the check. This may include a visit from the evaluators who will monitor the administration of the trial in order to

inform any necessary changes or to look at the teaching of reading or to run focus groups with pupils.

Taking part in the trial will take approximately three days of the teacher's time, depending on the size of the school, and we will work to minimise this as far as possible. The Department will pay £250 per school towards the costs of preparation and administration time. The schools will also receive data about the performance of their pupils during the trial, which could help to inform subsequent teaching of phonics.

It is important that we have input from schools so that we can develop a reliable and appropriate test which is manageable for schools to administer and we would appreciate your participation. The screening check will play a key role in helping to identify pupils who need extra support in phonic decoding, which is a fundamental building block for understanding and learning to read for enjoyment

If you would like to express interest in taking part in the trial, or if you have questions, please contact Sheila McCreary by email on ScreeningCheck.PHONICS@education.gsi.gov.uk or by telephone on 020 7340 7973 by 11 February. Once you have expressed interest, we will contact you to confirm whether or not you will be part of the trial. For those taking part in the trial, we will provide more information about all aspects of the process, including timings, in due course.

Annex D – Administration guide

About this document

This guide contains information on administering the Year 1 phonics screening check.

To make the administration of screening check consistent for all schools, schools should follow the administration instructions in this guide.

The screening check should be administered during the week of **13 – 17 June 2011**.

What is this document about?

The Administration guide provides key information for administrators, including:

- what to do before the screening check (page 3)
- what to do at the start and during the screening check (page 7)
- dealing with queries and issues during the screening check (page 9)
- what to do at the end of the screening check (page 9).

This guide should be taken into each room where the screening check will take place, as it contains important information for administrators and advice on what to do if things do not go according to plan.

Who is this document for?

The Administration guide provides information for anyone administering Year 1 phonics screening check as part of the 2011 pilot.

This must be a teacher and in order to put pupils at ease, we would expect that they will be known to the child. In this guide, 'administrator' refers to anyone responsible for, or involved with, administration of the screening check.

Contact details

The phonics section of the Department for Education website, www.education.gov.uk/schools/teachingandlearning/pedagogy/teachingstyles/phonics, contains more information about the introduction of the Year 1 phonics screening check and phonics more generally.

If you need any further help, then please contact the Phonics pilot helpline:

Phone number: 020 7340 7494

Email: ScreeningCheck.Phonics@education.gsi.gov.uk

1. What to do before the screening check

1.1 Preparing yourself

You should familiarise yourself with this guide as it contains all of the information you need to know to administer the screening check. All of this material will also be covered in the training event that will be attended by someone in your school. Your school will also be provided with an exemplar screening check at the training event which you can use to familiarise yourself with the sort of materials you will receive for the pilot. A number of different forms of the screening check are being administered across schools to ensure we have sufficient suitable items to construct future screening checks. Within each school, a maximum of five forms will be administered.

The screening check pupil materials will be a 10 page spiral-bound booklet with 4 words on each page. Each page will contain either 4 pseudo-words or 4 real words. The first 5 pages will contain words in section 1 and the last 5 pages will contain the words in section 2. All words will be lowercase. The first few pseudo-words in each section will be accompanied by a picture of an imaginary creature. This picture will be used to provide context when introducing pseudo-words (see section *What to do at the start of the screening check?*). There will also be a double sided practice sheet containing two words on each side (2 pseudo-words and 2 real words) to familiarise pupils with the task. It is for teachers to decide whether to use the practice sheet, based upon their knowledge of the pupil.

The words in the standard version of the screening check will be printed on white paper and written in the font 'Sassoon Infant™', font size 55. Since the pupil materials have been designed to be used with a number of pupils, each page is laminated with gloss laminate. Schools will be provided with an electronic version of the check on request in order that they may make modifications to suit the needs of specific pupils, including, for example, larger font sizes (see *Access arrangements, including modification* section below)

1.2 Receipt and storage of screening check materials

At your training session you will be advised when your screening check materials will be delivered. In your consignment you will receive:

- Cover page with school name
- Pupil list
- Pre-printed mark sheets for all pupils on the pupil list (including 5 blank mark sheets for pupils who have arrived in school since the January school census)
- 1 copy of the practice sheet

- A copy of each of the screening check materials required for use with pupils
- Return address mail bag

You should check the content of the pack carefully and telephone the Phonics pilot helpline immediately on **020 7340 7494** if you have anything missing.

If the pack does not arrive on the specified day, please telephone the Phonics pilot helpline on **020 7340 7494**.

Each pupil has been pre-assigned a form of the screening check and a pre-printed marksheet has been provided for each pupil. Pupil details were taken from the January 2011 School Census and therefore if a pupil arrived in school after the census date, they will not have a pre-printed form. Schools have been provided with 5 blank forms for pupils who arrived in school after the School Census was completed. Full pupil details need to be provided for these pupils to enable DfE to match the results to the correct pupil when returning results.

If there are insufficient blank forms for the number of newly arrived pupils, schools should telephone the Phonics pilot helpline immediately on **020 7340 7494**.

All screening check materials should be stored securely whilst they are in school.

1.3 Access arrangements, including modification

The Year 1 phonics screening check has been designed to ensure that pupils working at the expected standard at the end of Year 1 can access them. A small number of pupils may require additional arrangements to enable them to access the screening check appropriately, including the use of modified versions of the screening check materials.

To decide whether an individual pupil needs access arrangements for the Year 1 phonics screening check, schools should consider both:

- the pupil's assessment needs; and
- the type and amount of support that they receive as part of normal classroom practice.

The access arrangements described in this guide may be appropriate for:

- pupils with a statement of special educational needs (as described in the Special educational needs (SEN) code of practice) or a local equivalent such as an Individual Pupil Resourcing Agreement (IPRA)

- pupils for whom provision is being made in school as part of School Action or School Action Plus under the Special educational needs (SEN) code of practice
- pupils who require alternative access arrangements because of a disability that may or may not give rise to a special educational need
- pupils with English as an additional language (EAL) and who have limited fluency in English.

It is not possible to list all the different situations in which pupils may need access arrangements. Schools should contact the DfE Phonics pilot helpline on **020 7340 7494** for information on specific situations not covered by this document. The following list gives details of the sort of access arrangements that may be used.

- additional time and rest breaks

It is expected that the screening check should take no more than 5 minutes to administer for most pupils, although there is no time limit and pupils should be given sufficient time to respond to each item. However, if you believe a pupil will find it difficult to concentrate or may suffer fatigue during the screening check, you may use rest breaks to make the screening check more manageable for the pupil. Schools should consider the most appropriate time to administer a rest break, which may be between the two sections. Rest breaks may only be taken at the end of a full page.

- school based modifications

An electronic version of each form of the screening check will be made available on request. Since each pupil will be pre-assigned a form of the screening check, the school will need to provide details of which pupils require modification to ensure the correct form of the check is provided. Schools should telephone the Phonics pilot helpline on **020 7340 7494** by **8 June 2011** to organise for an electronic version to be provided. Pupil details **must** not be sent via email because of data protection issues. The types of modification that can be made by schools will include:

- changing the font
- changing the font size
- reducing the number of words on each page
- printing on different coloured paper
- printing to provide a non-laminated version

- return envelope.
- Braille versions of the check

Braille versions of the check are available on request. Braille versions will be provided in grade 1 Braille without pictures of imaginary creatures. Since each pupil will be pre-assigned a form of the screening check, the school will need to provide details of which pupils require Braille versions to ensure the correct form of the check is provided. Schools should telephone the Phonics pilot helpline on **020 7340 7494** by **8 June 2011** to organise for a Braille version to be provided. Pupil details **must** not be sent via email because of data protection issues.

- use of coloured overlays

Schools may use a coloured overlay if this is normal classroom practice for a pupil.

- use of British Sign Language (BSL)

Pupils using BSL will be able to respond to real words using sign language. In the pilot, we will explore how pupils with hearing impairments who use BSL are able to respond to the pseudo-words and develop advice for national roll-out accordingly. Pilot schools with pupils who use BSL should contact the Phonics pilot helpline on **020 7340 7494**.

- rephrasing of instructions including the use of gestures

The instructions provided on how to introduce the screening check (see section *What to do at the start of the screening check?*) have been written to ensure all pupils have a standardised introduction to the check. However, if a pupil is likely to be confused by the standard introduction, schools may choose to develop their own introduction to the screening check. The instructions may refer to the practice items but must not refer to the main screening check items themselves.

1.4 Preparing the room

The screening check should be administered in an appropriate room which is free from excessive noise and provides a comfortable well lit space for the pupil to take the screening check. You must prepare any rooms where the screening check will take place before pupils are admitted. You should remove or cover any displays or materials that could help pupils in the screening check.

1.5 Pupils who are working below the level of the screening check

It is expected that the vast majority of pupils will be able to access the screening check, in particular section 1. However, pupils who have not shown any understanding of grapheme-phoneme correspondences in class should normally be disapplied. Disapplication should always be based on documented evidence from previous assessments (see section *What to do at the start of the screening check?* for details on how to record disapplication)

Parents should be informed about the reasons for any disapplication, and the steps the school has put in place to help the child to learn to read, before a final decision is made.

1.6 Arrangements for pupils who cannot take the screening check at the scheduled time

The screening check can be administered on any day during the administration week (13 – 17 June 2011). During the pilot, schools will not be able to carry out a timetable variation to administer the screening check the following week and pupils should be recorded as absent on the marksheet (see section *What to do at the start of the screening check?* for details on how to record absence).

2. What to do at the start and during the screening check

2.1 Ensuring the correct form materials are available

Each pupil has been pre-assigned a form of the screening check and you should make sure that the pupil booklet containing the appropriate words matches the marksheet for that pupil. For pupils who arrived in school after the January School Census a blank form should be used. All blank forms relate to the same form of the screening check within each school.

There is also a practice sheet containing simple words to use to introduce the screening check to pupils. The same practice sheet is used with all forms of the screening check.

2.2 Completing the marksheet

On marksheets, a cross 'X' or a slash '/' should be used to indicate the appropriate response to each question and screening check item. Please use **black ink** only on the marksheet.

	Correct	Incorrect		Correct	Incorrect
Word 1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Word 1	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input checked="" type="checkbox"/>		<input type="checkbox"/>	<input checked="" type="checkbox"/>
Word 2			Word 2		

Before starting the screening check, you should complete the appropriate information boxes at the top of the form including the day of the week the check was completed or the reason why the check was not completed (either because the pupil left the school, was absent for the whole week or was disapplied from the check).

For blank marksheets, Full pupil names, including middle names, gender and date of birth should be written on the marksheet.

2.3 What to do if you make an error on the marksheet

If you make an error on the marksheet, please fill in the box completely and place a cross in the correct box as below

	Correct	Incorrect		Correct	Incorrect
Word 1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Word 1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

2.4 What to say at the beginning of the screening check

You should try to introduce the screening check in a consistent way for all pupils. However, administrators may need to provide more explanation to some pupils to ensure that they understand the task. The following text provides an example of how the screening check should be introduced for schools to use if they wish. This text uses the practice sheet to introduce pupils to the screening check. The use of the practice sheet is optional.

In this activity, I am going to ask you to read some words aloud.

Some of these words you may have seen before and others will be new to you.

You should try to read each word but don't worry if you can't. If it helps you, you may sound out the phonemes before trying to say the word.

This practice sheet shows you what the words will look like.

Have a go at these 2 words which you should know [at and in].

The words on this side [turn over practice sheet] are not real words. They are names for types of imaginary creatures. For the first one, you can see a picture of the creature.

Can you read the words on this page for me?

Ok, now we are going to start reading the words in this booklet and I'm going to note down what you say on my sheet.

In this booklet, there are 4 words on each page. I will tell you at the start of each page whether they are real words that you may have seen before or types of imaginary creatures.

The first page has types of imaginary creatures; the first two have pictures again.

Can you start reading the words to me?

For each page, remember to tell the pupil whether they are real words or types of imaginary creatures.

2.5 Scoring and recoding on the marksheet

The check should be scored by the administrator as they work through the check. At the training session, you were provided with some more detailed

notes to help you scoring the screening check. All administrators should have access to the notes before they start administering the screening check.

For each word, the administrator will record with a cross on the marksheet whether the pupil read the word correctly or not bearing in mind the following points:

- Pupils may sound out phonemes before blending.
- Pupils may elongate phonemes as long as they are blended to form the word. However, if pupils leave gaps between phonemes and do not blend them, this must be scored as incorrect.
- Alternative pronunciations must be considered when deciding whether a response is correct. For real words inappropriate grapheme-phoneme correspondences should not be marked correct (for example, reading *blow* to rhyme with *cow* would be incorrect). However, alternative pronunciations of graphemes will be allowed in pseudo-words.
- A pupil's accent should be taken into account when deciding whether a response is acceptable. There should be no bias in favour of pupils with a particular accent.
- Any pronunciation difficulties for a pupil should be taken into account when deciding whether a response is acceptable (for example, a pupil who is unable to form the 'th' sound and instead says 'fw' should have this scored correct).
- If a pupil makes an incorrect attempt and then corrects themselves, this should be marked as correct as the pupil has shown the ability to decode. However, pupils should not be prompted to 'have another go'. If a pupil makes several attempts at a word, the final attempt should be scored, even if this is incorrect and a previous attempt had been correct.
- You should not indicate whether a pupil has decoded a word correctly or incorrectly during the administration of the screening check but you may offer encouragement or support to ensure they remain focussed on the task.
- Pupils should be given as long as necessary to respond to a word, although in most cases, 10 seconds should be sufficient. The administrator should decide when it is appropriate to tell the pupil to move onto the next word, taking care not to try to move the pupil on if they are still trying to decode the word. It is acceptable for pupils to sound out the phonemes, as long as the pupil is then able to blend the word.

2.6 Stopping the screening check before the end

Most pupils should be able to attempt all words in the screening check. However, if a pupil is struggling with the check, the administrator may decide

to stop the check before the end. When making the decision to stop the screening check, the administrator should ensure that the pupil has been given full opportunity to show what they can do.

3. Dealing with queries and issues during the screening check

3.1 Disruption during the screening check

It is impossible to plan for every scenario. Whatever action you take, pupil safety and well-being must always be your first consideration. If you need to stop the screening check for any reason, note where the pupil has reached so that you can restart the screening check at a later time at the appropriate place.

3.2 Damaged or spoiled marksheets

If a pre-printed pupil marksheet is damaged or spoiled during the screening check, schools should stop the check and telephone the Phonics pilot helpline on **020 7340 7494**. Schools may carry on administering the screening check to other pupils whilst this issue is sorted out.

4. What to do at the end of the screening check

4.1 Completed marksheets and overall attendance register

Make sure that all necessary information is captured on the marksheet and that it is clear and legible. On the pupil list, please use the following codes to confirm the status of each marksheet being returned. If blank marksheets have been used, please fill the pupil name in the appropriate box.

- ✓ Pupil present
- A Pupil absent
- D Pupil disappplied

- L Pupil left school
- X Marksheet damaged/spoiled
- B Blank marksheet not used

4.2 Returning screening check materials

All screening check materials should be returned by schools at the end of the administration week. This includes all marksheets (including unused and damaged/spoiled marksheets) and pupil materials. At your training meeting, you were informed of the day the check materials would be collected from you and materials need to be packaged and ready for collection on that day. Please complete the checklist and make sure all materials are secured in the return envelope(s) provided until collection has taken place. If collection needs to be delayed or collection was not attempted on the day notified, please contact the Phonics pilot helpline on **020 7340 7494**.

5. Monitoring visits

As with national curriculum tests, a number of monitoring visits will be carried out during the pilot to ensure that administration is consistent between schools. The purpose of these monitoring visits is to improve the quality of information on administration that is provided to schools in future by highlighting where instructions were unclear. Monitors will have a list of things to check related to storage of materials and school administration arrangements related to the information in this guide and the visit should normally last around 30 minutes.

Monitoring visits will be unannounced during the screening check administration week and carried out by representatives of Sheffield Hallam University who are conducting an independent evaluation of the pilot. The person carrying out the monitoring visits will have identification to confirm who they are. If a school has a concern about the validity of the visit, or wishes to confirm the identity of the visitor, they should contact the Phonics pilot helpline on 020 7340 7494.

Annex E – Agenda for training day

09.30	Arrivals: registration, tea and coffee available
10.00	Welcome and introduction
10.20	Placing the screening check in context: phonics and the importance of reading for pleasure
10.35	Reading: decoding and understanding
	Pseudo-words: why we are including some pseudo-words, explaining pseudo-words to children, how pseudo-words can be used in the classroom most usefully
11.15	Break: tea and coffee
11.30	Test administration information: <ul style="list-style-type: none"> • Structure of check • Recording sheets • Access arrangements for pupils with SEN • Administration arrangements • Scoring • Reporting
12.50	Lunch
13.35	Break out groups: marking and scoring <ul style="list-style-type: none"> • Using recordings of pupils attempting the screening check to explore how to score different responses
14.35	Input into policy development: <ul style="list-style-type: none"> • We would like to get schools' feedback on the policy, in particular the steps schools might take to support pupils who struggle with the screening check to make progress in future
15.10	Preparing to administer the screening check at school Next steps, including information about the evaluation of the pilot
15:45	Close

Assessment Delivery Service Requirements (ADSR) 2011 Year 1 Phonics Screening Check

Where indicated by its security classification above or below, this document includes confidential or commercially sensitive information and may not be disclosed in whole or in part, other than to the party or parties for whom it is intended, without the express written permission of an authorised representative of the Department for Education.

Table of contents

- 1. INTRODUCTION**
- 2. DOCUMENT SCOPE AND OBJECTIVES**
- 3. ROLES AND RESPONSIBILITIES**
 - i. PDS**
 - ii. DfE**
- 4. REQUIREMENTS**
 - i. SCHOOL SAMPLE CONSTRUCTION REQUIREMENTS**
 - ii. PUPIL DATA FEEDS REQUIREMENTS**
 - iii. OUTPUT DATA FEEDS REQUIREMENTS**
 - iv. ADDITIONAL REQUIREMENTS**
- 5. QUALITY ASSURANCE**
- 6. FILE SPECIFICATIONS**
 - i. SCHOOL LEVEL INPUT FILE (FROM DfE TO PDS)**
 - ii. INTER-RATER RELIABILITY STUDY MARKER
DETAILS (FROM DfE TO PDS)**
 - iii. PUPIL LEVEL INPUT FILE (FROM DfE TO PDS)**
 - iv. INTER-RATER RELIABILITY STUDY PUPIL LEVEL
INPUT FILE (FROM DfE TO PDS)**
 - v. WORDS INPUT FILE (FROM DfE TO PDS)**
 - vi. INPUT DATA TIMING**
 - vii. OUTPUT FILE (FROM PDS TO DfE)**
 - viii. OUTPUT DATA TIMING**
 - ix. MATCHED OUTPUT FILE (FROM DfE TO QCDA)**
 - x. MATCHED OUTPUT DATA TIMING**
- 7. TIMING**
- 8. OPEN ISSUES**
- 9. APPENDICES- MARKSHEETS**

1. Introduction

Since May 2010, the Government has stated its commitment to raising children's achievement in reading, and has expressed the intention to establish a phonics screening check for children in Year 1. This will be a short, light-touch screening check designed to confirm that children have grasped the basics of phonic decoding and to identify those pupils who need extra help at an early stage, so that schools can provide support. The results of the screening check will provide valuable information to parents. The screening check will be part of the arrangements for the statutory assessment of children in respect of the first Key Stage. It was announced on 22 November 2010 that the check would be trialled in summer 2011 and introduced as a statutory assessment in 2012.

In order for the Department for Education (DfE) to introduce the screening check as a statutory assessment in 2012, the summer 2011 trial is being used to develop the content of the screening check and to demonstrate reliability and validity. The screening check consists of a 40 word test that takes 5-10 minutes to complete. In future years, DfE will use the check to monitor the proportion achieving the standard nationally and at local authority level with all maintained schools with a Year 1 cohort completing the assessment for all pupils. For 2011, a representative sample of schools will trial the check which will be marked internally with item level data captured externally to enable appropriate analysis to be carried out. Data from the trial will not be published by DfE as the trial is being used to develop the test for future years.

A sample of 450 schools was asked to participate in the trial, with a second sample of 459 also being invited. Schools then expressed their interest in participating and a representative sample (by geographical region, school type and KS1 reading attainment) was selected to take part in the trial. Statistical experts advised that a sample size of 10,000 pupils was needed to provide robust item level data for test development. This means that 302 schools are participating in the 2011 trial.

Schools participating in the trial were notified of their inclusion on 25 and 28 March 2011. The 2011 Year 1 Phonics Screening Check trial will be held from Monday 13 June to Friday 17 June 2011, and schools are able to decide when to administer the check. For larger schools, this may need to be over more than 1 day.

2. Document Scope & Objectives

This document outlines the requirements for, and use of, data by the 2011 Year 1 Phonics Screening Check trial. The document focuses specifically on the data that shall be exchanged between the Department for Education (DfE) and the Print and Distribution suppliers, Publishing Delivery Service (PDS) for Department for Education, delivered in partnership with Prolog & Communisis. The Qualifications and Curriculum Development Agency (QCDA) will work with DfE to ensure that the data requirements are fit for purpose, ahead of undertaking detailed data analysis.

The document covers 2011 only. Later years are out of scope for this document. The document details

- the roles and responsibilities of the PDS in gathering and providing data so that it is timely and fit for the purposes of DfE;

- the role of DfE as supplier, and customer to PDS;
- the required information to be sent by DfE to the PDS to be used for identifying schools and pupils and the information sent by PDS to the DfE for statistical purposes;
- the specifications for DfE's requirements in relation to the screening check trial, outlining the dependencies on the timely provision and quality of data from DfE, and its providers, in addition to the provision and quality of data provided by PDS;
- the required format for the screening check pupil level datafeeds to be sent by PDS to DfE for QCDA to undertake data analysis.

Proposed changes to this document will be considered through the screening check steering group. All agreed changes must not affect the overriding requirement to deliver the necessary information to the PDS and/or DfE on time, and to the requisite quality standards.

3. Roles and Responsibilities

3.1 PDS

PDS is responsible for:

- Secure print and distribution of marksheets and test materials; and
- Establishing and maintaining appropriate and robust processes and data capture systems to record pupil responses to items in the 2011 Year 1 Phonics Screening Check trial.

PDS will

- ensure they deliver pupil data in a timely manner and to DfE requirements;
- produce and agree a high-level plan covering the trial print and distribution (including mark sheets and check materials) operation, showing the milestones impacting on key deliverables and review this regularly at KITs;
- quality assure pupil attainment data at a pupil level– via
 - identifying and checking school data to ensure that the right test paper and spare test papers go to the correct school;
 - ensuring data collected from schools on the check sheets are valid and querying any abnormalities, such as significant differences in the expected number of pupils per school or non-collection of some materials; and
 - independently conducting checks of the component test data obtained from the OMR process using their own validation processes.

3.2 DfE

DfE is responsible for establishing and maintaining appropriate and robust systems for providing data about participating schools, and agreeing data requirements with QCDA. In this DfE's Data and Statistics Division will:

- provide a datafile containing pupil details, school name and test allocation to aid with the printing of test mark sheets and identification of pupils in sample schools;

- provide a datafile of test items per version of the check, in order for test materials and OMR marksheets with words of each specific version included, to be printed by PDS.
- provide access to any change to school addresses, via Edubase;
- specify, as clearly as possible, the nature, timing, coverage, quality and end use of the pupil data required;
- provide a partner role in advising and commenting on appropriate management and quality assurance measures to be employed by PDS;
- share how it conducts its own quality assurance as the end customer, and provide feedback to PDS, as appropriate;
- quality assure all datafeeds provided to PDS to ensure that they meet the agreed specifications and timings; and
- confirm acceptance (or otherwise) that datafeeds provided by PDS meet requirements.

4. Requirements

In order for DfE to develop the national Year 1 Phonics Screening Check for 2012, via a sufficient number of schools administering one of twenty forms of the trial screening check to their year one pupils in 2011, it is essential that DfE manages accurate and up to date information on schools, pupils, which version of the check they are sitting, if they are participating in validity and reliability studies, and PDS provides accurate pupil item level results. To achieve this, DfE and PDS are reliant on each other to provide various data sources, and for each party to manage an update this data and associated information.

The exact requirements on each party for end to end operations are detailed below, from the data needed to identify participating pupils and schools, through to that required for the psychometric analysis of results. The key vehicle for providing pupil level information will be via a Phonics Check pupil summary file (PUPSU). The Phonics Check PUPSU file covers DfE requirements for Year 1 Phonics Screening Check outcomes. Full detail of the Phonics Check PUPSU, and timings of datafeeds, are included in sections 6 and 7.

Please note that the focus of this section is on data requirements between PDS and DfE.

Please also note that other integral parts of the programme, such as communications to schools about the phonics screening check trial, are also outside the scope of this document. Detail of all activity, including communications, is available in the programme plan.

4.1 School sample construction requirements

DfE has constructed a representative sample as detailed in a separate paper on the recruitment of screening check pilot schools. DfE agreed that the sample should be

representative by geographical region, school type and Key Stage 1 reading attainment, with schools being stratified into 5 reading attainment bands.

4.2 Pupil data feeds requirements

The pupils required to complete the screening check in the trial are those at the end of Year 1 in summer 2011. DfE and PDS require information on pupils in advance of the trial administration week (13-17 June 2011) in order to build the required data systems and prepare schools. Please note that in this

- the number of mark sheets (see Annex A for mark sheets) sent to schools will be based on each schools year 1 cohort, as captured through the January 2011 school census;
- modified mark sheets will not be required as these are completed by teachers. For pupils requiring modified test materials, teachers will be able to request an electronic version and make necessary modifications in school; and
- PDS will capture the number of pupils not completing the test and reasons via the OMR process.

Data requirement

DfE will provide PDS with a list of Y1 pupils participating in the trial, their schools and which test they will be sitting, and whether they are participating in any validity and reliability studies, using the January 2011 School Census returns.

To mitigate the impact of changes to pupil numbers, blank test sheets (also at Annex A) pre-allocated to a particular test will be provided. Schools will be asked to provide information on changes to pupil numbers on the cover sheet enclosed with the test materials and mark sheets, which is then returned with the completed mark sheets and test materials. This will be in the format of a Phonics Check PUPSU file covering all participating schools. For the Phonics Check PUPSU file format see Section 6.

4.3 Output data feeds requirements

In order to manage information on the Year 1 Phonics Screening Check, the DfE requires detailed information on performance, at pupil level. The main output required by DfE is a datafeed containing pupils item level performance and inter-rater study item level marks on 4 July 2011. This is generated from the OMR process carried out by PDS. For the format of the data feed, please refer to section 6. For the timing of all datafeeds, please refer to section 7.

4.4 Additional requirements

In addition to the above data requirements, PDS is obliged to following certain DfE protocols. These are set out below.

4.41 Reducing Burdens

The Year 1 Phonics Screening Check trial data collection arrangements must take full account of the need to reduce burdens on schools. Arrangements must also account for the Data Sharing Protocol, subscribed to by DfE. Approval for school data collections rests with the Star Chamber External Scrutiny Board and approval for this collection was given by the Star Chamber Scrutiny Board at their meeting in January 2011.

The Year 1 Phonics Screening Check trial is sensitive to the needs of schools, including issues such as the responsiveness of communications, the timing of requests for information, volume of information requested and support arrangements provided, and the streamlining of collection systems. DfE will seek and respond to the views of schools in the delivery of the Year 1 Phonics Screening Check trial through separate monitoring and evaluation work.

4.42 Use of data for publication

DfE has to comply with protocols established by the UK Statistics Authority in providing evidence of the robustness of the process for producing and quality assuring results. The level of accuracy and confidence that can be placed in the data also determines its fitness for public accountability purposes. However the data will not be published for this year, as it is for test development purposes only. The provision, retention and use of data are governed by the provisions of the Data Protection Act 1998. DfE and PDS will ensure that it, and its contractors, comply with the requirements of this Act. All performance data prior to publication must be classified as 'restricted statistics' and accorded the appropriate level of security, as per the succeeding requirement.

4.43 Control of data files for collection of all data required

Please note that the Secretary of State for Education owns the data and retains the right to control access to datasets and is responsible for their final acceptance for use in publications and by third parties. Prior to publication all data outputs, no matter the coverage or quality, must be classified as 'RESTRICTED STATISTICS' and accorded the appropriate level of physical security. Following publication no results data or supporting documentation will be released to third parties without DfE's agreement to a quality specification and approval to release the data. The purpose of this control is to ensure that the origin of datasets is traceable and that only data of a known quality is released. The release of data will be undertaken in accordance with the Code of Practice for Official Statistics and be controlled by Jude Hillary, Head of Profession for Statistics in DfE.

5. Quality Assurance

PDS will undertake the quality assurance including, but not limited to:

- regular and thorough quality control checks of its work, and that of its contractors (if applicable), in relation to printing, distribution, data collection processes and key deliverables;
- pre-testing of the system to give assurance that the system will work, prior to live running.

6 File specifications

This section details the format of the data files necessary to deliver the output datafeeds.

6.1 School level Input file (From DfE to PDS)

DfE will provide a school level file input data file containing addressing details of participating schools.

6.1.1 Composition

This is taken from Edubase, a database that provides data for all Departmental mailings. The file will be sorted in ascending order of the Unique Reference Number.

6.1.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exception is the Establishment number, which must be four characters and therefore should be left-padded with zeros if less than 1000. The data should be ANSI encoded.

The files should be named: YR1 _PSC_SCH _<ccyymmdd>_<nnn>.TXT where <ccyymmdd> is the date the file was produced and the <nnn> is a sequence where multiple versions of a file are created on any given date. Normally sequence will be '001'.

The table below shows the required Edubase fields:

Field Name	Database Reference	Field Type	Field Length	Look-up Table	Change Date	Change Frequency	Description (where not clear from field name)
Unique Reference Number	URN	Number	6			Ongoing	
LA Number	LA	Number	3	Y		Ongoing	
LA Name	LANAME	Text		Y		Ongoing	
DfE Establishment Number	ESTAB	Number	4			Ongoing	
First name	HEADFIRSTNAME	Text	35			Ongoing	
Lastname	HEADLASTNAME	Text	35			Ongoing	
Organisation Name	SCHOOLNAME	Text	100			Ongoing	
House Number and street name	STREET	Text	100			Ongoing	
Address line 2	LOCALITY	Text	40			Ongoing	

Field Name	Database Reference	Field Type	Field Length	Look-up Table	Change Date	Change Frequency	Description (where not clear from field name)
Address line 3	ADDRESS_3	Text	40			Ongoing	
Town	TOWN	Text	30			Ongoing	
County	COUNTY	Text	30			Ongoing	
Postcode	POSTCODE	Text	8			Ongoing	
Establishment Type	ToE_DESC	Text	50			Ongoing	

6.2 Inter-rater Reliability Study Marker details (From DfE to PDS)

DfE will provide an input file containing school addressing details of Teachers/Markers in the inter-rater reliability study.

6.2.1 Composition

The file should contain 20 records for the 20 teachers/markers in the study. Address details should be taken from Edubase, a database that provides data for all departmental mailings. The file will be sorted in ascending order of the Unique Reference Number. The Type of Establishment field is included to preserve the addressing file format and should not contain any values

6.2.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exception is the Establishment number, which must be four characters and therefore should be left-padded with zeros if less than 1000. The data should be ANSI encoded.

The files should be named: YR1_IRS_SCH_<ccyymmdd>_<nnn>.TXT where <ccyymmdd> is the date the file was produced and the <nnn> is a sequence where multiple versions of a file are created on any given date. Normally sequence will be '001'.

The table below shows the required fields:

Field Name	Database Reference	Field Type	Field Length	Look-up Table	Change Date	Change Frequency	Description (where not clear from field name)
Unique Reference Number	URN	Number	6			Ongoing	
LA Number	LA	Number	3	Y		Ongoing	
LA Name	LANAME	Text		Y		Ongoing	

Field Name	Database Reference	Field Type	Field Length	Look-up Table	Change Date	Change Frequency	Description (where not clear from field name)
DfE Establishment Number	ESTAB	Number	4			Ongoing	
First name		Text	20				First Name of the teacher who will score the check in the inter-rater reliability study.
Last name		<BLANK >					Last Name of the teacher who will score the check in the inter-rater reliability study.
Organisation Name	SCHOOLNAME	Text	100			Ongoing	
House Number and street name	STREET	Text	100			Ongoing	
Address line 2	LOCALITY	Text	40			Ongoing	
Address line 3	ADDRESS_3	Text	40			Ongoing	
Town	TOWN	Text	30			Ongoing	
County	COUNTY	Text	30			Ongoing	
Postcode	POSTCODE	Text	8			Ongoing	
Establishment Type	ToE_DESC	Text	50			Ongoing	<LEAVE FIELD BLANK>

6.3 Pupil Level Input file (from DfE to PDS)

DfE will provide an input data file containing information on Year 1 pupils participating in the trial, which test they will be sitting, and whether they are participating in any validity and reliability studies. This information should be used for printing individual pupil test forms. See Section 9 - Appendices

6.3.1 Composition

The pupil results file contains information on all Year 1 pupils taking part in the trial. Pupil records will be identified within a school by the DfEReference number, a unique identifier made up of the test form number and a unique pupil number. The file should be sorted into ascending order on CurrDfEno and the DfEReference.

6.3.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exceptions are the Establishment number (field 2), which must be four characters and therefore should be left-padded with zeros if less than 1000; the Test/Test re-test form number which should be left padded by zero if less than 10; and The DfE reference (see business rules below) . The data should be ANSI encoded. The file should be named: YR1 _PSC_PUP _<ccyymmdd>_TXT where <ccyymmdd> is the date the file was produced .

Field Name	Data Type	Description	Use	Allowable characters	Nulls allowed?	Business Rules	Max Field Length
LA	Integer	The Local Authority (LA) three digit code identifies a particular LA	Identifies the LA the pupil is associated with.	0-9	N	Must be a valid LA code	3
Estab	Integer	The DfE Establishment number is a four digit reference number allocated to each school	Unique school identification number within an LA.	0-9	N	Must be a four digit number; left-padded with zeros if < 1000	4
URN	Integer	The Unique Reference Number, a six digit reference number allocated to each school	Unique school identifier	0-9	N	Must be a six digit number	6
Surname	Text	Full legal surname (derived from family, clan or marital association) of the child (as written) This Field must be left blank for blank forms only ----- OR – If pupil is in the inter-rater reliability study: The code allocated to the teacher who will score the check in the inter-rater reliability study. This will help capture the scores of pupils in the study without identifying the pupil	Provides the surname of the pupil. ----- OR - If pupil is in the inter-rater reliability study: Links scorer in the study to the check	Any alphanumeric	Y	None applied ----- OR - If pupil is in the inter-rater reliability study: Teacher 01 -Teacher 20	35
Forename	Text	Full forename of the child (not common contractions) This Field must be left blank for blank forms only ----- OR - If pupil is in the inter-rater reliability study: The unique pupil identifier assigned by	Provides the forename of the pupil. ----- OR - If pupil is in the inter-rater reliability study: Links pupil in the study to the main data file	Any alphanumeric	Y	None applied.	35

Field Name	Data Type	Description	Use	Allowable characters	Nulls allowed?	Business Rules	Max Field Length
		DfE to the pupil in the main study					
Middlenames	Text	In full, not shortened or familiar versions. If pupil has no middle name(s) then this field must be left blank	Provides the middle name(s) of the pupil.	Any alphanumeric	Y	None applied.	35
Test Form Number	Integer	Identifier for the test form taken by pupil	Links pupil to test form taken	01-20	N	Number must be in range 01-20	2
DfEReference	Text	Unique pupil identifier assigned by DfE	For future matching activities	0-9	N	Seven digit number in the range 0100001 - 2015000	7
Test re-test indicator	Integer	Identifier for whether the pupil is participating in the test re-test study 1 = Participating 0 = non-participating	Indicates whether pupil is participating in the test re-test study	0-1	N	Number must be in range 0-1	1
Test re-test form number	Integer	Identifier for the test re-test form taken by pupil. If the pupil is not taking part in the test-re-test study, this field should be left blank	Links pupil to the test form taken for the test re-test study	01-20	Y	Number must be in range 01-20	2
Inter-rater reliability indicator	Integer	Identifier for whether the pupil is part of the inter-rater reliability study 1 = Participating 0 = non-participating	Indicates if the pupil is part of the inter-rater reliability study	0-1	N	Number must be in range 0-1	1
Blank Test Form Indicator	Integer	1 = Blank Test Form 0 = Form allocated to specific pupil	For printing blank forms with DfEReference # and Test Form number only (No names included)	0-1	N	Number must be in range 0-1	1

6.4 Inter-rater Reliability Study Pupil Level Input File (from DfE to PDS)

DfE will provide the Inter-rater study input data file containing pupils in the inter-rater study this will be used.

6.4.1 Composition

The Inter-rater Study file contains information on pupils taking part in the study. Pupil records will be identified within a school by the DfEReference number, linked to the reference teacher/marker identifier.

6.4.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exceptions are the Establishment number (field 2), which must be four characters and therefore should be left-padded with zeros if less than 1000; the Test/Test re-test form number which should be left padded by zero if less than 10; and The DfE reference (see business rules below) .The data should be ANSI encoded. The file should be named: YR1 _IRS_PUP _<ccyymmdd>_TXT where <ccyymmdd> is the date the file was produced .

Field Name	Data Type	Description	Use	Allowable characters	Nulls allowed?	Business Rules	Max Field Length
LA	Integer	The Local Authority (LA) three digit code identifies a particular LA	Identifies the LA the pupil is associated with.	0-9	N	Must be a valid LA code	3
Estab	Integer	The DfE Establishment number is a four digit reference number allocated to each school	Unique school identification number within an LA.	0-9	N	Must be a four digit number; left-padded with zeros if < 1000	4
URN	Integer	The Unique Reference Number, a six digit reference number allocated to each school	Unique school identifier	0-9	N	Must be a six digit number	6
Surname	Text	The code allocated to the teacher who will score the check in the inter-rater reliability study. This will help capture the scores of pupils in the study without identifying the pupil	Links scorer in the study to the check	Any alphanumeric	N	Teacher 01 -Teacher 20	35
Forename	Text	The unique pupil identifier assigned by DfE to the pupil in the main stud	Links pupil in the study to the main data fil	Any alphanumeric	N	None applied.	35
Middlenames	Text	In full, not shortened or familiar versions. If pupil has no middle name(s) then this field must be left blank	Provides the middle name(s) of the pupil.	Any alphanumeric	Y	None applied.	35
DfEReference	Text	Unique pupil identifier assigned by DfE	For future matching activities	0-9	N	Seven digit number in the range 0100001 - 2015000	7
Test Form Number	Integer	Identifier for the test form taken by pupil	Links pupil to test form taken	01-20	N	Number must be in range 01-20	2

6.5 Words Input file (from DfE to PDS)

DfE will provide a file containing details of the words to be used in the screening check, linked to the test forms on which the words would be printed. PDS will then use these to print the marksheets and test materials.

6.5.1 Composition

The Words file would contain information on all the words to be tested in the screening check. Words would be linked to the test forms on which they would be printed. There would be 20 records in all, each containing 40 words.

Each form would be identifiable by a test form number and within the form, words would be identified by a word number between 1 and 40. The file would be sorted into ascending order on the test form identifier.

6.5.2 Format

The data file will be tab delimited and should be named: YR1 _PSC_WRD _<ccyymmdd>_TXT where <ccyymmdd> is the date the file was produced.

Field No.	Field Name	Length	NULL allowed?	Allowable characters or range	Field Description
1	TestFormNo	2	N	1-20	Unique test form identifier
2	W1		N		Phonics Screening Check word #1
.....	Repeat for each item		N		Phonics Screening Check word
40	W40		N		Phonics Screening Check word #40

6.6 Input data timing

DfE will provide the pupil details data file to PDS by 18 April 2011. This will include details for all pupils in Year group 1 pupils in the January 2011 census for participating schools, approximately 15,000 marksheets. At the same time, the data file containing words per version of the check for test materials and inclusion on marksheets will be provided.

DfE will provide the inter-rater study data file to PDS by 16 May 2011. This will include details for teachers and the pupils they are rating in the inter-rater study, following administration of the check in June 2011. Approx 650 marksheets will then need to be printed.

6.7 Output file(from PDS to DfE)

This is the results file containing item level data for the Year 1 Phonics Screening Check which will be provided by PDS

6.7.1 Composition

The pupil results file contains results data with pupil identification information.

Pupil records will be identified within a school by the DfEReference number, a unique identifier made up of the test form number and a unique pupil number. The file should be sorted into ascending order on CurrDfENo and the DfEReference.

6.7.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exceptions are the Establishment number (field 2), which must be four characters and therefore should be left-padded with zeros if less than 1000; the Test/Test re-test form number which should be left padded by zero if less than 10; and The DfE reference (see business rules below). The data should be ANSI encoded.

The files should be named YR1 _PSC_ITEM _<ccyymmdd>_<nnn>.TXT where <ccyymmdd> is the date the file was produced and the <nnn> is a sequence where multiple versions of a file are created on any given date. Normally sequence will be '001'.

The file will comprise 52 fields which may be considered, broadly, to fall into four parts:

- Part 1 – fields #1 to #4 contain school-level information
- Part 2 – fields #5 to #11 contain pupil level information
- Part 3 – fields #12 to #55 contain test specific information
- Part 4 – fields #56 to #98 contain test-retest specific information

The table below shows the result file as required from PDS

Field No.	Field Name	NULL allowed?	Allowable characters or range	Field Description
1	CurrLA	N	Three digit number in range: 201 to 938	Latest LA number of the school
2	CurrESTAB	N	Four digit number in range: 0001 to 9999	Latest establishment number of the school
3	CurrDfENo	N	Seven digit number in range: 2010001 to 9389999	Latest DfE Number of the school (concatenation of fields 2 and 3,
4	URN	N	Six digit reference number	Unique reference number of the school.
5	DfEReference	N	Seven digit number in the range 0100001 - 2015000	The unique pupil identifier assigned by DfE to the pupil in the main study

Field No.	Field Name	NULL allowed?	Allowable characters or range	Field Description
6	Surname ----- OR – If pupil is in the inter-rater reliability study: The code allocated to the teacher who will score the check in the inter-rater reliability study	N	Up to 35 characters, allowable are: A to Z, -, (,), ', 0-9 All leading and trailing spaces should be removed; only one space should appear between multiple surnames.	Surname of Pupil ----- OR – If pupil is in the inter-rater reliability study: The code allocated to the teacher who will score the check in the inter-rater reliability study
7	Middlenames	Y	Up to 35 characters, allowable are: A to Z, -, (,), ', 0-9 All leading and trailing spaces should be removed; only one space should appear between multiple middlenames.	Full Middlename(s) of pupil If pupil has no middle name(s) then this field must be left blank
8	Forename ----- OR – If pupil is in the inter-rater reliability study: The unique pupil identifier assigned by DfE to the pupil in the main study	N	Up to 35 characters, allowable are: A to Z, -, (,), ', 0-9 All leading and trailing spaces should be removed; only one space should appear between multiple forenames.	Full forename(s) of pupil ----- OR – If pupil is in the inter-rater reliability study: The unique pupil identifier assigned by DfE to the pupil in the main study
9	Month of Birth	Y	1-12	Month part of the Date of Birth of pupil.
10	Year of Birth	Y	Must be a 4 digit calendar year	Year part of the Date of Birth of pupil.
11	Gender	Y	One digit character from: M,F,?	Gender of pupil: M = Male, F = Female, ? = Unknown
12	Test FormNo	N	Two digit number between 01 and 20	Test Form identifier
13	Test Day	Y	One digit number between 1 and 5	1= Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5= Friday If the pupil did not complete the test, this field should be left blank
14	Test Status	Y	A, L, D	If pupil has completed the test, this field should be left blank Reason for non-completion of test A: If a pupil is absent L: if a pupil has left the school D: If a pupil is disapplied from the tests
15	PSC_W1	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
.....	Repeat for each item to end of paper	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present

Field No.	Field Name	NULL allowed?	Allowable characters or range	Field Description
54	PSC_W40	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
55	PSC_TotMark	Y	0-40	Total score out of 40. Indicates how many correct responses out of 40
56	ReTest FormNo	N	Two digit number between 01 and 20	Test Form identifier If the pupil did not participate in the test re-test study, this field should be left blank
57	RetestDay	Y	One digit number between 1 and 5	1= Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5= Friday If the pupil did not participate in the test re-test study, this field should be left blank
58	PSC_TS2_W1	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
.....	Repeat for each item to end of paper	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
97	PSC_TS2_W40	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
98	PSC_TS2_TotMark	Y	0-40	Total score out of 40. Indicates how many correct responses out of 40

6.8 Output data timing

PDS will provide a single pupil results and inter-rater study results data file to DfE on 4 July 2011.

6.9 Matched Output File (from DfE to QCDA)

This is the results file containing item level data for the Year 1 Phonics Screening Check matched to pupil characteristics, which will be provided to QCDA

6.9.1 Composition

The pupil results file contains results data with pupil identification and contextual information.

Pupil records will be identified within a school by the DfE Reference number, a unique identifier made up of the test form number and a unique pupil number. The file should be sorted into ascending order on CurrDfENo and the DfE Reference.

6.9.2 Format

The data file will be tab delimited and fields must not be padded with spaces or zeros. The exceptions are the Establishment number (field 2), which must be four characters and therefore should be left-padded with zeros if less than 1000; the Test/Test re-test form number which should be left padded by zero if less than 10; and The DfE reference (see business rules below). The data should be ANSI encoded.

The files should be named YR1 _PSC_MCH_ITEM _<ccyymmdd>_<nnn>.TXT where <ccyymmdd> is the date the file was produced and the <nnn> is a sequence where multiple versions of a file are created on any given date. Normally sequence will be '001'.

The file will comprise 55 fields which may be considered, broadly, to fall into four parts:

- Part 1 – fields #1 to #4 contain school-level information
- Part 2 – fields #5 to #15 contain pupil level information
- Part 3 – fields #16 to #59 contain test specific information
- Part 4 – fields #60 to #102 contain retest specific information

The table below shows the result file as required from DfE

Field No.	Field Name	NULL allowed?	Allowable characters or range	Field Description
1	CurrLA	N	Three digit number in range: 201 to 938	Latest LA number of the school
2	CurrESTAB	N	Four digit number in range: 0001 to 9999	Latest establishment number of the school
3	CurrDfENo	N	Seven digit number in range: 2010001 to 9389999	Latest DfE Number of the school (concatenation of fields 2 and 3,
4	URN	N	Six digit reference number	Unique reference number of the school.
5	DfEReference	N	Seven digit number in the range 0100001 - 2015000	The unique pupil identifier assigned by DfE to the pupil in the main study
6	Month of Birth	N	1-12	Month part of the Date of Birth of pupil.
7	Year of Birth	N	Must be a 4 digit calendar year	Year part of the Date of Birth of pupil.
8	Gender	N	M,F	Gender of pupil: M = Male, F = Female, ? = Unknown
9	EAL Status	Y	Three character Alphanumeric	Indicates the whether pupil's first language is English or other than English
10	FSMEligible	Y	0,1	Indicate whether pupil is eligible for Free School Meals 1: if pupil is eligible 0: if pupil is not eligible
11	EthnicGroup	Y	Four character AlphaNumeric	Codeset available on request
12	SEN status	Y	N,A,P,S	N = No Special Educational Need A = School Action P = School Action Plus S = Statemented
13	Primary SEN type	Y	TBC	TBC

Field No.	Field Name	NULL allowed?	Allowable characters or range	Field Description
14	Test-re-test indicator	Y	1 = Participating 0 = non-participating	Identifier for whether the pupil is participating in the Inter-rater reliability study
15	Inter-rater reliability indicator	Y	1 = Participating 0 = non-participating	Identifier for whether the pupil is participating in the test re-test study
16	Test FormNo	N	Two digit number between 01 and 20	Test Form identifier
17	TestDay	N	One digit number between 1 and 5	1= Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday
18	TestStatus	Y	A, L, D	Reason for non-completion of test A: If a pupil is absent L: if a pupil has left the school D: If a pupil is disapplied from the tests
19	PSC_W1	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
.....	Repeat for each item to end of paper	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
58	PSC_W40	N	NULL if pupil did not sit the test, 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
59	TotMark	Y	0-40	Total score out of 40. Indicates how many correct responses out of 40
60	ReTest FormNo	N	Two digit number between 01 and 20	Test Form identifier
61	RetestDay	Y	One digit number between 1 and 5	1= Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday
62	PSC_TS2_W1	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
.....	Repeat for each item to end of paper	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
101	PSC_TS2_W40	N	NULL if pupil did not take part in the test-retest study 1 = if the pupil answered correctly 0 = if the pupil answered incorrectly '-' = not attempted	Indicates whether OMR sheet response for this item is either correct, not correct, or not present
102	PSC_TS2_TotMark	Y	0-40	Total score out of 40. Indicates how many correct responses out of 40

6.10 Matched Output data timing

DfE will provide a single results data file to QCDA on **6 July 2011 including both pupil results and inter-rater study results, and an updated version on xx xxx 2011 with matched data.**

7 Timing

#	Date	Nature of datafeed	Detail
1	18 April 2011	School level input file	DfE to PDS: Addressing details of participation schools
2	16 May 2011	Inter-rater study marker details	DfE to PDS : Addressing details of Teachers/Markers taking part in the Inter-rater Reliability study
3	18 April 2011	Pupil level input file	DfE to PDS : Details of pupils in the trial, for printing of Test Forms
4	16 May 2011	Inter-rater Study pupil details	DfE to PDS: Details of pupils taking part in the Inter-rater Reliability study
5	4 July 2011	Output file	PDS to DfE : Item level results file
6	6 July 2011	Matched Output Data file	DfE to QCDA: Matched results file

8 Open Issues

#	Open Items	Description	Estimated date for issue resolution

Appendix A: Sample Test Form

[TestFormNumber][DfEReference]

Year 1 Phonics Screening Check – Answer Sheet

Pupil information

[Surname] [Middlenames] [Forename]

Day the test was completed:

Monday	Tuesday	Wednesday	Thursday	Friday
--------	---------	-----------	----------	--------

If not completed, reason for non-completion:

Left School	Absent	Disapplied
-------------	--------	------------

Screening check responses

Please tick the appropriate box for each word

Section 1		
Word	Correct	Incorrect
Word 1		
Word 2		
Word 3		
Word 4		
Word 5		
Word 6		
Word 7		
Word 8		
Word 9		
Word 10		
Word 11		
Word 12		
Word 13		
Word 14		
Word 15		
Word 16		
Word 17		
Word 18		
Word 19		
Word 20		

Section 2		
Word	Correct	Incorrect
Word 21		
Word 22		
Word 23		
Word 24		
Word 25		
Word 26		
Word 27		
Word 28		
Word 29		
Word 30		
Word 31		
Word 32		
Word 33		
Word 34		
Word 35		
Word 36		
Word 37		
Word 38		
Word 39		
Word 40		

Year 1 Phonics Screening Check – Answer Sheet

Pupil information

Full pupil name _____

Gender

<input type="checkbox"/>	Female	<input type="checkbox"/>	Male
--------------------------	--------	--------------------------	------

Date of Birth

M	M	Y	Y	Y	Y
---	---	---	---	---	---

Day the test was completed:

<input type="checkbox"/>	Monday	<input type="checkbox"/>	Tuesday	<input type="checkbox"/>	Wednesday	<input type="checkbox"/>	Thursday	<input type="checkbox"/>	Friday
--------------------------	--------	--------------------------	---------	--------------------------	-----------	--------------------------	----------	--------------------------	--------

Screening check responses

Please tick the appropriate box for each word

Section 1		
Word	Correct	Incorrect
Word 1		
Word 2		
Word 3		
Word 4		
Word 5		
Word 6		
Word 7		
Word 8		
Word 9		
Word 10		
Word 11		
Word 12		
Word 13		
Word 14		
Word 15		
Word 16		
Word 17		
Word 18		
Word 19		
Word 20		

Section 2		
Word	Correct	Incorrect
Word 21		
Word 22		
Word 23		
Word 24		
Word 25		
Word 26		
Word 27		
Word 28		
Word 29		
Word 30		
Word 31		
Word 32		
Word 33		
Word 34		
Word 35		
Word 36		
Word 37		
Word 38		
Word 39		
Word 40		

Annex G – Agenda for standard setting meeting

09.00	Arrivals: registration, tea and coffee available
09.30	Introduction overview of process
10.00	Performance level descriptor
11.15	Break
11.30	Introduction to standard setting procedures and practice
12.25	Complete baseline evaluation
12.30	Lunch (review evaluation forms)
13.15	Resolve any outstanding issues Lunch
13.25	Round 1 - individual ratings
14.10	Break (collating outcomes)
14.25	Discussion
14.45	Round 2 - small group ratings
15.30	Break (collating outcomes)
15.45	Discussion
16.05	Round 3 - whole group finalised rating (including review of impact data)
16.50	Complete final evaluation
17.00	Close

Annex H – Check-re-check guide

1. The purpose of the study

One of the concerns raised about introducing a check for pupils at the end of Year 1 is that pupils may perform differently on different days depending on a number of factors. This study will attempt to determine the extent to which this is true and the likely impact on outcomes for schools.

As each form of the screening check in the pilot is of different levels of difficulty, you should not expect pupils to score exactly the same on each form that you administer to them.

2. How to plan for the repeat screening check

2.1 Preparing yourself

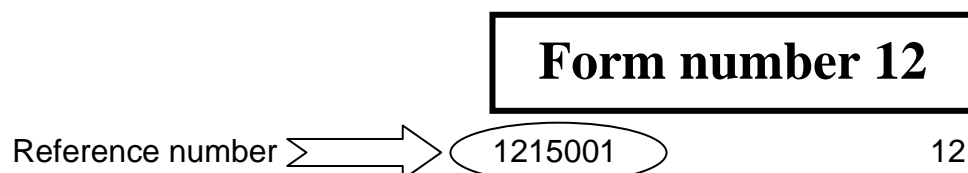
You should familiarise yourself with all of the information in the *Administration guide* as it contains all of the information you need to know to administer the screening check.

2.2 Receipt and storage of screening check materials

In addition to the main materials in your consignment you will also receive a second set of pupil marksheets for use in this study. These will be separated from the main pilot materials by a sheet of paper which says 'Repeat of the screening check – validity study'.

You should check the content of the pack carefully and telephone the Phonics pilot helpline immediately on **020 7340 7494** if you have anything missing.

Each pupil in your school has been pre-assigned two versions of the screening check and two pre-printed marksheets have been provided for each pupil. In total, no more than 5 versions have been assigned to your school. Pupil details were taken from the January 2011 School Census and therefore if a pupil arrived in school after the census date, they will not have a pre-printed form. Schools have been provided with 5 blank forms for pupils who arrived in school after the School Census was completed which correspond to 5 different blank forms for the repeat screening check validity study. If you use the blank marksheets you must make sure that you assign the correct second version to the correct pupil. To do this, make sure the second version administered to the pupil has the same reference number as the first version. The reference number can be found in the top right hand corner of the marksheet under the form number box.



If there are insufficient blank forms for the number of newly arrived pupils, schools should telephone the Phonics pilot helpline immediately on **020 7340 7494**.

All screening check materials should be stored securely whilst they are in school.

2.3 When to administer the repeat screening check

The repeat screening check should be administered in the same week as the main pilot (13 – 17 June) but on a different day. Where possible, schools should administer the first form of the screening check to all pupils before administering the second form. However, if pupils are absent, this may not be possible.

Since the words being trialled during the pilot each appear on two forms of the screening check, it is possible that some pupils will be asked to read the same words on both occasions. You should not draw pupils' attention to this, although if they notice and comment you may acknowledge that this is the case.

2.4 Access arrangements

You should use the same access arrangements for the repeat screening check as were used for the main pilot. If you have requested a Braille or electronic version for a pupil, you should make sure that you receive these for both forms.

3. How to administer the repeat screening check

You should administer the repeat screening check in exactly the same way as the main pilot using the information in the *Administration guide*.

3.1 What to say at the beginning of the repeat screening check

You should follow the instructions provided in the *Administration guide*. However, you may wish to add the following sentence to the instructions to pupils:

Today we are going to do the same activity that we did earlier in the week with some different words.

4. What to do at the end of the repeat screening check

You should pack all materials from the repeat screening check with those from the main pilot. Marksheets from the main study should be grouped together in register order followed by the marksheets from the repeat screening check validity study.

References

- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Cizek, Gregory J. and Bunch, Michael B. (2009). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*
- Coldwell, M., Shipton, L., Stevens, A., Stiell, B., Willis, B. and Wolstenholme, C. Centre for Education and Inclusion Research Sheffield Hallam University (DfE 2011). *Process evaluation for the Year 1 phonics screening check pilot*
- Davis, C. (2005). *N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics*
- Department for Education (2011). *DfE: School Workforce in England, November 2010 (Provisional)*
- Department for Education (2011). *Response to public consultation on the Year 1 phonics screening check*
- Department for Education (2011). *Year 1 phonics screening check: Framework for pilot in 2011*
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions (2nd ed.)*. New York: John Wiley.
- Medler, D.A. and Binder, J.R. (2005). *MCWord: An On-Line Orthographic Database of the English Language*. <http://www.neuro.mcw.edu/mcword/>
- Ofqual (2011). *Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage*
- Stuart, Masterson, Dixon & Quinlan (1993-1996). *Children's Printed Word Database*
- Yen, W. and Fitzpatrick, A.R. (2006). *Item Response Theory*. In *Educational Measurement, Fourth Edition*, Brennan (Editor)

© Crown copyright 2012

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or e-mail: psi@nationalarchives.gsi.gov.uk.

This document/publication is also available on our website at www.education.gov.uk

STA/12/5791

Standards and Testing Agency

An executive agency of the
Department for Education