Validity and sensitivity of the phonics screening check: implications for practice

Fiona J. Duff

Department of Experimental Psychology, University of Oxford, Oxford, UK

Silvana E. Mengoni

Centre for Education and Educational Technology, Open University, Milton Keynes, UK

Alison M. Bailey

City of York Local Authority, York, UK and Department of Psychology, University of York, York, UK

Margaret J. Snowling

St. John's College, University of Oxford, Oxford, UK

Background: Introduced in June 2012, the phonics screening check aims to assess whether 6-year-old children are meeting an appropriate standard in phonic decoding and to identify children struggling with phonic skills.

Aims: We investigated whether the check is a valid measure of phonic skill and is sensitive in identifying children at risk of reading difficulties.

Sample: We obtained teacher assessments of phonic skills for 292 six-year-old children and additional psychometric data for 160 of these children.

Methods: Teacher assessment data were accessed from schools via the local authority; psychometric tests were administered by researchers shortly after the phonics screening check.

Results: The check was strongly correlated with other literacy skills and was sensitive in identifying at-risk readers. So too were teacher judgements of phonics.

Conclusions: Although the check fulfils its aims, we argue that resources might be better focused on training and supporting teachers in their ongoing monitoring of phonics.

To become literate in a language such as English, it is necessary to crack the alphabetic code – that is, to understand that certain phonemes (sounds) in speech are represented by graphemes (letters) in writing (Byrne, 1998; Byrne & Fielding-Barnsley, 1989). Phonics instruction involves teaching this alphabetic principle and applying it to reading and spelling (Ehri et al., 2001). There is good evidence to support the efficacy of teaching phonics as a means of teaching reading (McArthur et al., 2012; National Reading Panel, 2000; Torgerson, Brooks & Hall, 2006). However, it is important to acknowledge that phonics

instruction is more effective when delivered in the context of a broader literacy curriculum, rather than in isolation (Camilli, Vargas & Yurecko, 2003; Stuebing, Barth, Cirino, Francis & Fletcher, 2008). These research findings are reflected in the current practice of most schools in England, following the *Independent Review of the Teaching of Early Reading* (Rose, 2006). Amongst the review's key recommendations was that phonics should be taught as the primary approach to learning to read and write and that such teaching should be embedded within a broad language and literacy curriculum (p. 70).

An additional recommendation following the Rose Review was that teachers assess children's phonic skills throughout the first 3 years of formal education (ages 4–7 years). Where implemented, this is typically achieved by teachers tracking children's progress through a series of developmental 'phonic phases', which are outlined in the teaching handbook *Letters and Sounds* (Department for Education and Skills, 2007). The phases move from demonstrating attention to sounds in the environment (Phase 1) to confident and fluent use of grapheme–phoneme correspondences for reading and spelling unfamiliar words (Phase 6). Moreover, although there are concerns about the ability of teachers to judge the reading skills of their pupils (see Madelaine & Wheldall, 2010, for a review), a number of UK studies have shown that teachers well versed in phonic strategies and monitoring procedures can provide reliable estimates of children's reading abilities as measured by objective tests (Snowling, Duff, Petrou, Schiffeldrin & Bailey, 2011; Snowling, Hulme, Bailey, Stothard & Lindsay, 2011; Snowling et al., 2009).

Despite the focus of government policy on the implementation of systematic phonics in recent years, the proportion of pupils leaving primary school with the expected level in English had stalled at around 80% (Department for Education, 2010). In response to this, in 2012, the UK coalition government introduced a statutory check on early reading progress – the phonics screening check. The stated purpose of the check is, 'To confirm whether individual pupils have learnt phonic decoding to an appropriate standard'. Consequently, 'Pupils who have not reached this standard at the end of Year 1 should receive support from their school to ensure they can improve their phonic decoding skills' (Department for Education, 2012a). Thus, the check forms part of an aim to identify and help struggling readers early in their literacy development (Department for Education, 2010).

The phonics screening check was administered in all maintained schools in England for the first time in June 2012. The check comprises 40 items – 20 real words and 20 pseudowords. All items are phonically regular, ranging from items with three letters in a consonant–vowel–consonant format (e.g., the pseudoword *pib*) to two-syllable words containing consonant clusters and vowel digraphs (e.g., the word *portrait*). To 'meet the standard', children had to read at least 32 of the items correctly without support or prompting. Figures show that 58% of pupils met this standard in 2012, and 69% in 2013 (Department for Education, 2012b, 2013b). Children's scores are reported to their parents/guardians and to the Department for Education; and each school's results can be used as evidence within school inspections.

The phonics screening check has been met with a significant amount of controversy. Educators have questioned its necessity, voicing concerns about whether the check will add any valuable information to what teachers already know about their pupils' progress (e.g., National Union of Teachers, 2012). There have also been objections to the statutory nature of the check, with concerns about the resource implications of mandatory testing and the negative consequences when such tests become 'high-stakes' (e.g., Association of Teachers and Lecturers, 2011; Brooks, 2010). Indeed, a survey of nearly 3000 teachers – conducted after the administration of the check but before its results – reported that 87%

of respondents did not agree with the statutory implementation of the check and thought that it should be discontinued (ATL/NAHT/NUT, 2012).

In this paper, we ask two critical questions about the phonics screening check: First, is the check valid (i.e., does it function as a useful measure of decoding skills); and second, is the check sensitive (i.e., can it identify children who are showing early signs of being at risk of a reading difficulty)? We also consider whether, given our findings, it is necessary. Answering the first question is relatively straightforward: We can investigate how well scores on the check correlate with reading skills measured by objective tests. The second is more difficult because there is no 'gold standard' for the identification of reading difficulty, and any cut-off between 'impaired' and 'normal' reading is arbitrary. The recently published American Psychiatric Association diagnostic manual (American Psychiatric Association, 2013) suggests low achievement is indicated by a score of at least 1.5 standard deviations (*SD*s) below average, translating to a standard score of 78 (p. 69). However, most agencies would contend that this is much below the desirable level of literacy for the UK population. We propose that an attainment level of 1 SD below the age norm (a standard score of 85) would be sufficient to place a child 'at risk' of significant underachievement, and hence, we use this as a cut-off against which to determine the sensitivity of the check.

To answer the preceding questions, we analysed attainment data from eight primary schools in York, selected to be representative of their particular local authority; these data included phonics screening check results and teacher assessments of children's phonic phase attainments. A subsample of children from these schools was seen on an individual basis to obtain psychometric data from tests of reading and associated skills, which were then related to the teacher's assessments.

Method

Participants

Eight primary schools participated in this study. They were chosen from 18 that volunteered and were selected to be representative of their local authority. The sample included schools in city centre and suburban settings, with catchment areas of varying socioeconomic status (SES). School-based assessment data were collected from 292 Year 1 children (5- to 6-year-olds); 160 of these children were also seen individually by researchers for more in-depth testing (mean age = 6 years 4 months [*SD* = 4 months]). No exclusion criteria were applied. Parental consent was obtained for all children to take part in the individual testing sessions. Background information was available for all children with respect to SES (measured by receipt of free school meals [FSM] and the Index of Multiple Deprivation [IMD] based on home postcode information), whether they had English as an additional language (EAL) and whether they were on the Special Educational Needs (SEN) register. For the full sample, 15% received FSM, 12% were on the SEN register and 6% had EAL, and the average IMD score was 15.32 (11.98) – where a higher score reflects greater deprivation. For the subsample, 10% received FSM, 8% were on the SEN register, 5% had EAL and the average IMD score was 13.32 (10.93). Averages for state-funded primary schools in England for 2011 were 15% FSM (Department for Education, 2011a), 19% SEN (Department for Education, 2011b) and 17% EAL (Department for Education, 2011a); and the average IMD in 2010 was 21.67 (Department for Communities and Local Government, 2011).

Materials

School-based assessments
• *Phonic phases.* Introduced in the Rose Review (Rose, 2006), this hierarchical scale of six phases provides teachers with descriptors of phonic knowledge and skills, from the earliest attention to sounds in the environment to decoding and spelling multi-syllabic words with alternative grapheme–phoneme correspondences. Teachers assess each pupil against the phases, making a 'best-fit' judgement of the phase that most accurately describes the pupil's current attainment in phonics. The expected level of attainment for children at the end of Year 1 is phonic Phase 5. Judgements for the current study were recorded in May 2012.

• *Phonics screening check.* As described earlier, this is a 40-item test, comprising 20 real words and 20 pseudowords, administered one to one by a teacher with no prompting or help. The pupil's responses are scored as correct or not (internal reliability, $\alpha = .96$; Standards & Testing Agency, 2012).

• *Spelling.* A spelling task was group administered by Year 1 teachers to all children in their class. This was adapted from the *British Ability Scales II* (Elliot, Smith & McCulloch, 1997). There were 14 items: on, and, go, sit, was, home, play, that, are, well, bird, boat, friend and know. Children were required to spell to dictation: Each word was said in isolation, embedded in a sentence and repeated in isolation. Children were not credited for reversed letters.

Individual assessments
• *Word reading.* The *Single Word Reading Test* from the *York Assessment for Reading Comprehension* (*YARC* – Snowling et al., 2009) was administered to assess word reading accuracy. This test comprises 60 words of increasing difficulty and includes both regular and exception words. Testing was discontinued after five consecutive errors (internal reliability, $\alpha = .98$).

• *Reading comprehension.* Reading comprehension was assessed via the *YARC Passage Reading* test (Snowling et al., 2009). All children were administered the Level 1 passage, which was devised to be suitable for children in Year 1. Children were timed while they read the passage aloud. They were then asked eight open-ended questions about the story. Measures of prose reading accuracy, rate and comprehension were derived. Reading was discontinued if more than 15 errors were made while reading the passage aloud. In these instances, an accuracy score of 16 errors was assigned, timing was not recorded and the comprehension questions were not administered, yielding a score of 0 (reliability coefficient, $\alpha = .64$).

• *Nonword reading.* The *Graded Nonword Reading Test* (Snowling, Stothard & McLean, 1996) was used to measure decoding skills. Children were presented with nonwords that increased in difficulty from 'hast' to 'sloskon' and were asked to read them aloud. There were five practice items and 20 test items, and the task was discontinued after six consecutive errors (internal reliability, $\alpha = .96$).

• *Phonological awareness.* Phonological awareness was measured using the *Sound Deletion* subtest from the *YARC* (Hulme et al., 2009). Children were presented with spoken words and

corresponding colour pictures, asked to repeat the word and then to delete a sound (e.g., say 'sheep' without the 'sh'). Children completed all 12 items, which tapped deletion of syllables and phonemes in initial, medial and final positions (internal reliability, $\alpha = .93$).

• *Expressive Vocabulary.* The *Expressive Vocabulary* subtest from the *Clinical Evaluation of Language Fundamentals IV* (Semel, Wiig & Secord, 2003) was used to assess production of single words. Children were asked to name pictures that increased in difficulty. There were 27 items, and the task was discontinued after seven consecutive errors (internal reliability, $\alpha = .85$).

• *Mathematics.* The *One Minute Addition Test* and *One Minute Subtraction Test* provided measures of children's mathematical skills (Westwood, Harris-Hughes, Lucas, Nolan & Scrymgeour, 1974). Children had 1 minute to complete as many additions as possible, followed by 1 minute for subtractions. There were 30 items on each test. Children were credited for reversed numbers (test/re-test reliabilities, $\alpha = .91$ and $.90$; K. Moll, personal communication, 23 January 2013). Children's scores on the two tests were summed to form a combined mathematics score.

Procedure

Children were tested individually for one 30-minute session in the 3 weeks after the national phonics screening check took place in 2012. The tasks were administered in the following order: *Expressive Vocabulary*, *Single Word Reading Test*, *Sound Deletion*, *Graded Nonword Reading Test*, *One Minute Addition* and *Subtraction Tests* and *Passage Reading*. The spelling task was administered by teachers on a group basis also within this 3-week period.

Results

The attainments of the full sample and subsample on measures of language, literacy and maths are summarised in Table 1. As a group, the subsample's performance on all standardised reading measures was above the population mean of 100 but within the average range. The distribution of the scores on the phonics screening check from our full sample is shown in Figure 1. This shows that the data are not normally distributed. Notably, although there was a long tail in the distribution, some 19% of the sample scored between 30 and 33, with a sudden jump in frequency for those attaining scores of 32–33 compared with 30–31 on the test. With respect to teacher judgements of phonic skill, the number of children distributed across each phase was as follows: Phase 1 = 2, Phase 2 = 15, Phase 3 = 28, Phase 4 = 88, Phase 5 = 111 and Phase 6 = 47.

Is the phonics screening check valid?

The validity of an instrument refers to whether it measures what it claims to. This was assessed by correlating the phonics screening check with teacher judgements of phonic skill and with various standardised measures (Table 2). Owing to the nonnormal distribution of the phonics screening check (Figure 1), Spearman's correlations were run. Only a limited number of variables were collected for the full sample. However, the pattern of correlations between these variables and the phonics screening check is very similar across the full sample and subsample. Thus, reporting will focus on the correlation coefficients from the subsample.

Table 1. Descriptive statistics for the full sample and subsample on language and literacy measures.

	N	Mean	SD	Range
Full sample				
Spelling[a] (14)	216	9.42	3.09	1–14
Phonic phase[a] (6)	291	4.48	1.08	1–6
Phonics screening check[a] (40)	292	30.84	9.28	0–40
Subsample				
Spelling[a] (14)	136	9.78	3.11	2–14
Phonic phase[a] (6)	160	4.67	1.01	1–6
Phonics screening check[a] (40)	160	32.52	7.63	0–40
Sound deletion[a] (12)	159	8.57	2.05	2–12
Sound deletion[b]	159	108.26	11.46	80–132
Single-word reading[a] (60)	160	24.04	10.83	1–50
Single-word reading[b]	160	110.31	14.55	69–131
Graded nonword reading[a] (20)	159	12.77	5.07	0–20
Expressive vocabulary[a] (54)	160	29.27	9.29	3–48
Expressive vocabulary[c]	160	10.22	3.00	1–17
One minute addition and subtraction[a] (60)	159	12.11	6.63	0–37
Prose reading accuracy[a] (16)	158	5.77	5.44	0–16
Prose reading accuracy[b]	158	110.49	11.79	82–131
Prose reading rate (seconds)[a]	138	93.11	83.59	20–560
Prose reading rate[b]	138	111.66	12.10	76–131
Prose reading comprehension[a] (8)	158	4.18	2.22	0–8
Prose reading comprehension[b]	157	105.39	13.77	75–131

Notes: For prose reading accuracy and rate, a higher raw score represents poorer performance.
[a]Raw score.
[b]Standard score.
[c]Scaled score.

The phonics screening check correlates strongly with teacher judgements of phonic phases ($r = .72$) and with standardised measures of reading accuracy (nonword reading, single-word reading and prose reading accuracy, r's = .75–.83) and spelling ($r = .72$). It also correlates well with phoneme awareness, prose reading rate and comprehension (r's = .57–.68). In contrast, there are more moderate correlations between the phonics screening check and vocabulary and maths (r's = .45), indicating that the check is more specific to the domain of literacy and does not simply measure general abilities. Thus, the phonics screening check shows convergent and discriminant validity.

Is the phonics screening check sensitive?

A score of 32 out of 40 was set as the threshold for the phonics screening check in 2012 and 2013 – that is, children at or above this level are said to have met the required standard of phonic decoding (Department for Education, 2012b, 2013b). Nationally, 58% of pupils
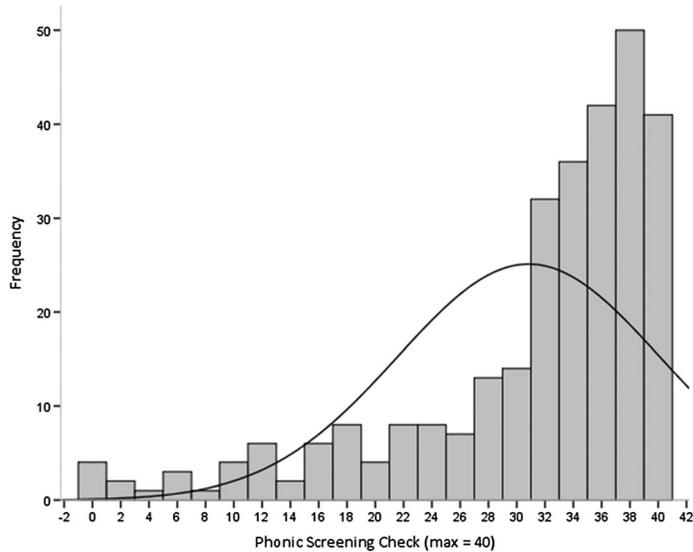
Figure 1. Distribution of scores on the phonics screening check ($N = 292$), with normal distribution curve.

met this standard in 2012 (Department for Education, 2012b). In the current sample, 194 of 292 children met the standard (66%); in the subsample, 115 of 160 met the standard (72%).

As with any new screening test, it is important to compare its ability to identify risk for difficulties with that of more established approaches. Thus, the classification function of the phonics screening check was compared with that of standardised measures of reading, and a routine teacher assessment (phonic phase judgements). In terms of the standardised measures, the subsample was performing in the high-average range. On these standardised measures, we defined risk of a reading difficulty as a score of more than 1 *SD* below the population mean (i.e., a standard score of <85) on one of two tests: single-word reading or prose reading rate.[1] In this way, 16 out of 160 children (10%) met the criteria for being at risk of reading difficulties (nine children based on single-word reading accuracy and seven (non-overlapping) children based on prose reading rate). With respect to teacher judgements of phonic phases, risk of a reading difficulty was defined as scoring less than the expected level for children at the end of Year 1 (i.e., <5).[2] This method identified 133 children (46%) in the full sample and 62 children (39%) in the subsample.

Table 3 shows the classification of children according to whether or not they 'miss' the expected standard on the phonics screening check (<32) as a function of their risk of reading difficulties (i.e., whether or not they 'miss' the expected standard on standardised tests of reading [<85] or on the teacher assessment of phonic phases [<Phase 5]). Sensitivity gives a measure of those who are identified as at risk of a reading difficulty on the basis of the standardised reading tests or phonic phases and who are also identified as being at risk on the phonics screening check (a raw score of <32) – that is, the rate of true positives. Specificity gives a measure of those who are not defined as at risk of a reading difficulty on the basis of the standardised reading tests or phonic phases, nor on the phonics screening check (i.e., the rate of true negatives).

When the classification function of the phonics screening check (i.e., its ability to detect risk for reading difficulty) is compared with that of standardised tests of reading, the phonics screening check identifies all but two of the 16 children considered at risk of reading

Table 2. Bivariate Spearman's correlations between the demographic, language, literacy and mathematics variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Phonics check	—	.29***	.01	.72***	.72***	.79***	.83***	−.75***	−.68***	.57***	.57***	.46***	.46***
2. Age	—	—	−.11	.22**	.25**	.22**	.21**	−.26**	−.09	.22**	.14	.18*	.37***
3. Deprivation score	.02	—	—	−.07	−.13	−.05	−.11	.09	.05	−.09	−.09	−.27**	−.20*
4. Phonic phase	.71***	—	−.06	—	.72***	.67***	.81***	−.79***	−.74***	.66***	.61***	.59***	.51***
5. Spelling	.70***	—	−.11	.70***	—	.70***	.81***	−.71***	−.70***	.57***	.53***	.52***	.48***
6. Nonword reading						—	.83***	−.78***	−.63***	.54***	.59***	.47***	.48***
7. Single-word reading							—	−.83***	−.88***	.64***	.64***	.60***	.52***
8. Prose reading accuracy								—	.78***	−.69***	−.63***	−.51***	−.55***
9. Prose reading rate									—	−.53***	−.54***	−.53***	−.48***
10. Prose reading comprehension										—	.53***	.55***	.46***
11. Sound deletion											—	.53***	.52***
12. Expressive vocabulary												—	.49***
13. Maths													—

Notes: Correlations are above the diagonal for the subsample (n's = 117–160) and below the diagonal for the full sample (n's = 216–291). For prose reading accuracy and rate, a higher raw score represents poorer performance.

***$p < .001$,
**$p < .01$,
*$p < .05$.

Table 3. Classification of children according to whether they meet the expected standard on the phonics screening check and in phonic phases, as a function of risk for reading difficulties.

		Phonic phases				Standardised reading tests	
		Full sample (N = 291)		Subsample (N = 160)		Subsample (N = 160)	
		Miss (<Phase 5)	Meet (≥Phase 5)	Miss (<Phase 5)	Meet (≥Phase 5)	Miss (<SS of 85)	Meet (≥SS of 85)
Phonics screening check	Miss (<32)	81	16	37	8	14	31
	Meet (≥32)	52	142	25	90	2	133
Phonic phases	Miss (<Phase 5)	—	—	—	—	15	47
	Meet (≥Phase 5)	—	—	—	—	1	97

Notes: Risk for reading difficulties is defined as children 'missing' the expected standard on standardised tests of word reading accuracy or prose reading fluency (i.e., a standard score [SS] of <85) or 'missing' the expected standard in phonic phases (i.e., <Phase 5).

difficulties; its sensitivity is .88 (Table 3). However, it also identifies 31 children who were not classified as at risk (its specificity is .82). This suggests the phonics screening check is good at identifying children who are truly at risk of reading difficulties but slightly over-estimates the number of children at risk. In comparison, when the classification function of the phonics screening check is compared with that of teacher judgements of phonic phases, the check shows reduced sensitivity in identifying risk (.60 for the subsample and .61 for the full sample), but increased specificity (.92 for the subsample and .90 for the full sample).

Finally, we evaluated how effectively teachers might identify risk for reading difficulty, compared with identification through standardised reading tests. This was carried out by comparing rates of classification of risk from the phonic phases to those from the standardised reading tests. The phonic phases identify all but 1 out of 16 children who are at risk of reading difficulties (its sensitivity is .94) but also identify 47 children who are not deemed at risk (its specificity is .67). This suggests that when using phonic phases, teachers are very good at identifying children truly at risk of reading difficulties but that the number of children at risk is overestimated – more so than the phonics screening check.

Is the phonics screening check necessary?

We were also interested in whether the phonics screening check added any value to what teachers who were well trained in phonic teaching and assessment already know about children's literacy skills, on the basis of their routine assessments. The strength of the cor-relations that the phonics screening check and the phonic phase judgements had with all the literacy measures was compared (from Table 2). There were no significant differences in how the assessments correlated with spelling, single-word reading, prose reading accu-racy or prose reading rate (differences in $r = .00$–.06, Z's $= -1.43$ to 0.00, p's $= .152$ to $<.001$). However, the phonics screening check correlated more strongly with nonword reading than did phonic phase judgements (difference in $r = .12$, $Z = 3.24$, $p = .001$); but the phonic phase judgements correlated more strongly with reading comprehension than did the phonics screening check (difference in $r = -.09$, $Z = -1.99$, $p = .047$).

Discussion

In June 2012, the UK government introduced a new statutory assessment to assess the pho-nic decoding skills of Year 1 pupils. The aims of this phonics screening check are to ascer-tain whether pupils' phonic skills are at an appropriate standard, to identify children who may be struggling in this area and ultimately to raise literacy standards. We obtained the results of the phonics screening check from a sample of Year 1 pupils in one local author-ity, together with teacher assessments of phonic phases and psychometric test data. In light of the various criticisms levelled at the phonics screening check, we assessed whether the check was valid and sensitive; we also considered its necessity.

The validity of an instrument refers to whether it measures what it claims to measure. Our analyses show that the phonics screening check is a highly valid measure of children's phonic skills. The check showed convergent validity by correlating strongly with other measures of phonic skills (e.g., teacher judgements of phonic ability and psychometric tests of nonword reading and spelling) and with broader measures of reading (e.g., single-word reading accuracy, prose reading accuracy and comprehension). It also demonstrated

discriminant validity, by showing weaker correlations with more distal skills (e.g., vocabulary and maths).

Furthermore, the phonics screening check seemed to be sensitive with respect to identifying children at risk of reading difficulties. We compared the check's ability to identify struggling readers (those who scored <32) with identification rates based on preexisting standardised tests of reading accuracy and fluency (for which we defined risk of a reading difficulty as a standard score of <85 – 1 *SD* below the age-expected mean). It is worth noting that caution should be applied when interpreting these analyses, which have involved dichotomising continuous data and identifying a group of putative poor readers when dyslexia is essentially a dimensional disorder (Pennington, 2006; Rose, 2009). Note also that categorisation according to the standardised reading tests only identified a small number of at-risk readers ($n = 16$, 10%), potentially limiting the extent to which our findings can be generalised, and that sensitivity and specificity rates are intrinsically linked to where the cut-point between 'typical' and 'at-risk' reading is drawn. Nevertheless, the phonics screening check classified most of the same children as at-risk readers as the reading measures had done, resulting in a high sensitivity of 88%. The check also had an acceptable specificity of 82%. This lower figure demonstrates that the check slightly overestimated the prevalence of at-risk readers; however, this might be considered a good property for a screening instrument, given that it is preferable to overidentify the children who might need additional help in order to catch all those who certainly do.

Taken together, these observations lead us to conclude that the phonics screening check is a valid instrument for measuring word-level reading ability and sensitive in identifying young children at risk of a reading difficulty – although this needs to be verified by follow-up data. However, it is important to note that analysis of the distribution of scores on the phonics screening check raises some doubts about its integrity. Notably, there was a sudden spike in the frequency of scores at 32 – the threshold for meeting the standard. This trend is accentuated in the national data obtained in both 2012 and 2013 (Department for Education, 2012b, 2013b). Reasons suggested for this unusual distribution include the release of the score for meeting the standard prior to the test administration and subsequent marking up of scores that fell on the borderline of meeting the standard (e.g., Bishop, 2012; Standards & Testing Agency, 2012; Townley & Gotts, 2013). If this interpretation is correct, it questions the objectivity of the instrument and the utility of the national data for tracking standards over time (Bishop, 2013). It is noteworthy that the threshold for meeting the standard in 2014 will not be released until after the check has been administered (Standards & Testing Agency, 2013, p. 17).

A survey of nearly 3,000 teachers (ATL/NAHT/NUT, 2012) found that 91% thought the phonics screening check failed to tell them anything they did not know already about children's reading abilities. To the contrary, the Department of Education (2013a) maintains that schools with good reading assessment systems benefited from the phonics screening check because it helped to identify children whose difficulties with phonics were hidden by strong sight word reading skills. This argument implies that teacher assessments of phonic skills may not be well correlated with objective assessments of sight word reading. However, our results showed that well-trained teachers were as good at judging children's reading skills as was the phonics screening check in so far as both tended to correlate to the same extent with standardised reading measures – notably with tests of sight word reading and prose reading accuracy. There was one significant difference in favour of the phonics screening check, such that it correlated more highly with a standardised test of nonword reading. This might be a reflection of the fact that both these

tasks are highly focused on the reading aloud of fully decodable items. In contrast, the phonic phases correlated more strongly with reading comprehension, which suggests that, although the assessment entirely focuses on phonic skills, the phonic phase judgements are also capturing broader aspects of literacy.

Furthermore, teachers' judgements showed at least as good sensitivity in terms of identifying children who were defined operationally from standardised reading measures as at risk of reading difficulties (94% sensitivity of phonic phases, compared with 88% sensitivity of the phonics screening check). However, the phonics screening check was less likely to overestimate the prevalence of at-risk readers (82% specificity of the phonics screening check compared with a 67% specificity of the phonic phases). Care must be taken when interpreting this apparent tendency of teachers to overidentify risk of reading difficulties. First, we chose to define risk as not yet attaining the phonic phase expected by the end of Year 1 (i.e., not attaining phase 5). Because teacher judgements were recorded in May 2012, it is possible that more children would have been judged to have reached Phase 5 by the end of that summer term (July 2012). Second, this cut-off point had been imposed on the data at the point of analysis; sensitivity of teacher judgements may well have been higher had teachers been asked explicitly to state which pupils they considered to be at risk.

On balance, given the strong stability of reading once the system is set up (Lervåg & Hulme, 2009), we concur that a rigorous assessment of phonic skill is important for early identification of children at risk of reading difficulties. However, we argue that when teachers are well educated about the cognitive mechanisms involved in reading, and they have training in the teaching and assessment of phonics, their judgements are sufficient for this purpose, and a mandatory phonics screening check is not necessary. Although a likely counter-argument is that well-monitored phonics teaching is not systematically implemented on a national level, we believe that the use of resources to better equip teachers to conduct ongoing phonic assessments would be more cost-effective, not least because this would place them in the best position to intervene before reading difficulties set in.

There are some additional limitations to acknowledge. First, the sample size for our study was relatively small for the validation of a national screening measure. Second, the focus on a local authority in which teachers have been well educated in order to implement the phonics strategy and in which pupils were highly attaining may have introduced bias. Third, as teacher judgements of children's phonic phases influence the level at which they are taught, an underestimate of phonic ability could constrain children's phonic development and in turn their performance on the check. It might be for this reason that teacher judgements and performance on the phonic check correlate well. The risk of this is reduced, however, where teachers have been well trained in phonic assessment and teaching, as in this study. These limitations aside, our main point is to emphasize what *can* be achieved by teachers who are properly trained and empowered to implement systematic phonics in their classrooms as well as to accurately and consistently monitor their pupil's progress.

Conclusion

We have shown that the new phonics screening check is a valid measure of phonic skills and is sensitive to identifying children at risk of reading difficulties. Its slight tendency to overestimate the prevalence of at-risk readers (as compared with standardised tests of reading accuracy and fluency) is arguably a favourable property for a screening instrument.

We agree that early rigorous assessment of phonic skills is important for the timely identification of word reading difficulties. However, combining our observations about the integrity of the national phonics screening check data with our findings that teachers perform reliable and sensitive assessments of phonics progression, we argue in favour of using resources to continue to train and support teachers in the knowledge, assessment and teaching of early literacy skills.

Notes

1. Note that prose reading rate scores are not calculable for children with the poorest prose reading accuracy scores; children who made >15 reading errors were not given a time score. Thus, our estimation of the number of children at risk is conservative.
2. With respect to the school-based assessments, we have taken age-related expectations as determined by the Department for Education as cut-off points and used these to distinguish between those considered at risk or not at risk of a reading difficulty. However, it is important to note that the Department for Education guidelines do not specify that scores below the cut-offs constitute a reading difficulty.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th edn). Arlington, VA: American Psychiatric Publishing.

Association of Teachers and Lecturers (2011, February). Year one phonics screening check: Response from the Association of Teachers and Lecturers. Retrieved 3 January 2013, from http://www.atl.org.uk/Images/2011% 20%20Y1_phonics_test_consultation_response.pdf

ATL/NAHT/NUT (2012, July). Teachers' and head teachers' views of the year one phonics screening check. Retrieved 23 November 2011, from http://www.teachers.org.uk/phonics

Bishop, D.V.M. (2012, October). Data from the phonics screen: A worryingly abnormal distribution. Retrieved 3 January 2013, from http://deevybee.blogspot.co.uk/2012/10/data-from-phonics-screen-worryingly.html

Bishop, D.V.M. (2013, October). Good and bad news on the phonics screen. Retrieved 9 October 2013, from http://deevybee.blogspot.co.uk/2013/10/good-and-bad-news-on-phonics-screen.html

Brooks, G. (2010, December). The government's proposed decoding test for 6-year-olds. Retrieved 3 January 2013, from http://www.ukla.org/news/professor_greg_brooks_critiques_the_governments_proposed_decoding_test_for_/

Byrne, B. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle*. Hove, UK: Psychology Press.

Byrne, B. & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology*, 81, 313–321. DOI: 10.1037//0022-0663.81.3.313

Camilli, G., Vargas, S., & Yurecko, M. (2003, May). Teaching children to read: The fragile link between science and federal education policy. *Education Policy Analysis Archives*, 11. Retrieved 9 July 2009, from http://epaa.asu.edu/epaa/v11n15/

Department for Communities and Local Government (2011, March). English indices of deprivation 2010: Overall. Retrieved 2 January 2013, https://www.gov.uk/government/publications/english-indices-of-deprivation-2010

Department for Education (2010). *The importance of teaching – The schools white paper*. London: The Stationery Office.

Department for Education (2011a, June). Schools, pupils and their characteristics, January 2011. Retrieved 2 January 2013, from http://www.education.gov.uk/rsgateway/DB/SFR/s001012/index.shtml

Department for Education (2011b, June). Special educational needs in England: January 2011. Retrieved 2 January 2013, from http://www.education.gov.uk/rsgateway/DB/SFR/s001007/index.shtml

Department for Education (2012a, May). Phonics screening check for Year 1 pupils. Retrieved 3 January 2013, from http://www.education.gov.uk/schools/teachingandlearning/pedagogy/phonics/a00197709/developing-a-new-year-1-phonics-screening-check

Department for Education (2012b, September). Phonics screening check and national curriculum assessments at Key Stage 1 in England, 2011/2012. Retrieved 3 January 2013, from http://www.education.gov.uk/rsgateway/DB/SFR/s001086/sfr21-2012.pdf

Department for Education (2013a, July). Phonics screening check FAQs. Retrieved 28 October 2013, from http://www.education.gov.uk/schools/teachingandlearning/pedagogy/a00198207/faqs-year-1-phonics-screening-check#faq24

Department for Education (2013b, October). Phonics screening check and national curriculum assessments at Key Stage 1 in England, 2012/2013. Retrieved 11 October 2013, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/245813/SFR37-2013_Text.pdf

Department for Education and Skills (2007). *Letters and sounds: Principles and practice of high quality phonics*. London: Department for Education and Skills.

Ehri, L., Nunes, S.R., Willows, D.M., Schuster, B.V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36, 250–287. DOI: 10.1598/RRQ.36.3.2

Elliot, C.D., Smith, P., & McCulloch, K. (1997). *British Ability Scales II*. Windsor: NFER Nelson.

Hulme, C., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E. et al. (2009). *York Assessment of Reading for Comprehension: Early reading*. GL Assessment.

Lervåg, A. & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, 20, 1040–1048. DOI: 10.1111/j.1467-9280.2009.02405.x

Madelaine, A. & Wheldall, K. (2010). Teacher judgment of reading performance. In K. Wheldall (Ed.), *Developments in educational psychology* (2nd edn, pp. 196–216). London: Routledge.

McArthur, G., Eve, P.M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T. et al. (2012). Phonics training for English-speaking poor readers (Review). *Cochrane Database of Systematic Reviews 2012*, Issue 12. Art. No.: CD009115. DOI: 10.1002/14651858.CD009115.pub2.

National Reading Panel (2000). *Report of the National Reading Panel: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development Clearing House.

National Union of Teachers (2012, April). 'Five years old is too young to fail!': Year one phonics screening check report. Retrieved 3 January 2013, from http://www.teachers.org.uk/node/15556

Pennington, B.F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101, 385–413. DOI: 10.1016/j.cognition.2006.04.008

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. London: Department for Education and Skills Publications.

Rose, J. (2009). Identifying and teaching children and young people with dyslexia and literacy difficulties. DCSF Publications, DCSF 00659-2009.

Semel, E., Wiig, E.H., & Secord, W.A. (2003). *Clinical evaluation of language fundamentals* (4th edn). London: Psychological Corporation.

Snowling, M.J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A.M. (2011). Identification of children with dyslexia: The validity of teacher judgements using 'Phonic Phases'. *Journal of Research in Reading*, 34, 157–170. DOI: 10.1111/j.1467-9817.2011.01492.x

Snowling, M.J., Hulme, C., Bailey, A.M., Stothard, S.E., & Lindsay, G. (2011). *Better communication research project: Language and literacy attainment of pupils during early years and through KS2: Does teacher assessment at five provide a valid measure of children's current and future educational attainments? DFE-RR172a*. London: Department for Education. Retrieved 3 January 2013, from https://www.education.gov.uk/publications/RSG/publicationDetail/Page1/DFE-RR172A

Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E. et al. (2009). *York Assessment of Reading for Comprehension: Passage reading*. GL Assessment.

Snowling, M.J., Stothard, S.E., & McLean, J. (1996). *Graded Nonword Reading Test*. Bury St Edmunds, UK: Thames Valley Test Company.

Standards & Testing Agency (2012). Phonics screening check: 2012 technical report. Retrieved 23 January 2013, from http://media.education.gov.uk/assets/files/pdf/p/sta%20phonics%202012%20technical%20report_final.pdf

Standards & Testing Agency (2013). 2014 key stage 1 assessment and reporting arrangements. Retrieved 9 October 2013, from http://media.education.gov.uk/assets/files/pdf/k/2014_ks1_assessmentandreportingarrangementsara_26_sept.pdf

Stuebing, K.K., Barth, A., Cirino, P.T., Francis, D.J., & Fletcher, J.M. (2008). A response to the recent reanalyses of the National Reading Panel report: Effects of systematic phonics instruction are practically significant. *Journal of Educational Psychology*, 100, 123–134. DOI: 10.1037/0022-0663.100.1.123

Torgerson, C.J., Brooks, G., & Hall, G. (2006). *A systematic review of the research literature on the use of systematic phonics in the teaching of reading and spelling*. London: Department for Education and Skills Publications.

Townley, L. & Gotts, D. (2013). *Topic note: 2012 phonics screening check – Research report May 2013*. London: Department for Education.

Westwood, P., Harris-Hughes, M., Lucas, G., Nolan, J., & Scrymgeour, K. (1974). One minute addition test–one minute subtraction test. *Remedial Education*, 9, 70–72.

Fiona Duff completed her PhD and first postdoctoral position at the University of York before moving to her current post as a Research Associate at the University of Oxford.

Silvana Mengoni also completed her PhD at the University of York and is now a Research Associate at the Open University.

Alison Bailey recently completed her PhD at the University of York and continues to work as a Primary Literacy Consultant for the City of York Local Authority.

Margaret Snowling, having held a Personal Chair as a Professor in Psychology at the University of York, is now the President of St. John's College, University of Oxford.

Address for correspondence: Fiona Duff, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, UK OX1 3UD. E-mail: *fiona. duff@psy.ox.ac.uk*