Australian Journal of Learning Difficulties

ISSN: 1940-4158 (Print) 1940-4166 (Online) Journal homepage: http://www.tandfonline.com/loi/rald20

Do nonword reading tests for children measure what we want them to? An analysis of year 2 error responses

Anne Castles, Vince Polito, Stephen Pritchard, Thushara Anandakumar & Max Coltheart

To cite this article: Anne Castles, Vince Polito, Stephen Pritchard, Thushara Anandakumar & Max Coltheart (2018) Do nonword reading tests for children measure what we want them to? An analysis of year 2 error responses, Australian Journal of Learning Difficulties, 23:2, 153-165, DOI: 10.1080/19404158.2018.1549088

To link to this article: https://doi.org/10.1080/19404158.2018.1549088

Published online: 24 Nov 2018.

AUSTRALIAN JOURNAL OF LEARNING DIFFICULTIES
2018, VOL. 23, NO. 2, 153–165
https://doi.org/10.1080/19404158.2018.1549088

Routledge
Taylor & Francis Group

EMINENT RESEARCHER AWARD OF LEARNING DIFFICULTIES AUSTRALIA, 2017

Do nonword reading tests for children measure what we want them to? An analysis of year 2 error responses

Anne Castles, Vince Polito, Stephen Pritchard, Thushara Anandakumar and Max Coltheart

Department of Cognitive Science, ARC Centre of Excellence in Cognition and its Disorders Macquarie University, Sydney, NSW, Australia

ABSTRACT

Nonword reading measures are widely used to index children's phonics knowledge, and are included in the Phonics Screening Check currently implemented in England and under consideration in Australia. However, critics have argued that the use of nonword measures disadvantages good readers, as they will be influenced by their strong lexical knowledge and err by making word errors (e.g. reading flarm as "farm"). We tested this claim by examining the errors made by a group of 64 Year 2 children when reading aloud a set of simple nonwords. We found that stronger word readers were less likely to make a word error response than weaker word readers, with their most prevalent type of error being another nonword that was highly similar to the target. We conclude that nonword reading measures are a valid index of phonics knowledge, and that these tests do not disadvantage children who are already reading words well.

ARTICLE HISTORY
Received 12 November 2018
Accepted 12 November 2018

The importance of the explicit teaching of phonics skills to children during initial reading instruction is now widely agreed on by reading scientists (for a recent review, see Castles, Rastle, & Nation, 2018). The fundamental reason that teaching such skills is considered to be valuable is quite a simple one: it falls out of the nature of the English alphabetic writing system. This writing system is, to a significant degree, a code for sound: the letters on the page represent the sounds of spoken language. Therefore, if children are taught phonics, they are taught to crack the code. They can translate letters or groups of letters (known as graphemes) into sounds (or phonemes). If the pronunciation generated is in their oral vocabulary, they can then independently go from the sound of the word to its meaning, without having to ask a teacher or guess from context. As well, the phonics knowledge children have will generalise beyond individual words—a child with the knowledge required to read *went*, for example, will also be able to read *ten, wet*, and *net*. Although not all words in English follow standard grapheme-phoneme correspondences, basic phonics will allow children to reach a close phonological approximation of many words, which can then be adjusted using oral vocabulary knowledge (Elbro, de Jong, Houter, & Nielsen, 2012; Tunmer & Chapman, 2012).

CONTACT Anne Castles anne.castles@mq.edu.au

The weight of the research evidence has underpinned widespread recommendations to adopt systematic phonics instruction methods across the United States (National Reading Panel, 2000), the United Kingdom (Rose, 2006), and Australia (Rowe, 2005). Following the conclusions of Rose's (2006) review, state-funded schools in England now have a statutory duty to provide systematic phonics instruction from the commencement of school. The success of implementation is measured via a *Phonics Screening Check* given to all children at the end of the second year of reading instruction (Year 1). This Check requires children to read aloud 20 words and 20 nonwords (also referred to as pseudowords). The nonwords are considered to be crucial because they assess "pure" phonic knowledge without any impact of the children's existing knowledge of individual words.

The introduction of the Phonics Screening Check in England has resulted in dramatic improvements in phonics knowledge, at least as assessed by scores on the test: 58% passed the check in 2012 when it was first introduced, increasing to 82% in 2018 (U.K. Department for Education, 2018). Thus, the policy appears to have been successful in increasing schools' compliance in delivering systematic phonics instruction and in improving the quality of the delivery. These promising results have fuelled a push for the adoption of a similar policy in Australia, with the Federal Government recently proposing a National Literacy and Numeracy Check for Year 1 children, including a phonics screen (see Buckingham, 2016).

However, the introduction of the Phonics Screening Check in England has by no means been uncontroversial. A number of commentators have objected strenuously to the test, and in particular to its inclusion of nonwords such as *tay* and *plock* (e.g. Clark, 2016; Gibson & England, 2016). The concern has been that nonword reading tasks discriminate against good word readers and children who come to school already reading. Critics claim that such children will err by pronouncing the nonwords based on their strong word knowledge. For example, in a broad critique of the Check, Gibson and England (2016) review a teacher survey conducted by the U.K. Literacy Association (2012), summarising it as follows:

> "…many teachers thought that the Check misidentified pupils who came to school already seeing themselves as readers and who were beyond the stage of phonetic decoding: 'In several cases successful, fluent readers did less well in the Check than emergent readers…. Most schools surveyed indicated that the phonics Check seriously disadvantaged, and in some cases impeded, successful readers' (UKLA, 2012, p. 3. See also Lewis & Ellis, 2006, p. 15). The general view was that some above-average readers…looked for and substituted real or semantically meaningful words for pseudowords, e.g. 'shame' for shan, 'farm' for flarm or 'storm' for strom." (p. 494)

Embedded within the concerns raised here are some basic empirical questions. Specifically, in the initial stages of learning to read: (1) Do good word readers substitute real words for nonwords when trying to read aloud nonwords and, if so, do they do so more frequently than weaker word readers? (2) Do the errors of good word readers indicate lower reliance on basic grapheme-to-phoneme conversion strategies than less-skilled word readers? (3) More broadly, do children's error patterns on nonword reading tests provide evidence for the validity of this measure, in that errors consistent with the use of grapheme-phoneme conversion strategies are associated with higher total scores on the test? Although there have been numerous reports of nonword reading deficits in impaired readers (see e.g. Rack, Snowling, & Olson, 1992), and also some investigations of differences in strategies used by children during nonword reading (Bowey & Underwood, 1996; Brown & Deavers, 1999; Coltheart & Leahy, 1992), detailed analyses of children's errors when reading aloud nonwords that would provide answers to the questions posed above are lacking.

The aim of this study was to address these questions. We did so by analysing the nature of the errors that a group of Australian Year 2 children made when reading aloud a set of simple nonwords—similar to those used in the Phonics Screening Check—and examining how this related to their word reading abilities.[1] Such an investigation is timely, not just because of its implications for the validity of the use of nonwords in a phonics screen, but also more broadly for theories of reading and its acquisition. In the case of skilled adult readers, there has been an increasing awareness of the enormous variability that exists in nonword reading responses, not just between but even within individuals (Andrews & Scarratt, 1998; Coltheart & Ulicheva, 2018; Pritchard, Coltheart, Palethorpe, & Castles, 2012; Schmalz et al., 2014). These findings present a challenge for theoretical and computational models of reading, which typically represent the behaviour of just a single "ideal" reader. Therefore, gaining a better understanding of how developing readers approach the task of reading aloud nonwords, and the factors that influence their responses, can be informative for understanding the acquisition of phonics skills as well as for designing measures to effectively index them.

Our analyses focused on two key dimensions of the children's error responses in reading nonwords. First, we coded the *lexicality* of their responses: An error response that is a word (e.g. *rint* read as "rent") as opposed to another nonword (e.g *rint* read as "ront") would be considered evidence of the influence of lexical processes on nonword reading. If stronger word readers make a higher proportion of such responses than weaker word readers, this would provide some support for concerns raised by Gibson and England (2016) that good word readers are disadvantaged by their superior word-level knowledge when reading aloud nonwords. Secondly, we attempted to capture the degree of overall *similarity* of the responses to the target nonwords: *High Similarity* errors were those that differed by only one letter or phoneme from the target items while *Low Similarity* errors were those that differed by more than one letter or phoneme. If stronger word readers are applying phonics knowledge when they read nonwords, they would be expected to make a higher proportion of High Similarity errors (and a lower proportion of Low Similarity errors) than weaker word readers.

Method

Participants

The participants were 64 children (35 female; 29 male) from three different primary schools in Melbourne who were participating in a larger longitudinal study of reading being conducted by the first author. They were tested at around the middle of Year 2 (mean age = 8y,0m; range = 6y,11m – 8y,9m). One of the three schools was situated in an inner urban area, of relatively high socio-economic status; one was in an outer, semi-rural area, of medium socio-economic status, and one was in an outer urban area of relatively low socio-economic status. Reading instruction methods in all three schools came from a predominantly "whole language" tradition, although all three schools reported teaching some phonics and adopting a balanced approach to reading instruction.

Materials and procedure

For the error analysis, we examined the children's responses to a subset of 20 monosyllabic nonwords extracted from the nonword reading list of Castles and Coltheart (1993) word/nonword reading test. We chose to focus on the monosyllabic items since these are most comparable to the nonwords children would receive in a phonics screen. The nonwords were of mean length of 4.5 letters, with a range of 3–6 letters. The items, together with details on the number of children who made errors on each item and the number of different error responses, are presented in the Appendix.

The children read aloud these nonwords as part of the full administration of the first edition of the Castles and Coltheart test, which includes 30 regular words, 30 irregular words and 30 nonwords. The items were presented on individual cards in randomly mixed order. The children were told that some of the items were real words and some were nonsense words, and that they were to read them out loud as accurately as they could. Responses were untimed, and corrections were allowed—the child's final answer was coded as their response.

As an independent index of the children's word reading ability, we used the Word Identification subtest of the Woodcock Reading Mastery Test-Revised (*WRMT*; Woodcock, 1987). This consists of a series of words for reading aloud, ordered by difficulty and with a discontinuation rule applying after six consecutive errors. This test was administered during the same testing period (lasting approximately two weeks), but on a different day, as the Castles and Coltheart test.

Error analysis

A child's response in reading aloud a nonword was classified as an error if any grapheme was given a phoneme which was not the most common phonemic correspondence for that grapheme. First, errors that involved no response or a "don't know" response were identified. The remaining errors were then categorised based on the lexicality of response: (a) an error response which was a *Word*, otherwise known as a lexicalisation error or (b) an error response which was another *Nonword*. Within these categories, the errors were then further subcategorised according to their similarity to the target nonword. Specifically:

1. *High Similarity error*. The error response differed by only one phoneme or letter from the target nonword. This included errors where a letter or phoneme was substituted, and also errors where a letter or phoneme was added or deleted (e.g. *doash*—"doosh"; *farl* -> "foul"; *splatch*—"spatch"). We did not differentiate between visual similarity and sound-based similarity in this categorisation since, for most error responses, it was not possible to distinguish between these two possible sources of error (for example, if a child reads the nonword *doash* as "doosh", there is no way of knowing whether the has child misperceived the letter *a* as an *o* or whether they have incorrectly translated the grapheme *oa* into the phoneme/oo/).

2. *Low Similarity error*. The error response differed by more than one letter or phoneme from the target nonword (e.g. *delk* -> "drank"; *doash*—"druch"; *farl* -> "father")

The errors were categorised separately by two independent raters. Initial agreement between the raters was greater than 90%. Responses for which different classifications had been given by the two raters were discussed and a preferred classification agreed upon. For a very small number of items (fewer than 5%), a classification could not be agreed on or the transcription of the error could not be interpreted—these were placed in the "No Response/don't know" category.

Results

Errors in reading aloud nonwords

Overall, the children made a mean of 8.67 errors on the 20 monosyllabic nonwords (SD = 6.02). The frequency histogram is presented in Figure 1, which demonstrates a large spread in the total number of errors. One child made no errors and one child erred on all 20 items.

Results for different levels of word reading ability

To examine whether the types of error varied as a function of a child's word reading ability, we carried out a median split based on the children's raw scores on the Word Identification Test of the *WRMT*. The median raw score on this test was 48, so children with scores over 48 were classified as *Stronger Word Readers* and those with scores up to and including 48 as *Weaker Word Readers*. Descriptive statistics on all the reading measures for the two groups are presented in Table 1. As well as performing significantly more poorly on the categorisation measure—the Woodcock Word Identification test—the Weaker Word Readers performed more poorly than the Stronger Word Readers on

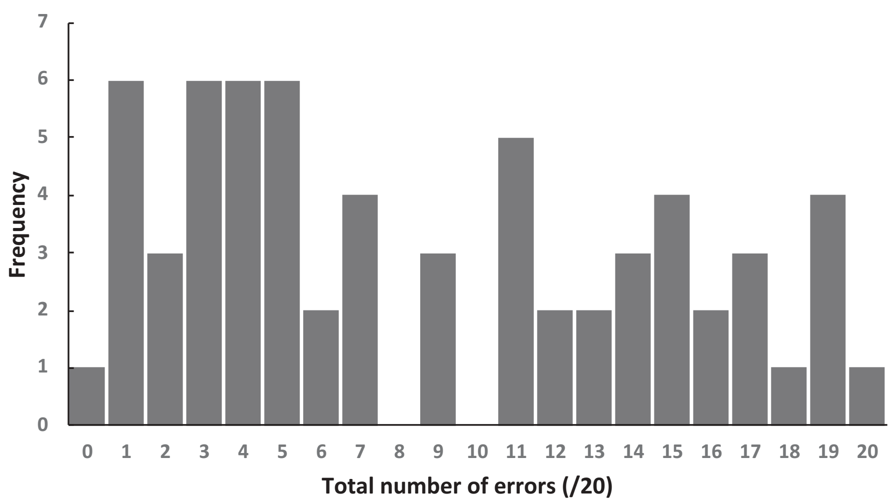

Figure 1. Distribution of error scores on the nonword reading test.

Table 1. Performance on reading measures for the stronger and weaker word reader groups.

	Stronger word reader (N = 31)		Weaker word reader (N = 33)	
Test	*M*	*SD*	*M*	*SD*
Woodcock Word Identification (raw)	61.29	7.58	40.03	7.92
Castles and Coltheart test				
Regular Word Accuracy (/30)	27.33	2.66	16.12	6.50
Irregular Word Accuracy (/30)	18.13	2.48	11.30	3.83
Nonword Accuracy (/30)	23.68	4.86	8.86	6.51
Monosyllabic Nonword Errors (/20)	3.94	2.97	13.12	6.02

all of the Castles and Coltheart subtests, and made a significantly higher number of errors on the monosyllabic nonwords (all p values < .001).

Overall, there was only a small proportion of no responses or "don't knows" (M = 3.6%; SD = 15). The Stronger Word Readers never produced such a response, while the Weaker Word Readers did so on an average of approximately 7% of occasions, but this difference was not significant, $F(1, 61) = 3.31$, $p = .074$. These responses were not examined further (although they were included in the total error count).

The nonword reading errors for the two groups, broken down by the lexicality of response (Word error, Nonword error), are presented as box and whisker plots in Figure 2. The differences in the proportions of these errors made by the Stronger and Weaker Word Readers were analysed using one-way ANOVAs, with Reader Group as a between-subjects factor.[2] The Stronger Word Readers were significantly less likely to produce lexicalisation errors than were the Weaker Word Readers, $F(1, 61) = 4.88$, $p = .031$. As can be seen, by far the predominant error type made by the Stronger Word Readers was to produce another, incorrect nonword (mean of 72% of total

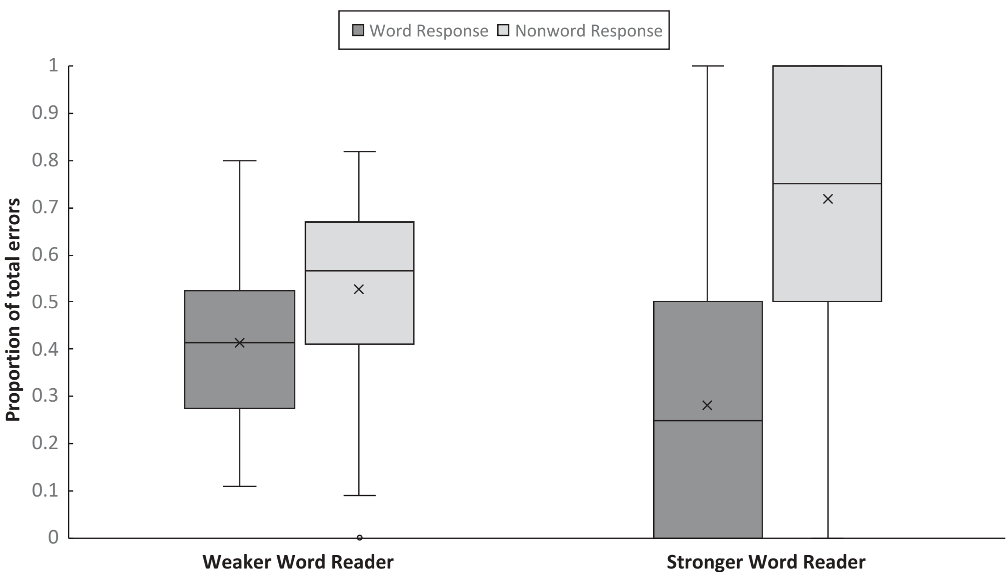

Figure 2. Box-and-whisker plot of lexicality of error responses as a function of word reading ability. The boxes represent the interquartile range and the whiskers the outer range, with outliers shown as circles. Means are represented as Xs.

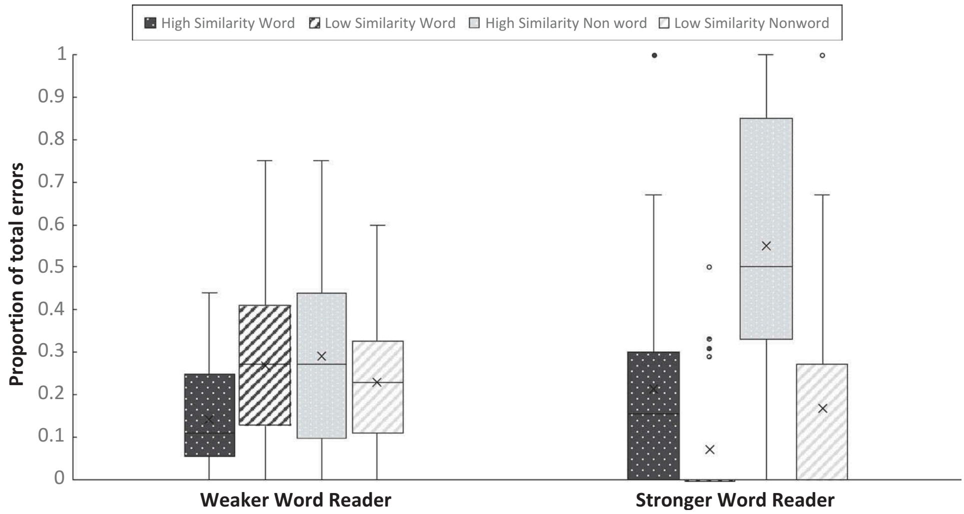

Figure 3. Box-and-whisker plot of similarity and lexicality of error responses as a function of word reading ability. The boxes represent the interquartile range and the whiskers the outer range, with outliers shown as circles. Means are represented as Xs.

errors), and they did so on a significantly greater proportion of occasions than the Weaker Word Readers, $F(1, 61) = 10.17$, $p = .002$.

The results further broken down by the similarity of the error response to the target nonword are presented in Figure 3. The Stronger Word Readers made a greater proportion of High Similarity errors than the Weaker Word Readers, $F(1, 61) = 24.32$, $p < .001$. As well, as can be seen, these High Similarity errors were predominantly other nonwords that were highly similar to the target nonword (mean of 55% of total errors), rather than similar word responses (mean of 21%). Reflecting this, the two groups differed significantly in the proportion of High Similarity nonword responses that they produced, $F(1, 61) = 14.82$, $p < .001$, but not in the proportion of High Similarity word responses, $F(1, 61) = 1.69$, $p = .19$. Correspondingly, the Weaker Word Readers produced a greater proportion of Low Similarity word responses than the Stronger Word Readers, $F(1, 61) = 23.59$, $p < .001$, but the groups did not differ in the proportion of Low Similarity responses that were other nonwords, $F(1, 61) = 1.29$, $p = .26$.

To convey the pattern across the entire sample, Figure 4 presents a scatterplot of the proportion of word error responses when reading the nonwords as a function of word reading ability (Woodcock Word Identification). A negative correlation is evident ($r = -0.24$, $p = .051$), with the proportion of lexicalisations tending to decrease with increasing word reading scores. Eleven children made no error responses that were lexicalisations, and all of these participants scored above the median on the Woodcock Word Identification test. The two children who made 100% lexicalisation responses also had above-median Woodcock Word Identification scores, but it should be noted that each of these children made only one error on the nonword reading test, to which they produced a high similarity word response. When we excluded the six children who only made one error on the nonword reading test from the analysis, the negative correlation increased to $r = -0.37$ ($p = .004$).

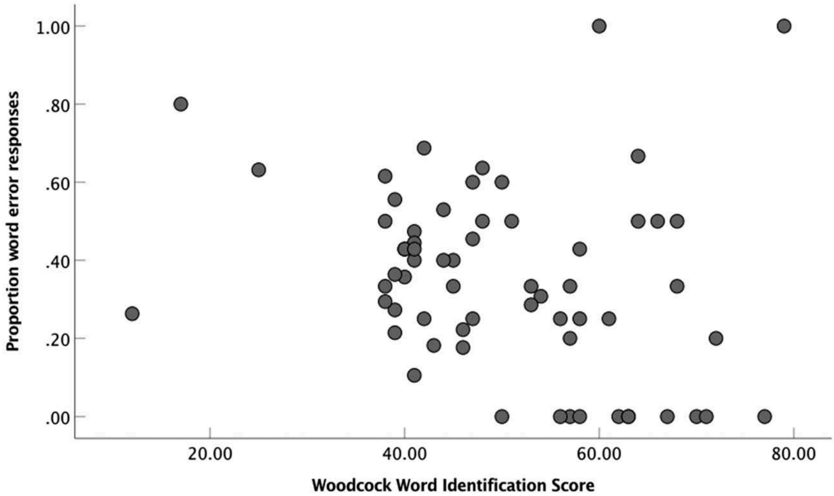

Figure 4. Scatterplot of lexicalisation responses as a function of word reading ability.

Table 2. Pearson correlations between nonword reading score and proportion of error types.

	Error response (% of total errors)			
	Word		*Nonword*	
	High Similarity	*Low Similarity*	*High Similarity*	*Low Similarity*
Nonword Score (/20)	.21	−.62**	.59**	−.27*

Key: * Significant at $p < .05$; ** Significant at $p < .01$

Error types as a function of total nonword reading score

Finally, we examined what kinds of errors were typical of those obtaining high versus low overall accuracy scores on the nonword reading test. Table 2 presents Pearson correlations between total nonword score and the proportion of each kind of error response. As can be seen, total nonword reading scores were strongly positively associated with the production of *High Similarity Nonword* error responses and strongly negatively associated with the production *Low Similarity Word* responses.

Discussion

The analyses reported here were conducted to address questions surrounding the way in which good and poor word readers in a typical early primary school setting approach the task of reading nonwords, and the implications of this for the validity of the Phonics Screening Check. Specifically, we examined claims that children who are already strong word readers will be inclined to draw on this lexical knowledge when reading aloud nonwords, making a high proportion of lexicalisation responses (Clark, 2016; Gibson & England, 2016). We asked the following questions: (1) Do good word readers substitute real words for nonwords when trying to read aloud nonwords and, if so, do they do so more frequently than weaker word readers? (2) Do the errors of good word readers indicate lower reliance on basic grapheme-to-phoneme conversion strategies than less-skilled word readers? (3) More broadly, do children's error patterns on nonword reading provide evidence for the validity of this measure, in that errors consistent with reliance on grapheme-phoneme conversion are associated with higher total scores on the test? We discuss the results in relation to each of these questions in turn.

In relation to our first question, children who were relatively strong word readers—as indexed by the Word Identification Test of the *WRMT*—did produce some lexicalisation error responses when reading simple monosyllabic nonwords. However, this type of error comprised on average less than one third of their total error responses. When they made an error, stronger word readers were far more likely to produce another, incorrect nonword than they were a word. As well, they were significantly *less* likely to make lexicalisation errors than were the weaker word readers in the sample. These results are therefore inconsistent with the proposal (Gibson & England, 2016; UKLA, 2012) that good word readers, who are progressing well in reading generally, are disadvantaged by the inclusion of nonwords in the Phonics Screening Check because they are inclined to substitute real words for nonwords. Rather, they are consistent with the conclusion of the official UK report on the first three years of the Phonics Screen, which states "there is no identifiable pattern of poorer performance on the PSC than expected in those children who are already fluent readers." (UK Department for Education, 2015, p. 25).

There were a very small number of children (n = 2) who were good word readers and made only a single error, but that one error was a lexicalisation. It is possible that students such as these are particularly noticeable to teachers in a classroom setting, leading to the impression that word reading skill may impair performance on nonwords. Considered as whole, however, these data overwhelmingly show that increased word reading ability is associated with less rather than more lexicalisation errors when reading nonwords.

It should also be noted that these results were obtained in a task where the non-words were presented to the children intermixed with words (the Castles and Coltheart test), such that they had no way of knowing which was which. In the English Phonics Screening Check, the word and nonword items are clearly delineated (with an alien picture next to each nonword), presumably making lexicalisation responses overall less likely than in the present study. A further point to note is that the present pattern of results was observed across three schools whose teaching methods were largely from a "whole language" tradition. This should have maximised the chance of observing evidence for children drawing on lexical knowledge when reading nonwords, and yet little evidence for this was found. Our expectation would be that children being taught via explicit phonics instruction methods would show, if anything, less evidence of an influence of lexical knowledge on their nonword reading errors.

Turning to our second question, a further implication of the claim that good word readers are inclined to make lexicalisation responses when reading nonwords is that they must be relying less on their phonics knowledge when performing the task than weaker word readers. However, our analysis of the similarity of the children's error responses to the target nonwords was also inconsistent with this suggestion. Stronger word readers were far more likely to make an error response that was very similar to the target than were weaker word readers. Such responses differed by only a single letter or phoneme from the correct target, suggesting that these children had simply made an error during the process of phonologically decoding the nonword. Even stronger support for this account comes from the finding that by far the predominant error for the stronger word readers was a highly-similar *nonword* response, comprising on average more than 50% of their total error responses. This is exactly the type of error that would be expected from a child attempting, if imperfectly, to read a nonword via the application of grapheme-phoneme correspondences.

Thus, we conclude that the good word readers in our sample were largely applying their phonics knowledge to read nonwords, and that they were doing so more successfully than the weaker word readers. This is not surprising examination of Table 1 reveals that, although selected on the basis of their word reading, these children were also on average better nonword readers than the weaker word readers; that is, they were readers who were progressing well in reading generally and who were advanced in both their lexical and phonics knowledge. It is important to note that there does exist a subset of children—those with pure *phonological dyslexia*—who have well-developed word reading skills but poor phonics knowledge (Castles & Coltheart, 1993; Petersen, Pennington & Olson, 2013). We would expect these children to show quite a different profile from the overall pattern showed by the good word reader group in our sample, obtaining low scores on nonword reading and producing the kinds of lexicalisation responses referred to by critics of the Phonics Screening Check. But the existence of such children does not challenge the present findings. On the contrary, our results converge on the following conclusion: when children have phonics knowledge, they typically apply this knowledge to the task of reading nonwords, regardless of their level of word reading ability; when they do *not* have phonics knowledge, as in the case of phonological dyslexia, they perform poorly on nonword reading tests and make errors indicative of the use of alternative strategies.

This brings us to the final question posed in the present study: is there evidence from our analyses that nonword reading tests are valid measures of phonic knowledge, in that the errors made by children on these tests are consistent with them having attempted to apply grapheme-phoneme correspondences? The answer here is a clear yes: children who had higher total scores on the nonword reading test tended to make errors that varied only marginally in letters and/or phonemes from the target items, and that were other nonwords rather than words. In contrast, children who had lower total scores on the nonword reading test were more likely to provide responses that were highly dissimilar from the target and that were words rather than nonwords. These results suggest that, at least for these early readers, nonword reading scores are governed largely by children's success in applying basic grapheme-phoneme correspondences.

We examined nonword error responses as a function of children's word reading ability. It is of course possible that the findings would have been different if we had indexed children's reading progress using another measure, such as reading comprehension. Perhaps children who are fluently reading for meaning in Year 1 or Year 2 are more inclined to draw on their lexical and semantic knowledge when reading nonwords? Although this is an empirical question, we think that this is unlikely to be the case: studies of reading acquisition show that children's reading comprehension in the early stages of reading is very strongly predicted by their word reading ability (e.g. Juel, 1988). Therefore, we would expect that those children in our sample who were reading single words at an above-average level for their age would also have been overall better-than-average in reading comprehension.

In summary, the error analyses reported here support the use of nonword reading tests to assess phonics knowledge in early readers. We find that these measures are a valid index of such knowledge, and find no evidence for the claim that strong word readers are disadvantaged by their use. Although there was considerable variability in the children's error responses, consistent with that reported in skilled readers (Coltheart & Ulicheva, 2018; Pritchard et al., 2012), overall the pattern of responses observed was best accounted for in terms of children attempting to apply their knowledge of grapheme-phoneme correspondences to read aloud the novel letter strings. Therefore, the use of a Phonics Screening Check, such as that currently used in England, is likely to be an effective means of tracking the acquisition of this important skill in children, and identifying those who require additional assistance.

Notes

1. Unfortunately, data were not available for Year 1 children. However, given the variability in reading skills in the sample, we believe the findings would be comparable across these year levels.
2. The proportions of each error type could not be compared directly as these are non-independent.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1052–1086.

Bowey, J. A., & Underwood, N. (1996). Further evidence that orthographic rime usage in nonword reading increases with word-level reading proficiency. *Journal of Experimental Child Psychology*, *63*, 526–562.

Brown, G. D. A., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*, *73*, 208–242.

Buckingham, J. (2016). *Focus on phonics: Why Australia should adopt the year 1 phonics screening check*. Retrieved from https://www.cis.org.au/app/uploads/2016/11/rr22.pdf

Castles, A., & Coltheart, M. (1993). Varieties of Developmental Dyslexia. *Cognition*, *47*, 149–180.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51.

Clark, M. M. (2016). Learning to be literate: Insights from research policy and practice. *Improving Schools*, *19*, 129–140.

Coltheart, M., & Ulicheva, A. (2018). Why is nonword reading so variable in adult skilled readers? *PeerJ*. doi:10.7717/peerj.4879

Coltheart, V., & Leahy, J. (1992). Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 718–729.

Elbro, C., de Jong, P. F., Houter, D., & Nielsen, A.-M. (2012). From spelling pronunciation to lexical access: A second step in word decoding? *Scientific Studies of Reading*, *16*(4), 341–359.

Gibson, H., & England, J. (2016). The inclusion of pseudowords within the year one phonics 'screening check' in English primary schools. *Cambridge Journal of Education*, *46*, 491–507.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447.

Lewis, M., & Ellis, S. (2006). *Phonics: Practice, research and policy*. London: Paul Chapman.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health & Development. Retrieved from https://www1.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

Peterson, R. L., Pennington, B., & Olson, R. K. (2013). Subtypes of developmental dyslexia: Testing the predictions of the dual-route and connectionist frameworks. *Cognition*, *126*, 2–38.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1268–1288.

Rack, J., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*, 28–53.

Rose, J. (2006). Independent review of the teaching of early reading final report. *Independent review*. Retrieved from www.standards.dcsf.gov.uk/phonics/report.pdf

Rowe, K. (2005). *Teaching reading: National inquiry into the teaching of literacy*. Training, Australian Council for Educational Research. Retrieved from https://research.acer.edu.au/tll_misc/5/

Schmalz, X., Marinus, E., Robidoux, S., Palethorpe, S., Castles, A., & Coltheart, M. (2014). Quantifying the reliance on different sublexical correspondences in German and English. *Journal of Cognitive Psychology*, *26*, 831–852.

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading*, *16*, 122–140.

U.K. Department for Education. (2015). *Phonics screening check evaluation: Final report*. Retrieved from https://www.nfer.ac.uk/publications/YOPC03/YOPC03.pdf

U.K. Department for Education. (2018). *Phonics screening check and key stage 1 assessments: England 2018*. Retrieved from https://www.gov.uk/government/statistics/phonics-screening-check-and-key-stage-1-assessments-england-2018

U.K. Literacy Association. (2012). *UKLA analysis of schools' response to the year 1 phonics screening check*. Retrieved from www.teachers.org.uk/files/y1psc-survey-october-2012.pdf

Woodcock, R. (1987). *Woodcock reading mastery tests – Revised*. Circle Pines, MN: American Guidance Service.

Appendix

The nonword items, with number of correct and number of different error responses

Nonword	Number of correct responses (/64)	Number of different Error responses
(1) baft	26	24
(2) bick	45	9
(3) borp	31	18
(4) brinth	33	17
(5) crat	38	12
(6) delk	37	20
(7) doash	13	18
(8) farl	36	20
(9) framp	34	22
(10) gop	48	9
(11) gurve	28	23
(12) hest	48	8
(13) norf	51	9
(14) peef	42	12
(15) peng	43	16
(16) pite	35	13
(17) pofe	20	23
(18) rint	46	6
(19) spatch	42	16
(20) trope	29	15