

British Educational Research Journal

**The Importance of Early Phonics Improvements for Predicting Later Reading Comprehension**

<sup>1</sup>Kit S. Double

<sup>1</sup>Joshua A. McGrane

<sup>1</sup>Jamie C. Stiff

<sup>1</sup>Therese N. Hopfenbeck

<sup>1</sup>Department of Education, University of Oxford, Oxford, UK

Correspondence to:

Kit Double

15 Norham Gardens, Oxford, OX26PY

Department of Education, University of Oxford

[kit.double@education.ox.ac.uk](mailto:kit.double@education.ox.ac.uk)

### Abstract

The role of phonics instruction in early reading development has been the subject of significant conjecture. Recently, England implemented a phonics screening check to assess the phonetic decoding of 6-year old students to ensure that all students master this foundational literacy skill and attain adequate phonemic awareness in the early years of primary schooling. Students who fail this check are obliged to retake the assessment the following year. In this article, we compare the performance of students who initially pass this check (pass) and students who fail the original assessment but pass the retaken assessment (fail-pass), with students who fail both the original and retaken assessments (fail-fail). Using data from the Key Stage 1 assessment of reading and the Progress in International Reading Literacy Study (PIRLS), we examined the reading comprehension performance of these students approximately 1 and 4 years after their first phonics screening. Results suggested that fail-pass students performed substantially better than fail-fail students, even after performance on the initial phonics check was controlled for. While fail-pass students do not appear to entirely catch up to pass students in reading comprehension, their relatively better performance underscores the importance of intervening for those students who are identified as having problems with phonetic decoding to increase their likelihood of success at reading comprehension in later schooling.

Keywords: reading comprehension; phonics; PIRLS; early intervention

## The Importance of Early Phonics Improvements for Predicting Later Reading Comprehension

The importance of sound blending and phonetic decoding for reading has long been established (Richardson, DiBenedetto, & Bradley, 1977), but its role in the teaching of reading has been contentious. The main focus of the so-called “reading wars” has been whether phonemic awareness, that is, the understanding of the relationship between letters (graphemes) and phonemes (sounds), is an essential precursor to the teaching of reading, or whether reading should instead be first taught at the whole-word level (Kim, 2008; Pearson, 2004). To some extent, a treaty has been called on this war with an increasing recognition in the research literature that the explicit teaching of phonics is an important component in the teaching of reading (Castles, Rastle, & Nation, 2018). This has been supported by evidence that early phonics interventions effectively promote reading development (Ehri, Nunes, Stahl, & Willows, 2001; Slavin, Lake, Davis, & Madden, 2011; Snowling & Hulme, 2012), and that phonemic awareness at early ages has moderate correlations with later literacy (Missall et al., 2007; Snowling, Lervåg, Nash, & Hulme, 2019). On the back of this debate, educational policy makers in England and elsewhere have moved to include phonics screening as part of national testing policy in the hope that early identification of students with phonics difficulties may facilitate early intervention and improve later reading performance. In this article, we examine whether ameliorating early phonetic decoding difficulties predicts improvements in reading comprehension at later stages of schooling.

### Phonics screening in England

The increased public and research acknowledgment of the importance of early phonics education prompted the English government to introduce a compulsory phonics check for state schools in 2012. The introduction of the phonics check was an extension of the English government’s push for teachers to focus on systematic synthetic phonics instruction and was seen as a way to ensure schools had a rigorous phonics program in place (Darnell, Solity, & Wall, 2017). In the phonics check, the child is assessed independently by their teacher and asked to read aloud 20 words, and 20 pseudowords. The inclusion of pseudowords is seen as important for ensuring that the children are relying on phonetic decoding skills, rather than their ability to recognise whole words that they are already familiar with (Standards & Testing Agency, 2017). To date, all versions of the phonics screening checks have required the student to correctly read 32 of the 40 words to have met the expected standard. Those students that do not meet this standard are required to take the phonics check again the following year, with their teachers being strongly encouraged to provide guided interventions to ensure that students reach the standard by their second attempt. This ‘retake’ has the overarching aim of ensuring that all children complete Key Stage 1 (the first two formal years of primary education in England) with sufficient phonemic awareness so that they can independently decode the text that they read through the rest of their primary education. Since the introduction of the assessment, the proportion of students reaching the expected standard on their first attempt has continued to rise, from 58% in 2012 to 82% in 2018 (Department for Education, 2018).

Nonetheless, the phonics check remains controversial. Although the check was piloted in the year prior to the first assessment (Coldwell et al., 2011), the Department for Education was criticised for first implementing the check without sufficient evidence of its predictive validity or formative effects (Goldstein, Moss, Sammons, Sinnott, & Stobart, 2018). While it has since been shown that the check has strong concurrent validity with other phonics assessments and broader measures of reading including comprehension tests, the need for a statutory assessment of phonics is still often questioned regardless (Duff, Mengoni, Bailey, & Snowling, 2015). The inclusion of pseudowords has also drawn criticism, as it has been argued that pseudowords are particularly challenging for students who are already capable of

reading for meaning (Coldwell et al., 2011). Furthermore, critics argue that the marking guidance provided to teachers for pseudowords is both overly restrictive in terms of acceptable pronunciations, and ambiguous about the allowances to be made for different accents (Gibson & England, 2016). Gilchrist and Snowling (2018) also found that performance on the phonics check correlates more strongly with measures of pseudoword reading rather than other more general measures of phonemic awareness that correlate more strongly with later reading comprehension. The check has also been criticised for being unfair to students who enter school at a younger age, as well as those from lower socioeconomic backgrounds, with both groups overrepresented in the group scoring below the expected standard (Clark, 2013).

With respect to the predictive validity of the phonics screening check, the need for longitudinal data has been recognised, but the opportunities thus far have been limited (Gilchrist & Snowling, 2018). However, a subset of students who sat the phonics check when it was first introduced have now completed two formal assessments of reading comprehension during primary school; the Key Stage 1 assessment of reading, taken near the end of Year 2 (age 6-7), and the Progress in International Reading Literacy Study (PIRLS), taken near the end of Year 5 (age 11-12). PIRLS is an international comparative study of the reading comprehension of primary-aged students, which in England, are representatively sampled from the population of Year 5 students across the country (McGrane, Stiff, Baird, Lenkeit & Hopfenbeck, 2017). The students that participated in PIRLS 2016 were among those that also sat the first phonics screening check in 2012. Therefore, both the Key Stage 1 assessment of reading and PIRLS 2016 present a unique opportunity to explore the longitudinal predictive validity of the phonics screening check. This, in turn, enables the examination of the impact of phonics interventions for students who initially fail to meet the expected standard on later reading comprehension performance.

### **Critical Periods and Early Phonics Intervention**

There is evidence that phonics interventions are more effective at improving reading performance during the early stages of reading development. For example, Ehri et al. (2001) found that the effect of phonics interventions was moderate ( $d = .55$ ) when performed before first grade, but only small when performed after first grade ( $d = .27$ ). For example, in a longitudinal experimental design, Hatcher, Hulme, and Ellis (1994) found that 7-year olds with reading difficulties benefited from reading interventions that incorporated explicit phonics teaching compared to reading interventions alone. Although phonological processing is important even for skilled readers, phonics appear to play an especially important role when children are first developing reading skills (Castles et al., 2018). While skilled readers are able to rapidly extract information from graphemes including phonological information (Rastle & Brysbaert, 2006), individuals who do not develop into skilled readers are less likely to routinely extract critical information from graphemes in adulthood (Veldre & Andrews, 2015, 2016). The importance of early intervention to improve phonics difficulties suggests that the phonics check in England, performed in Year 1, is well timed to identify children who are struggling with phonics so that steps can be taken during this critical period to remedy their phonics difficulties.

### **Research Questions and Hypotheses**

Importantly, the aim of this study is not to evaluate the impact of any particular phonics intervention, but instead to simply examine the longitudinal reading comprehension performance of students who are able to increase their phonetic decoding to the expected standard with those students that have equivalent difficulties, but are unable to improve their phonetic decoding performance to this

standard within a year<sup>1</sup>. Based on the importance of early phonemic awareness for later reading success and the importance of early phonics intervention for ameliorating long term effects on reading (e.g. Hatcher et al., 1994; Slavin et al., 2011; Snowling & Hulme, 2012), we hypothesise that students who fail the first phonics check but go on to pass the second phonics check will perform better at reading comprehension compared with those who do not meet the expected standard at both time points, even after controlling for baseline differences in phonics check performance.

## **Method**

### **Databases**

Data was sourced from the English Department for Education and included the National Pupil Database (NPD) and PIRLS 2016 results. The NPD is a database of student variables aged 2 to 21 in state education within the United Kingdom. Variables of particular interest (described below) in the NPD were the results of the two phonics checks and the KS1 Reading results. The NPD was matched with the PIRLS dataset using name and date of birth. Complete data was available for 4,569 students from 162 English schools (202 classrooms).

### **Measures**

#### ***Demographic variables***

Demographic variables are recorded in the NPD. We extracted the student's age (in months) and gender as well as their free school meals status (FSM). FSM is a widely-used proxy for socioeconomic disadvantage which has often been linked to educational underachievement (Taylor, 2018). Because of the longitudinal nature of this study's design, we used a measure of whether students were eligible for free meals in the last 6 years rather than their current FSM status.

#### ***Phonics Screening Check***

As previously mentioned, the phonics screening check was introduced in England in 2012 as a method to assess the phonetic decoding skills of students in Year 1 (6 to 7 years old). It is a school-based check that requires students to read aloud 20 words and 20 non-words (pseudowords). The check is administered by a teacher who performs the assessment without prompting or assisting the student. The check aims to identify students who are unable to decode the sounds or blend them correctly. As discussed above, students who correctly read at least 32 out of the 40 words were deemed to be reading at the expected standard. Those students who do not meet this standard (i.e., scored < 32) were obliged to retake the phonics check in Year 2 (McGrane, Stiff, Baird, Lenkeit, & Hopfenbeck, 2017).

#### ***Key Stage 1 Reading (KS1 Reading)***

Within the English educational context, KS1 refers to Years 1 and 2 (typically aged 5-7). The methods for assessing reading at the end of KS1 have changed multiple times in the past decade. The method used for assessing students in 2013, when the students included in this study were completing KS1, consisted solely of teacher judgments. Teachers were asked to rate their students into five competency levels, ranging from Level 1 (least competent) to Level 3 (most competent), with three sub-levels at Level 2 (C, B and A).

---

<sup>1</sup> While we adopt the idea put forth by the English government that a pass mark of 32 out of 40 in the phonics check signifies 'acceptable standards', we acknowledge that this appears to be an arbitrary cut-off and little justification for this particular criterion has been provided by those responsible for the phonics check.

## PIRLS

PIRLS (2016) is an international reading assessment directed by the International Association for the Evaluation of Educational Achievement (IEA) and taken by 50 countries. In England, PIRLS aims to assess the reading attainment and attitudes of Year 5 students. PIRLS requires students to answer a series of reading comprehension questions before filling out a number of self-report surveys about their attitude towards reading, language spoken at home, books at home and home possessions. In each country, a representative sample of students is selected to participate. We used the PIRLS data from 2016, as this was the first PIRLS assessment that could be linked to the phonics check (which was introduced in 2012), and thus PIRLS represents a standardized test of reading comprehension taken 4 years after the initial phonics check. The 2016 PIRLS dataset for England consisted of 5,095 Year 5 students from 170 primary schools. Like many international assessments, PIRLS utilises ‘plausible values’ based on multiple imputation estimation of the latent construct of interest (i.e. reading comprehension). For the analysis we used the first plausible value as the criterion variable.<sup>2</sup>

## Classification

Given the dichotomous (i.e., pass or fail) interpretation of the phonics check, we used the two assessment timepoints (first and retake) to form a 3-category classification of phonics performance. Students who passed the first phonics check (i.e., scored  $\geq 32$ ) were classified as the *pass* category. Students who did not pass the first phonics check, but then went on to pass the phonics check on their second attempt were classified as *fail-pass*. Finally, students who failed both the first and second phonics check were classified as *fail-fail*.

## Data Analysis

We sought to examine the association of phonics category with later measures of reading comprehension (KS1 and PIRLS reading scores). Of particular interest was the difference between the *fail-fail* and *fail-pass* categories, as these categories represent students who showed initial phonics deficits, with the differentiating factor being whether students were able to improve to the expected standard within 12 months or not.

For each of the criterion variables (KS1 reading, PIRLS reading) we first ran an *unconditional model* comparing the three groups on the criterion variable. We subsequently performed a *conditional model* comparing the *fail-pass* and *fail-fail* groups, conditioning on their first phonics check score. This allowed us to compare students with equivalent initial phonics check performance who either progressed to meet the expected standard by the time of the retaken phonics check (*fail-pass*), or those who did not (*fail-fail*). We did not include the pass group in the conditional model, as there is a logical dependence between the pass group category and the conditioning variable (first phonics check score), because students in this category did not retake the phonics check. Finally, we re-analysed both models using *contextualised models* that included additional covariates, including age, gender, and FSM status, to control for these extraneous variables that are also known to be associated with phonics performance on entry to school and later reading comprehension performance (McGrane et al., 2017; Tymms, Merrell, Hawker, & Nicholson, 2014).

---

<sup>2</sup> The PIRLS data set provides 5 plausible values for each student’s performance but using one plausible value is known to provide unbiased estimates of population parameters (Jerrim, Lopez-Agudo, Marcenaro-Gutierrez, & Shure, 2017; OECD, 2009).

We performed multilevel models using the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.4.3 (R Core Team, 2017). A two-level random intercepts model was used, with random intercepts for school and classroom entered into the model. In addition, because the KS1 reading ability score was not continuous (students were rated against criteria on a 5-level scale), we applied an ordinal multilevel model for this criterion variable, which again accounted for the clustering of students in classrooms and schools.

## Results

### Descriptive Statistics

The phonics performance categorisation was determined using the above classification procedure into *pass* ( $n = 2,798$ ), *fail-pass* ( $n = 1,273$ ), and *fail-fail* ( $n = 498$ ). Descriptive means and bivariate correlations are presented in Table 1. As shown in Figure 1, the distribution of phonics check scores is non-normally distributed. For example, 13.96% of students received a mark of 32-33 whereas only 4.09% of students received a mark of 30-31. Therefore, Spearman's rank correlations were used when phonics check scores were included in the examination of association between variables.

[insert Figure 1 approximately here]

As shown, both the first and second phonics check scores showed moderate to large correlations with the KS1 and PIRLS reading marks, with the correlation slightly smaller between PIRLS (approximately 4 years later) and the phonics check compared with KS1 reading (approximately 1-2 years later) and the phonics check. Together these results suggest that the phonics check is a significant predictor of later reading comprehension performance, albeit one that dissipates over time. In terms of demographic differences, age was negatively correlated with phonics performance. Furthermore, males showed significantly poorer performance than females on the first phonics check ( $t(4639) = 6.10, p < .001, d = .49$ ), but not the second phonics check ( $t(1798) = 1.03, p = .303, d = .05$ ). Additionally, students with FSM eligibility had significantly poorer performance on both the first ( $t(4639) = 15.30, p < .001, d = .18$ ) and second ( $t(1798) = 5.56, p < .001, d = .27$ ) phonics checks.

### Key-stage 1 Results

Multilevel ordinal models relevant to KS1 reading are presented in Table 2. 19 participants who had incorrectly entered KS1 Reading data were removed from these analyses.

#### **Unconditional Model**

Performance on KS1 reading as a function of phonics check performance category are presented in Figure 2. A multilevel model was performed with KS1 reading level as the criterion variable and phonics check category as the predictor variable. The results indicated that the *fail-pass* were significantly more likely to receive a higher grade in KS1 reading than the *fail-fail* group ( $OR = .17, z = 15.64, p < .001$ ). Furthermore, the *pass* group was significantly more likely to receive a higher grade in KS1 reading than the *fail-pass* group ( $OR = 13.20, z = 33.91, p < .001$ ). For completeness, we re-ran the model with the *fail-fail* group as the reference category. The results suggested that students who passed the initial phonics check were significantly more likely than the *fail-fail* group to have a higher KS1 reading grade,  $OR = 76.17 (z = 35.89, p < .001)$ .

#### **Conditional Model**

As expected, the *fail-pass* group outperformed the *fail-fail* group on the first phonics check ( $\beta = -0.35, t = 15.69, p < .001$ ). This is most likely the result of *fail-pass* students being closer to the pass cut-

off to begin with thereby being more likely to pass on their second attempt. As such, it is possible that any differences in subsequent reading success are driven by these initial differences, rather than the lessening of phonics difficulties between the first and second phonics checks. Therefore, we ran a conditional model comparing the performance of the *fail-fail* group and the *fail-pass* group, but conditioning on students' first phonics check performance. Essentially this model tests, controlling for baseline phonics check performance, whether those that went on to pass the second phonics check show at higher grades of reading comprehension at KS1 compared to those students who do not go on to pass the second phonics check. The results suggest that, even after controlling for baseline phonics check performance, the *fail-pass* group was significantly more likely to have a higher grade in KS1 reading than the *fail-fail* group ( $OR = 0.32, z = 9.35, p < .001$ ).

### **Contextualised Models**

Finally, we re-ran both the unconditional model and the conditional model with additional demographic variables (age, gender, FSM status) entered as covariates. As shown in Table 2 all results were practically unchanged, suggesting that resolution of phonics deficits is an important predictor of future reading performance, even after these background factors are controlled for.

## **PIRLS Results**

### **Unconditional Models**

PIRLS reading performance as a function of phonics check performance category is presented in Figure 3. The unconditional model again showed that the *fail-pass* group significantly outperformed the *fail-fail* group ( $\beta = -0.18, t = 12.80, p < .001$ ). Furthermore, the *pass* group performed significantly better than the *fail-pass* group ( $\beta = .34, t = 23.12, p < .001$ ). Re-running the model with the *fail-fail* group as the reference group indicated that the *pass* group significantly outperformed the *fail-fail* group ( $\beta = .62, t = 29.81, p < .001$ ), as expected.

### **Conditional Models**

We again ran a model that conditioned on performance on the first phonics check. As with KS1 Reading, this model showed that even after students' baseline phonics check performance was controlled for, the *fail-pass* group significantly outperformed the *fail-fail* group in PIRLS reading comprehension performance ( $\beta = -.16, t = 7.10, p < .001$ ).

### **Contextualised Models**

As with KS1 Reading, we re-ran all PIRLS models with age, sex, and FSM status as covariates in the model (see Table 3). Including these covariates in the models did not substantially change the results for the phonics check performance groups.

## **Discussion**

This study was the first to systematically examine the relative performance of students who pass versus fail the retake phonics check. When considering students who fail the original phonics assessment, our results show that even for children with the same level of phonics difficulties at the time of the first phonics assessment, those who go on to reach adequate standards by the time of the second phonics check fair comparably well four years later. These results suggest that it is critically important that phonics difficulties are addressed and alleviated rapidly in order for reading performance in later schooling years.

The phonics screening check was implemented in 2012 to identify children at risk of developing difficulties with later reading comprehension. Despite an increasing recognition of the benefits of the explicit teaching of phonics, the push for systematic synthetic phonics proposed by the Department for Education in England, and particularly the use of the phonics screening check, have remained controversial. The present results firstly suggest that the phonics check predicts long-term reading comprehension ability, supporting the predictive validity of the check. Many of the early criticisms of the phonics screening check were concerned with the lack of evidence of predictive validity. That is, whether a student's performance in a test of phonetic decoding in year 1 of primary school would be able to predict their reading outcomes later in school. The results of this study support other preliminary studies of the predictive validity of the phonics check (Duff et al., 2015). Importantly, this study is the first that provides evidence of predictive validity of the phonics check at a point where the vast majority of students are now reading to learn, as opposed to learning to read (Mullis, Martin, Foy, & Hooper, 2017).

Secondly, these findings suggest that the phonics check may be a valuable tool for identifying children in need of early intervention and that those students who are able to ameliorate their early phonics difficulties fair substantially better in later reading tests compared with children with equivalent difficulties who do not improve in the first year of schooling. These findings support the idea that there is a critical period for the development on phonetic decoding skills around the period of development that occurs at school entry (Ehri et al., 2001; Castles et al., 2018). It should, however, be noted that we do not have evidence of how much additional support the children who failed the first phonics check received in the year before the retaken assessment, but these results suggest that the progress made by students in the year between the two checks has long term implications on reading comprehension development.

Often critics of large-scale national assessments have argued that they offer little formative value, that is they do not provide adequate feedback to teachers and students that would allow them to modify teaching and learning in a manner that would improve future academic outcomes (Sellar & Lingard, 2014). The phonics check introduced in England has been introduced explicitly as a formative tool, which is designed to identify students who need additional phonics support (Department for Education, 2018). The current results suggest that the phonics check can be an important tool for diagnosing early difficulties and monitoring progress in the amelioration of those difficulties. The implementation of large scale *formative* assessments is still relatively rare; by assisting teachers in identifying students with phonics difficulties and providing clear guidance for additional assistance to assist students who are identified as having these deficits, the implementation of the phonics check provides a clear example of how large-scale standardized assessments can be used in a formative way to improve academic outcomes.

While the current results suggest that ameliorating phonics deficits early is important for closing the gap in later reading comprehension, the current data cannot speak to what steps educators took to address this gap or the efficacy of various phonics interventions. Phonics interventions of course differ in efficacy and while the English government prescribes additional support and intervention for students who fail the first phonics check, it is largely agnostic about the form that this additional support should take. Evidence suggests that incorporating phonics into reading teaching is more effective than reading alone at improving reading ability, but phonics teaching alone is more effective at improving phonics ability (Hatcher et al., 1994). It is therefore important that teachers incorporate any additional phonics teaching within reading instruction for students with phonics deficits, rather than focusing on phonics in an isolated manner.

The current results suggest that ameliorating early phonics difficulties predicts better reading comprehension up to four years later. Importantly, students who overcome early deficits in phonetic decoding are able, at least in part, catch up to their peers who do not present phonics difficulties. These results underscore the importance of utilising large-scale national assessments in a formative fashion, that is for intervention and support rather than simply as a tool for diagnosis and ranking.

## References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, 67(1), 1-48.
- 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. (2013). The phonics check for Year 1 children in England: Unresolved issues of its value and validity after two years. *Education Journal*, 177, 13-15.
- Coldwell, M., Shipton, L., Stevens, A., Stiell, B., Willis, B., & Wolstenholme, C. (2011). *Process evaluation of the Year 1 Phonics Screening Check Pilot* (DFE-RR159). Retrieved from Sheffield, UK:
- Darnell, C. A., Solity, J. E., & Wall, H. (2017). Decoding the phonics screening check. *British educational research journal*, 43(3), 505-527.
- Department for Education. (2018). National curriculum assessments at key stage 1 and phonics screening checks in England, 2018. Retrieved from <https://www.gov.uk/government/publications/phonics-screening-check-and-key-stage-1-assessments-england-2018/national-curriculum-assessments-at-key-stage-1-and-phonics-screening-checks-in-england-2018>
- Duff, F. J., Mengoni, S. E., Bailey, A. M., & Snowling, M. J. (2015). Validity and sensitivity of the phonics screening check: implications for practice. *Journal of Research in Reading*, 38(2), 109-123.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of educational research*, 71(3), 393-447.
- 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(4), 491-507. doi:10.1080/0305764X.2015.1067289
- Gilchrist, J. M., & Snowling, M. J. (2018). On the validity and sensitivity of the phonics screening check: erratum and further analysis. *Journal of Research in Reading*, 41(1), 97-105.
- Goldstein, H., Moss, G., Sammons, P., Sinnott, G., & Stobart, G. (2018). *A baseline without basis: The validity and utility of the proposed reception baseline assessment in England*. London: British Educational Research Association
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child development*, 65(1), 41-57.
- Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. J. E. o. E. R. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. 61, 51-58.
- Kim, J. S. (2008). Research and the reading wars. In F. M. Hess (Ed.), *When research matters: How scholarship influences education policy* (pp. 89-111). Cambridge, MA: Harvard Education Press.
- McGrane, J., Stiff, J., Baird, J.-A., Lenkeit, J., & Hopfenbeck, T. N. (2017). *Progress in International Reading Literacy Study (PIRLS): National Report for England*. Oxford, UK: Oxford University Centre for Educational Assessment (OUCEA)
- Missall, K., Reschly, A., Betts, J., McConnell, S., Heistad, D., Pickart, M., . . . Marston, D. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review*, 36(3), 433.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Retrieved from <http://timssandpirls.bc.edu/pirls2016/international-results/>
- OECD. (2009). *PISA 2006 Technical Report*. . OECD Publishing.
- Pearson, P. D. J. E. p. (2004). The reading wars. 18(1), 216-252.

- R Core Team. (2017). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
- Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, 53(2), 97-145.
- Richardson, E., DiBenedetto, B., & Bradley, C. M. (1977). The relationship of sound blending to reading achievement. *Review of educational research*, 47(2), 319-334.
- Sellar, S., & Lingard, B. J. B. E. R. J. (2014). The OECD and the expansion of PISA: New global modes of governance in education. 40(6), 917-936.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, 6(1), 1-26.
- Snowling, M. J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders*, 47(1), 27-34.
- Snowling, M. J., Lervåg, A., Nash, H. M., & Hulme, C. (2019). Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Developmental science*, 22(1), e12723.
- Standards & Testing Agency. (2017). Assessment framework for the development of the Year 1 phonics screening check.
- Taylor, C. (2018). The Reliability of Free School Meal Eligibility as a Measure of Socio-Economic Disadvantage: Evidence from the Millennium Cohort Study in Wales. *British Journal of Educational Studies*, 66(1), 29-51. doi:10.1080/00071005.2017.1330464
- Veldre, A., & Andrews, S. (2015). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 586.
- Veldre, A., & Andrews, S. (2016). Semantic preview benefit in English: Individual differences in the extraction and use of parafoveal semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(6), 837.

Table 1.

Descriptive statistics and bivariate Pearson correlations for study variables

Statistic	N	Mean	St. Dev.	Min	Max	Phonics Mark 2 <sup>3</sup>	KS1 Reading	PIRLS Reading	Engage	Liking	Confidence	Sex	Age	Free Lunch
Phonics Mark 1 <sup>4</sup>	4,641	29.739	9.871	0	40	0.50***	0.70***	0.52***	0.01	0.10***	0.36***	-0.06***	0.17***	-0.21***
Phonics Mark 2 <sup>4</sup>	1,800	32.632	7.17	0	40		0.54***	0.38***	0.02	0.03	0.27***	-0.01	0.02	-0.11***
KS1 Reading	4,622	16.787	3.563	3	27			0.64***	0.01	0.11***	0.43***	-0.11***	0.19***	-0.25***
PIRLS Reading	4,641	557.6	76.757	240.48	776.03				0.04*	0.17***	0.48***	-0.10***	0.11***	-0.24***
Sex	4,641	1.494	0.5	1	2								0	-0.01
Age	4,641	10.254	0.294	9.58	11.08									-0.01
Free Lunch	4,641	0.293	0.455	0	1									1

<sup>3</sup> Spearman rank correlations were used for correlations involving phonics mark 1 and 2 due to the non-normal distribution of the phonics check marks.

Table 2.

Ordinal Multilevel models analysing the relationship between phonics category and KS1 Reading. *Fail-pass* is set as the reference group for all models

Predictors	Unconditional Model				Conditional Model				Contextual Model			
	OR	b	OR CI	p	OR	b	OR CI	p	OR	b	OR CI	p
Fail-Fail	.17	-1.75	.14, .22	<.001	3.13	1.14	2.46, 3.97	<.001	.18	1.74	.14, .22	<.001
Pass	13.2	2.58	11.37, 15.32	<.001					11.8	4.2	1.15, 13.72	<.001
Phonics mark 1					1.2	.19	1.19, 1.22	<.001				
Sex									.66	-.41	.59, .74	<.001
Age									2.61	.96	2.15, 3.18	<.001
Free meal									.53	-.64	.46, .60	<.001

Note: OR = odds ratios

Table 3.

Multilevel models analysing the relationship between phonics category and PIRLS Reading. *Fail-pass* is set as the reference group for all models

	Unconditional Model				Conditional Model				Contextualised Unconditional Model				Contextualised Conditional Model			
<i>Predictors</i>	<i>b</i>	$\beta$	<i>CI</i>	<i>p</i>	<i>b</i>	$\beta$	<i>CI</i>	<i>p</i>	<i>b</i>	$\beta$	<i>CI</i>	<i>p</i>	<i>b</i>	$\beta$	<i>CI</i>	<i>p</i>
(Intercept)	-.35			<.001	1.23			<.001	-1.88			<.001	-2			.005
Pass	.69	.34	.31, .36	<.001					.63	.31	.28, .34	<.001				
Fail-fail	-.59	-.18	-.21, -.16	<.001	-.35	-.16	-.21, -.12	<.001	-.56	-.18	-.20, -.15	<.001	-.34	-.16	-.20, -.12	<.001
Phonics mark 1					.04	.35	.31, .39	<.001					.04	.33	.28, .37	<.001
Sex									-.14	-.07	-.10, -.05	<.001	-.08	-.04	-.08, -.00	.042
Age									.17	.05	.02, .07	<.001	.09	.03	-.01, .07	.187
Free Meal									-.27	-.12	-.15, -.10	<.001	-.21	-.11	-.15, -.06	<.001
<b><u>Random Effects</u></b>																
$\sigma^2$		.7				.66				.69				.66		
$\tau_{00}$		.06 IDCLASS				.07 IDCLASS				.05 IDCLASS				.07 IDCLASS		
		.02 IDSCHOOL				.03 IDSCHOOL				.01 IDSCHOOL				.02 IDSCHOOL		
ICC		.07 IDCLASS				.09 IDCLASS				.07 IDCLASS				.09 IDCLASS		
		.03 IDSCHOOL				.04 IDSCHOOL				.02 IDSCHOOL				.03 IDSCHOOL		

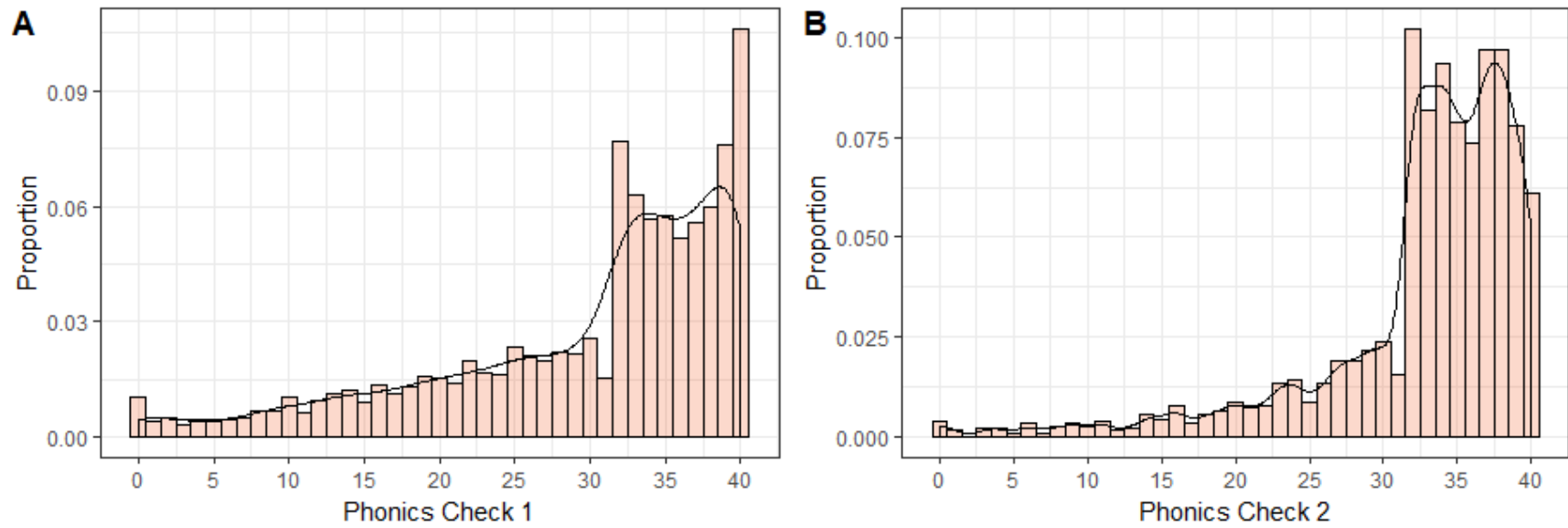


Figure 1. Distribution of scores on phonics check 1 and 2. Note, a pass mark of 32 is set.

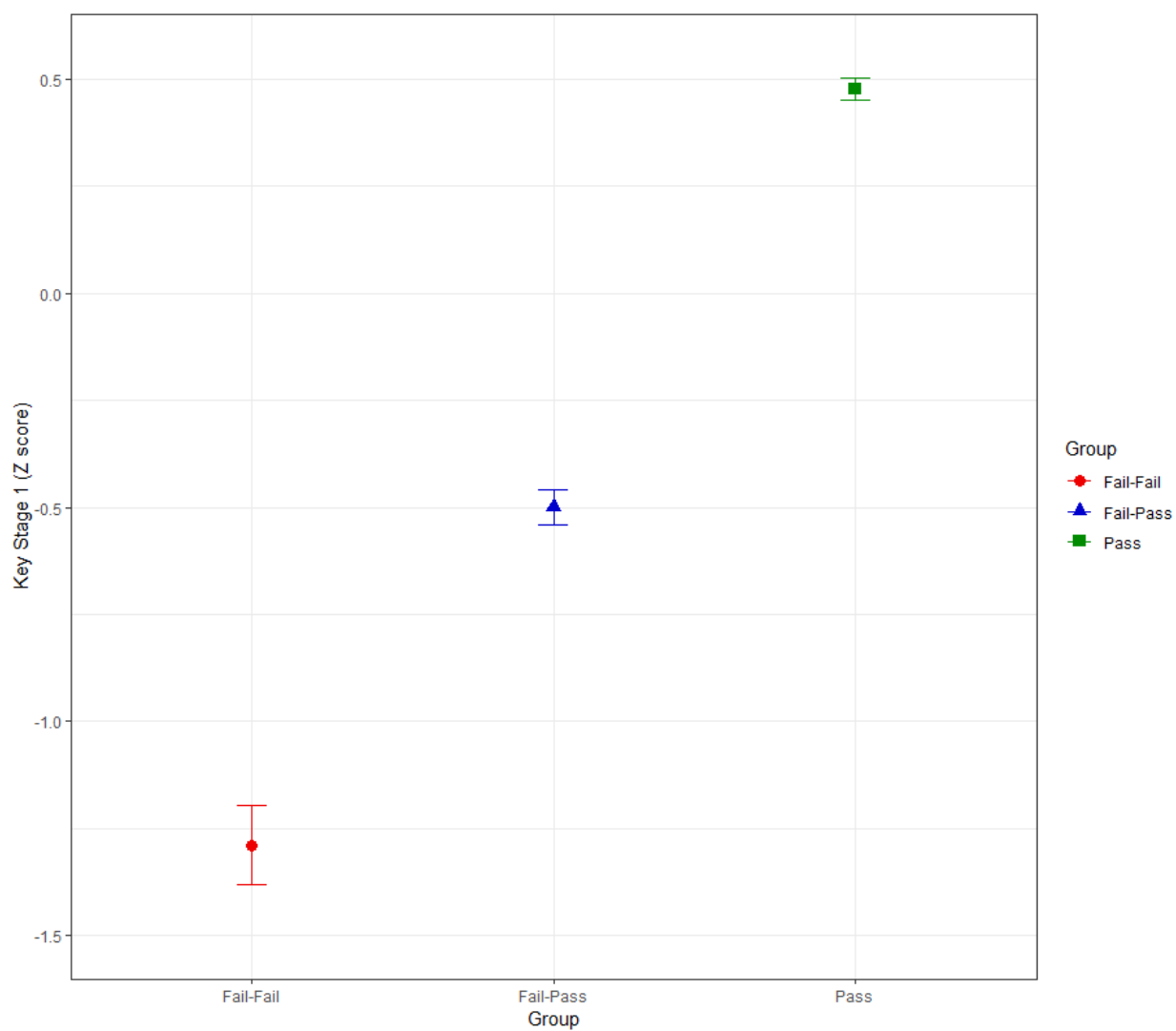
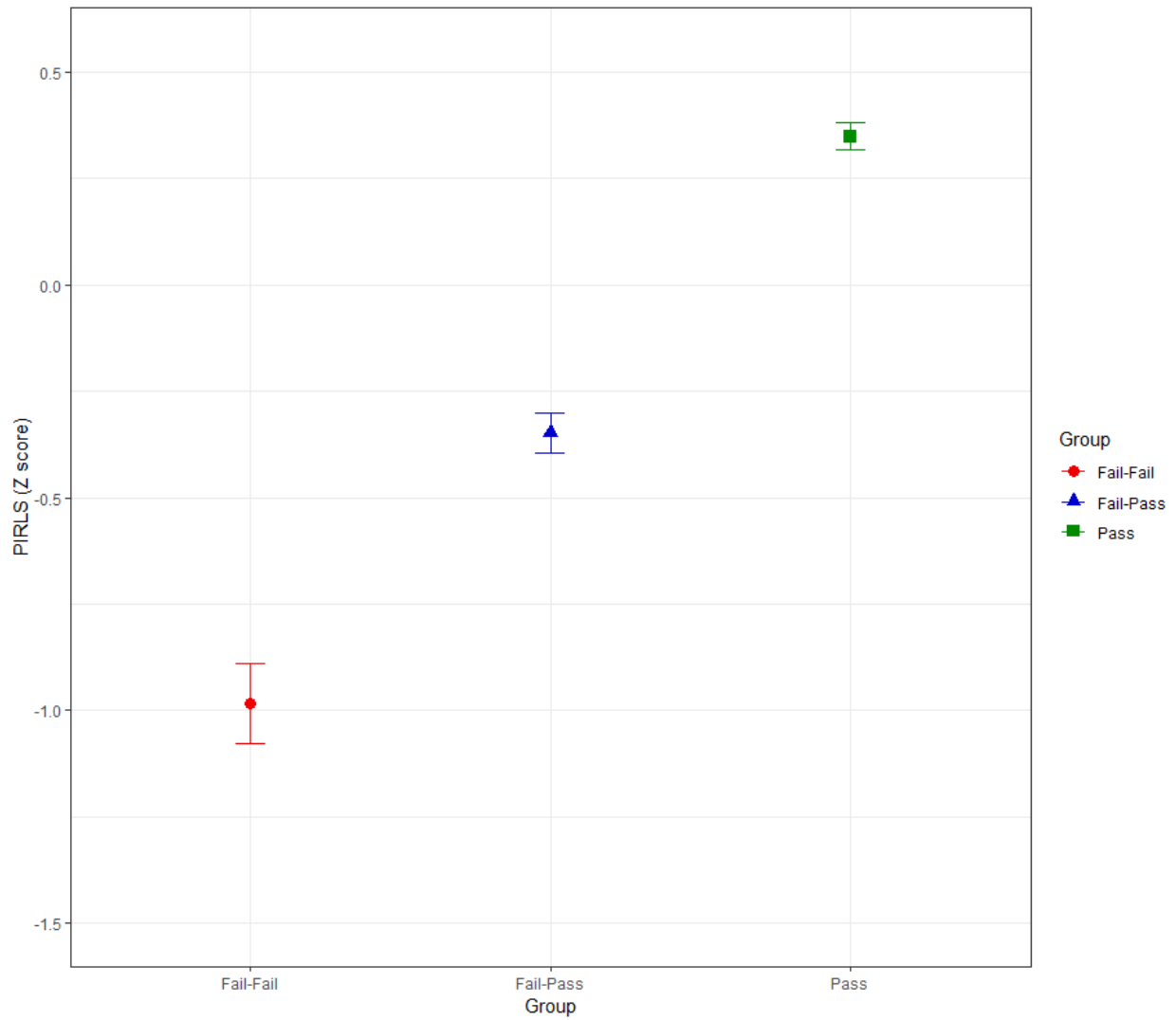


Figure 2. KS1 Reading performance as a function of phonics category. Error bars show the 95% confidence interval.



*Figure 3.* PIRLS Reading performance as a function of phonics category. Error bars show the 95% confidence interval.