

There is still little or no evidence that systematic phonics is more effective than common alternative methods of reading instruction: Response to Brooks (2023).

Jeffrey S. Bowers  
School of Psychological Science  
University of Bristol,  
12a Priory Road, Bristol, BS8-1TU  
j.bowers@bristol.ac.uk

Key Words: Phonics; Reading instruction; Whole language; PIRLS

Acknowledgements: I would like to thank Dr. Patrica G. Bowers and Dr. Peter N. Bowers for commenting on an earlier draft of this work.

### **Abstract**

Brooks (2023) rejects Bowers' (2020) conclusion that there is little or no evidence that systematic phonics is more effective than alternative teaching methods common in schools. He makes his case based on challenging my analysis of 4 or the 12 meta-analyses reviewed in Bowers (2020). I show his criticisms are flawed and conclusions are unwarranted. I also briefly review the more recent PIRLS results that have been taken to support the claim that mandated synthetic systematic phonics has improved reading comprehension in England. This conclusion is also shown to be unjustified. I conclude there is still no reliable evidence that systematic phonics is best practice, that researchers should stop making strong claims based on the current evidence, and that the field needs to explore alternative approaches.

### **Context and implications:**

Brooks (2023) challenged the analysis of Bowers (2020) and concluded there is good evidence that systematic phonics is more effective than common alternative approaches to reading instruction. I show that all his points are mistaken and conclude there is still no reliable evidence that systematic phonics is best practice. It is important to correct the widespread mischaracterisation of the current evidence because it makes it more difficult for researchers and educators to develop and assess alternative approaches to reading instruction.

Both Bowers (2020, 2021) and Wyse and Bradbury (2022) challenged the evidence taken to support the conclusion that systematic phonics is more effective than alternative forms of instruction common in schools (such as whole language and balanced literacy). In this journal, Brooks (2023) claimed that these “critiques fail in their attempts to show that the evidence on phonics is unreliable”. Here I focus on Brooks’ critique of Bowers (2020) and show that his claims are mistaken and conclusion unjustified.

In the Bowers (2020) article I reviewed all meta-analyses designed to assess the efficacy of phonics (12 in total) and the reading outcomes in England following the legal mandate to teach phonics in state schools in 2007. To avoid any confusion, I did not claim there is good evidence for alternative forms of instruction common in school (whereas Wyse and Bradbury, 2022, argue that the evidence supports ‘balanced instruction’ that characterized most instruction in England prior to 2007). Rather, my point was (and is) that the strong claims regarding the efficacy of systematic phonics compared to other forms of instruction are unjustified, and that more research should be devoted to considering alternative approaches to reading instruction, such as Structured Word Inquiry (Bowers & Bowers, 2017; Bowers & Kirby, 2010) that teaches grapheme-phoneme correspondences in a morphological and etymological context.

Brooks’ response only considered four of these 12 meta-analyses. First, he criticized my analysis of Galuschka et al. (2014) who reported similar effect sizes for phonics ( $g' = 0.32$ ), phonemic awareness instruction ( $g' = 0.28$ ), reading fluency training ( $g' = 0.30$ ), auditory training ( $g' = 0.39$ ), and color overlays ( $g' = 0.32$ ). Despite the similar effect sizes, Galuschka et al. concluded that phonics was the “most” effective method because only phonics had a significant effect (because the phonics condition included many more studies). I noted that this conclusion requires significant statistical interactions, with larger effects of phonics compared to the other methods. But the interactions were not reported, and they would not be significant given the similar effect sizes.

In response, Brooks argues that it is not appropriate to look for interactions between different forms of instruction as there was only one independent variable in the study, namely, type of instruction. On his view, an interaction requires to independent variables, and here there is only one. This is mistaken. If you want to call the different forms of instruction different levels of a single variable (type of instruction) you still need to determine whether the different levels interact, with better outcomes for phonics. And the answer is clear: There is no statistic that can show that an effect size of  $g' = 0.32$  for phonics is larger than an effect size of  $g' = 0.32$  for colour overlays, for example. Of course, this is not an endorsement for colour overlays, or any of the alternative methods. It is simply a statistical point: The results from Galuschka et al. (2014) do not support the conclusion that phonics is more effective than these alternative methods. One might appeal to other empirical findings or make a theoretical case for why phonics is preferable to colour overlays, but this meta-analysis does not provide evidence for this conclusion.

Next, Brooks criticizes my brief analysis of Han’s (2009) unpublished PhD thesis. Again, I concluded that this study failed to provide evidence that systematic phonics is more effective than alternative approaches because similar results were reported across a range of different forms of instruction, with effect sizes as follows: phonics (.33), phonemic awareness (.41), fluency (.38), vocabulary (.34), and comprehension (.32). Brooks was critical of my analysis because I failed to mention a flaw in the study, writing “[Han] gives a list of 11

‘Instructional activities’ which are classified as phonics—but only one, or at most two, deserve that label.” But this is irrelevant to my conclusion. If Han (2009) meta-analysis is flawed it remains the case that it provides no evidence that phonics is more effective than alternative methods. Indeed, if the Han (2009) misclassified many of these studies, not only is the .33 effect for phonics not larger than the alternative methods, but it is also a flawed analysis that should not be used to draw any conclusions.

Next, he criticizes my brief analysis of another unpublished PhD thesis concerned with reading instruction for struggling readers in Grades 5-12 (Sherman, 2007). I claimed it failed to obtain a significant impact of phonics on reading outcomes whereas Brooks claims it did, writing:

“The top line of data in Sherman’s table 20 (p. 69) shows an ES of 0.33 for the impact of phonics on literacy overall. The confidence interval for the ES is given as 0.13 Lower, 0.52 Upper; since this does not cross zero, the ES must be significant at least at  $p < 0.05$ , even though Sherman does not give a probability value or discuss this result.”

Brooks is correct regarding the top line of Table 20 in the Sherman (2007) thesis, but on the 6<sup>th</sup> line of the same table, that summarizes the outcomes when outlier studies are excluded, the confidence intervals do overlap with zero, showing no significant effect.

Is it appropriate to exclude the outlier studies? Yes it is. Sherman (2007) herself excluded three studies that reported effect sizes of 2.9, 5.15 and 7.69. No intervention could plausibly obtain such large effect sizes. More critically, the meta-analysis does not provide evidence for phonics even if the outlier studies are included, as made clear from the following passage from the thesis:

“Because of the small number of studies and the variability of the population studied, the alpha level was relaxed to 0.25 to explore statistical significance of main effects or interaction effects at this level. The impact of group size and reading level on effect size was significant in many of the analyses at a 0.25 alpha level.”

So, even if we consider the full dataset that includes a study with a Cohen  $d$  of 7.69, the analysis only shows that phonics had a significant effect at the .25 level.

Finally, Brooks criticizes my analysis of Camilli et al. (2003) who had re-analyzed the studies from the National Reading Panel (2000) or NRP meta-analysis. Camilli et al. noted that the NRP compared systematic phonics to a control condition that included a heterogeneous set of studies, some of which included no phonics and others that included unsystematic phonics (as characteristic of whole language and balanced literacy). Camilli et al. noted that this control condition is inappropriate if the question one wants to address is whether systematic phonics is more effective than the forms of instructions common in schools (forms of instruction that include some degree of unsystematic phonics, as noted by the NRP itself). When Camilli et al. (2003) compared systematic phonics to studies that included unsystematic phonics, the effect size was roughly half the size reported in the NRP ( $d = .24$  vs  $d = .41$ ). Interestingly, the analysis also found significant and numerically larger effects of systematic language activities ( $d = 0.29$ ) and tutoring ( $d = 0.40$ ). A subsequent

meta-analysis of the NRP studies by Camilli et al. (2006) that considered additional moderator variables revealed an even smaller effect of systematic phonics ( $d = 0.12$ ) that was no longer significant. Furthermore, as noted by Bowers (2020), even these small effect sizes are overestimates of the impact of systematic phonics instruction compared to non-systematic phonics: These effects are largely driven by word and nonword decoding measures and are reduced for word reading accuracy (a measure of regular and irregular word naming), fluency, and reading comprehension, the effects reflect the short-term rather than long-term impacts of systematic phonics, only 13 of the 38 studies in the NRP used randomized controlled designs, and there is evidence for publication bias amongst these RCT studies, amongst other problems. All these factors will have inflated the small effects observed in the Camilli et al. (2003, 2006) re-analyses.

In his response, Brooks (2023) criticized how Camilli et al. (2003) classified studies in the control condition as containing unsystematic and no phonics, and wrote:

When I pointed out the fragility of Camilli et al.'s (2003) analysis, Bowers (personal communication, 9 March 2023) replied:

'My critique does not hinge on the Camilli et al. findings (there is little evidence for phonics even if you ignore his [sic] point).' Despite what he says, Bowers' argument does in fact make considerable use of 'the Camilli et al. findings'.

Brooks has selectively quoted my email to give the impression that I conceded his point regarding Camilli et al. (2003). I did not. Here is what I wrote:

"My critique does not hinge on the Camilli et al. findings (there is little evidence for phonics even if you ignore his point), but I don't understand your criticism of the study. There are essentially no forms of instruction used in school that use NO phonics, so studies that completely ignore phonics are not appropriate to include in a control condition if you want to claim that systematic phonics is needed to improve existing classroom instruction."

It is reasonable to ask whether Camilli et al. were able to reliably classify studies from the NRP as containing no phonics vs unsystematic phonics (sometimes the relevant information was not clearly provided in the studies), but the fact remains that the control condition in the NRP included a heterogeneous set of studies, including studies that included no phonics. Accordingly, the control condition in the NRP does not allow an assessment of systematic phonics compared to common alternative methods used in schools. To provide a more straightforward comparison, Bowers (2020) identified all the studies from the NRP that specifically compared synthetic systematic phonics (the form of systematic phonics mandated in England) to whole language. Only four studies included in the NRP made this comparison, and the effect sizes in order of magnitude were  $d = 0.91$ ,  $d = 0.12$ ,  $d = 0.07$ , and  $d = -0.47$ . Furthermore, none of these four studies employed randomized controlled designs, nor assessed the long-term effects of systematic phonics. Clearly, these four studies provide little evidence that systematic synthetic phonics is more effective than whole language.

More generally, researchers should stop citing the NPR as providing strong evidence for systematic phonics given that is composed of studies that are all now 25 years old or older

and there are many more recent meta-analyses that incorporate more recent research. Not only do these more recent meta-analyses provide little or no support for the claim that systematic phonics is more effective than a mixture of alternative methods when it comes to reading accuracy, fluency, or reading comprehension, they all ignored the important point of Camilli et al. regarding control conditions. Consequently, they do not even test the claim that systematic phonics is more effective than the alternative forms of instruction common in schools.

Given that recent meta-analyses do not compare systematic phonics to common alternative methods, perhaps the most relevant evidence comes from the experience in England where systematic phonics has been mandated in state schools since 2007, and where a Phonics Screening Check was introduced in 2012 that assesses whether phonics is being effectively implemented in schools. Although Brooks (2023) did not consider this work, Bowers (2020, 2021) reviews the evidence and shows that systematic phonics is indeed being well implemented in schools (the results of the Phonics Screening Check have gone up), but there is little or no evidence that phonics has improved reading outcomes in England as measured by Standard Assessment Tests (SAT), Progress in International Reading Literacy Study (PIRLS), and Program for International Student Assessment (PISA) tests.

Since Bowers (2020, 2021), the results of the PIRLS 2021 have recently been released (Lindorff, Stiff, & Kayton, 2023). This is a test of reading comprehension, and England ranked 4<sup>th</sup>, up from 8<sup>th</sup> in 2016. Several authors have claimed that the excellent performance in this last round supports the conclusion that systematic synthetic phonics has improved the reading outcomes. For instance, professor Kathy Rastle in a podcast entitled “Has England just become a reading superpower”, says “the power of [phonics] instruction really shines through in the PIRLS results” See: <https://open.spotify.com/episode/5fx3JlkALT6eQtEj8LCUxH>. (go to: 8:36). Similarly, Buckingham (2023) writes:

“The UK Government made the Year 1 Phonics Screening Check mandatory in English schools in 2012. There is good evidence suggesting that the Phonics Screening Check played a significant role in England’s improved performance in the most recent PIRLS assessment”.

Do the recent PIRLS results provide evidence that the mandated systematic phonics and the Phonics Screening Check have improved reading outcomes in England? Here are the results across all years, with the date followed by the score: 2001: 553; 2006: 539; 2011: 552; 2016: 559; 2021: 558. Not only are the PIRLS results not improved in the most recent round (the score goes down by one point), with little improvement reported between 2001 (pre-mandated phonics) and 2021, but the biggest improvement in PIRLS (between 2006-2011) predates the introduction of the PSC in 2012. In addition, Singapore, Ireland, and Northern Ireland have consistently outperformed England when assessing children in English despite not legally requiring phonics nor adopting the Phonics Screening Check. Part of the reason why England went up in the most recent rankings is that Ireland and Northern Ireland were excluded from the comparison. In fact, both countries scored better in the 2021 PIRLS but were excluded because they delayed PIRLS testing by a few months – due to COVID – and consequently, the children were a few months older. It is also the case that England delayed PIRLS testing for a full year due to COVID so that children were of the appropriate age (most countries carried out PIRLS in 2021 as originally planned), and this may also have

contributed to England's higher ranking. Indeed, Lindoff et al. (2023) in the "PIRLS 2021: National Report for England Research report" write:

The global COVID-19 pandemic had an impact on data collection for PIRLS as well as on normal school operations in England and other education systems internationally. The available data do not allow us to measure that impact in a precise way, which complicates both international comparisons and trends over time for PIRLS in the 2021 cycle.

There is also a conceptual problem with Rastle's and Buckingham's interpretation of the results. England scored better than Italy and Spain (and many other countries) that have writing systems with consistent grapheme-phoneme correspondences. These children would score near 100% (at a younger age) if they were presented with a Phonics Screening Check in their native language (see Seymour, Aro, & Erskine, 2003). However effective mandated phonics has been in improving the naming of English regular words (and nonwords), English children are still not as good as Italian and Spanish children at naming words (and nonwords) in their own languages. Accordingly, the higher English scores in PIRLS (that measures reading comprehension) must reflect something other than phonics. This is worth exploring. But the alternative reason(s) why children in England have done so well on PIRLS over many years is not being explored because the results are being attributed to systematic synthetic phonics.

In summary, Brooks only challenged my interpretation of four of the 12 meta-analyses reviewed in Bowers (2020), and as shown above, they all support my conclusion: The Galuschka et al. (2014), Han (2010), and Sherman (2007) meta-analyses provide no evidence that systematic phonics is more effective than common alternative methods, and Camilli et al. (2003)'s point regarding the control condition of the NRP is correct and applies to all subsequent meta-analyses (other than Camilli et al., 2003, 2006). Consequently, all the meta-analyses taken to support systematic phonics in the "reading wars" are not even designed to test whether systematic phonics is more effective than the methods commonly used in schools. Finally, the recent claim that mandated systematic synthetic phonics in England has improved reading outcomes on PIRLS 2021 is again unjustified.

To reiterate, this is not an endorsement of whole language, balanced literacy, let alone colour overlays. Rather, my claim is that the field has seriously mischaracterized the strength of evidence for systematic phonics. If the research community continues to claim that the "science of reading" strongly supports systematic phonics (e.g., Buckingham, 2020; Crawford et al, 2023; Fletcher et al., 2021), then there should be some persuasive responses to the various problems that have been identified with the evidence (Bowers, 2020, 2021; Bowers & Bowers, 2021; Wyse, & Bradbury, 2022). And if the evidence for systematic phonics is much weaker than widely claimed, then it is important to acknowledge this because it makes it more difficult to obtain funding and publish research that focuses on various alternative promising alternatives, including Structured Word Inquiry (Bowers & Bowers, 2017, 2018; Bowers & Kirby, 20210).

### References:

- Bowers, J.S. (2020) Reconsidering the Evidence that Systematic Phonics is more Effective than Alternative Methods of Reading Instruction. *Educational Psychology Review*, 32, 681-705.
- Bowers, J.S. (2021). Yes children need to learn their GPCs but there really is little or no evidence that systematic or explicit phonics is effective: A response to Fletcher, Savage, and Sharon (2020). *Educational Psychology Review*. doi.org/10.1007/s10648-021-09602-z
- Bowers, J. S., & Bowers, P. N. (2017). Beyond phonics: the case for teaching children the logic of the English spelling system. *Educational Psychologist*, 52, 124–141. <https://doi.org/10.1080/00461520.2017.1288571>.
- Bowers, J. S., & Bowers, P. N. (2018). Progress in reading instruction requires a better understanding of the English spelling system. *Current Directions in Psychological Science*, 27(6), 407-412.
- Bowers, J.S., & Bowers, P.N. (2021). The science of reading provides little or no support for the widespread claim that systematic phonics should be part of initial reading instruction: A response to Buckingham. doi.org/10.31234/osf.io/f5qyu
- Bowers, P. N., & Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing: An Interdisciplinary Journal*, 23, 515–537. <https://doi.org/10.1007/s11145-009-9172-z>.
- Buckingham, J. (2020). Systematic phonics instruction belongs in evidence-based reading programs: A response to Bowers. *The Educational and Developmental Psychologist*, 37(2), 105-113.
- Buckingham, J. (2023). Response to the NSRA consultation paper Review to Inform a Better and Fairer Education System. Retrieved from: <https://multilit.com/wp-content/uploads/2023/08/NSRA-Submission-MultiLit-2023.pdf>
- Brooks, G. (2023). Disputing recent attempts to reject the evidence in favour of systematic phonics instruction. *Review of Education*, 11(2), e3408.
- Camilli, G., Vargan, S., & Yurecko, M. (2003). Teaching children to read: the fragile link between science and federal education policy. *Education Policy Analysis Archives*, 11(15), 1–51.
- Camilli, G., Wolfe, M., & Smith, M. L. (2006). Meta-analysis and reading policy: perspectives on teaching children to read. *The Elementary School Journal*, 107(1), 27–36. <https://doi.org/10.1086/509525>.

- Crawford, M., Raheel, N., Korochkina, M., & Rastle, K. (2023, September 15). Global Literacy Goals Constrained by Inadequate Foundational Decoding Skills for Pupils in Low- and Middle-Income Countries. <https://doi.org/10.31234/osf.io/2qxm9>
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PLoS One*, 9(2), e89900. <https://doi.org/10.1371/journal.pone.0089900>.
- Han, I. (2009). Evidence-based reading instruction for English language learners in preschool through sixth grades: a meta-analysis of group design studies. Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/54192>.
- Lindorff, A., Stiff, J., & Kayton, H. (2023). PIRLS 2021: National Report for England. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1156633/PIRLS\\_2021\\_-\\_national\\_report\\_for\\_England\\_\\_May\\_2023.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1156633/PIRLS_2021_-_national_report_for_England__May_2023.pdf)
- 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. Bethesda, MD: National Institute of Child Health and Human Development.
- Seymour, P. H. K., Aro, M. and Erskine, J. M. 2003. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94: 143–174.
- Sherman, K. H. (2007). A Meta-analysis of interventions for phonemic awareness and phonics instruction for delayed older readers. University of Oregon, ProQuest Dissertations Publishing, 2007, 3285626.
- Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading. *Review of education*, 10(1), e3314.