



# 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)

Jeffrey S. Bowers<sup>1</sup>

Accepted: 28 January 2021 / Published online: 13 March 2021  
© The Author(s) 2021

## Abstract

Fletcher, Savage, and Sharon (Educational Psychology Review, 2020) have raised a number of conceptual and empirical challenges to my claim that there is little or no evidence for systematic phonics (Bowers, Educational Psychology Review, 32, 681–705, 2020). But there are many mistakes, mischaracterizations, and omissions in the Fletcher et al. response that not only obscure the important similarities and differences in our views but also perpetuate common mischaracterizations of the evidence. In this response, I attempt to clarify a number of conceptual confusions, perhaps most importantly, the conflation of phonics with teaching GPCs. I do agree that children need to learn their GPCs, but that does not entail a commitment to systematic or any other form of phonics. With regard to the evidence, I respond to Fletcher et al.'s analysis of 12 meta-analyses and briefly review the reading outcomes in England following over a decade of legally mandated phonics. I detail why their response does not identify any flaws in my critique nor alter my conclusion that there is little or no support for the claim that phonics by itself or in a richer literacy curriculum is effective. We both agree that future research needs to explore how to combine various forms of instruction most effectively, including an earlier emphasis of morphological instruction, but we disagree that phonics must be part of the mix. I illustrate this by describing an alternative approach that rejects phonics, namely, Structured Word Inquiry.

**Keywords** Phonics · Systematic phonics · Grapheme-phoneme correspondences · GPCs · Structured Word Inquiry · SWI

I am pleased that Fletcher et al. (2020) have responded to my article entitled “Reconsidering the Evidence That Systematic Phonics Is More Effective Than Alternative Methods of

---

✉ Jeffrey S. Bowers  
j.bowers@bristol.ac.uk

<sup>1</sup> School of Psychological Science, University of Bristol, 12a Priory Road, Bristol BS8-1TU, UK

Reading Instruction” published in this journal (Bowers 2020). I hope this exchange will encourage researchers to more fully consider the evidential basis for the widespread claim that systematic phonics is effective. I am also pleased that there are important points of agreement between us. For example, we both agree that the strength of evidence for phonics has been exaggerated in some circles, and most importantly, we are both advocating an approach of instruction that makes morphology an important part of early reading instruction.

But there are many mistakes, mischaracterizations, and omissions in the Fletcher et al. response that not only obscure the important similarities and differences in our views but also perpetuate common mischaracterizations of the evidence (also see Buckingham 2020a, b and response by Bowers and Bowers 2021). I expect many readers will take this commentary not only as a challenge to many of my claims but also as a justification to continue to emphasize systematic phonics at the early stages of reading instruction rather than consider new approaches. I show that these conclusions are unjustified. Apart from our contrasting characterizations of the evidence, there are important differences in the new directions we are advocating. It is important to highlight these differences in order to move the field forward with a firm understanding of the data and a clear understanding of the alternatives.

My response is organized as follows. First, I address a set of conceptual issues that I think have been misunderstood by Fletcher et al. This includes a misunderstanding of the goal of my critique and a conflation of phonics with teaching GPCs. Second, I respond to their characterization of the meta-analyses they take to support systematic phonics. Third, I briefly summarize the impact of requiring systematic phonics in all English state schools since 2007. This was a key part of the Bowers (2020) review but not considered by Fletcher et al. (2020). I briefly revisit these findings because they speak to one of the main claims of Fletcher et al., namely, the value of combining systematic phonics with additional forms of reading instruction. Finally, I outline the different possible ways forward given the limited evidence for phonics thus far, including teaching GPCs in a qualitatively different way to phonics.

## **What Was My Main Motivation for Critiquing Systematic Phonics and What Conclusions Do I Draw?**

Fletcher et al. characterized my review as being primarily concerned with whether phonics or whole language is more effective, and they argue that contrasting the superiority of phonics versus alternative reading methods is the wrong question for twenty-first century science of teaching. But this is a misunderstanding. In fact, I was first motivated to look into the evidence for systematic phonics because of the difficulties I had in publishing a paper co-authored with my brother that detailed an alternative to both systematic phonics and whole language called Structured Word Inquiry or SWI (Bowers and Bowers 2017). Central to SWI is the claim that children should learn how their writing system works, including GPCs and the meaningful sub-lexical regularities in spellings (through morphology and etymology), and how these regularities interact, in order to teach reading, spelling, and vocabulary at the same time. My (subjective) take on the difficulty in publishing our work was that reviewers were so committed to the importance of phonics that they were not open to considering alternative approaches, even an alternative that rejects whole language and balanced literacy.

And that is why it is so important to recognize that the evidence for phonics is weak: it should motivate researchers to look beyond the phonics/whole language debate. And although Fletcher et al. and I agree that phonics is not enough, our positions are different. Fletcher et al.

(and almost all advocates of phonics) argue that phonics is necessary, but not sufficient. The question from this perspective is what needs to be added to phonics to most effectively teach children to read (for an excellent summary of this position, see Castles et al. 2018). My position is that although there is indeed good evidence that teaching GPCs is essential, there is little evidence that phonics is effective, either by itself or in combination with other forms of instruction. My conclusion is that more research should be carried out on alternative approaches such as SWI that reject phonics altogether (e.g., Bowers and Bowers 2017, 2018a). This claim is often dismissed by researchers and teachers because they conflate phonics with teaching GPCs, as described next.

## What is Systematic Phonics?

The authors of the National Reading Panel (2000), or NRP, assessed the effectiveness systematic phonics that they contrasted with unsystematic phonics. Shanahan (2005), one of the members of the NPR, described this distinction as follows: “*Systematic phonics* is the teaching of phonics with a clear plan or program, as opposed to more opportunistic or sporadic attention to phonics in which the teacher must construct lessons in response to the observed needs of children.” [italics in original] (p. 11). This is an important distinction because most alternative methods of reading instruction, including whole language and balanced literacy, include unsystematic phonics (National Reading Panel, 2000).

What is unclear in this definition, and the NRP itself, is whether the “systematic” in “systematic phonics” also requires a set of letter-sound correspondences to be taught in a fixed order, a practice often described as “scope and sequence.” In fact, the authors of the NRP considered a wide range of different types of phonics as systematic, including synthetic phonics, analytic phonics, embedded phonics, analogy phonics, and onset-rime phonics. Although the different methods all teach letter-sound correspondences explicitly, they vary in the degree to which the correspondences are taught in a specific sequence. Fletcher et al. are right to point out that the NRP did not find any significant differences between the different types of systematic phonics, and accordingly, there is little evidence that a planned sequence of teaching GPCs is relevant to the outcomes (the same conclusion was more recently drawn by Castles et al. 2018). Given the lack of evidence that GPCs or other letter-sound correspondences should be taught in a fixed order, Fletcher et al. argue that the NRP and subsequent meta-analyses provide evidence for “explicit phonics” as opposed to systematic phonics. I adopted the term “systematic” phonics following the terminology of the NPR and all subsequent meta-analyses described below, but I am happy with the term “explicit” phonics as well. The change in terminology makes no difference to my analyses of the outcomes of meta-analyses.

A key point emphasized by Bowers (2020) is that all forms of systematic (or explicit) phonics teach letter-sound correspondences before and independently of the meaning-based regularities in spellings (morphology and etymology), what Bowers and Bowers (2018b) called the “phonology first” assumption. That is, teaching of GPCs in phonics is not informed by the fact that English spellings encode both phonological and meaningful regularities, with morphology constraining GPCs. The centrality of this claim to phonics is not only highlighted by the many statements that morphological instruction should follow phonics (e.g., Adams 1990; Castles et al. 2018; Ehri and McCormick 1998; Larkin and Snowling 2008; Frith 1985; Rastle 2019; Rastle and Taylor 2018; for a challenge to these claims, see Bowers and Bowers

2018b) but also by the fact that not a single phonics intervention in all the meta-analyses described below taught children the interaction between morphology and phonology. In addition, advocates of systematic phonics often justify the phonology first hypothesis on the basis of the “simple view of reading” (Hoover and Gough 1990), according to which children need to first decode written words before accessing word meanings via their verbal language system, and on the basis of the “alphabetic principle” according to which children need to first “crack the alphabetic code” that maps graphemes to phonemes (e.g., Castles et al. 2018).

But phonics is not the only way to explicitly teach GPCs. For example, Structured Word Inquiry (SWI) teaches GPCs in the context of morphology in an attempt to teach GPCs more effectively. SWI is not an example of adding morphology after teaching GPCs, or even an example of teaching GPCs and morphology at the same time but independently. Rather, it is an example of teaching GPCs and morphology together, from the start, at the same time, because they interact in ways that make sense of GPCs. To illustrate, consider the word <does> that is considered an “exception” in the phonics perspective (<does> should rhyme with “foes,” “goes,” “hose,” “nose,” “pose,” “toes,” etc.). SWI highlights how <does> has a consistent spelling that maintains the spelling of its base morpheme (the spelling of <do> is preserved in <does>, <doing>, and <done>). A SWI lesson might begin with a conversation about the meanings of words <does> and <do> in the context of spoken sentences (“She **does** her work,” and “I **do** my work”) and then use simple graphical methods (e.g., the word sums and morphological matrices) to highlight the consistent spelling-meaning connections in morphological families. In this context, children can learn the mapping between <d> → /d/, that the <o> grapheme can write more than one phoneme, and that although the <s> grapheme can spell the phoneme /s/ in words like <spell> or <cats>, a common job of the <s> grapheme is to spell the /z/ phoneme in words like <dogs>. All GPCs can be taught in this way while children learn word spellings and vocabulary at the same time (for illustration of how SWI can be taught with young children, see Anderson et al. (2019); for a theoretical motivation of SWI that rejects the “alphabetic principle,” see Bowers and Bowers (2018a)).

In sum, systematic phonics is committed to the claim that GPCs and other letter-sound correspondences are first taught in isolation of the meaningful spelling regularities of English, whereas SWI is committed to teaching phonological and semantic regularities together because they interact with semantic regularities helping to make sense of GPCs. Distinguishing between these different approaches of teaching GPCs is not a terminological quibble: Phonics and SWI adopt a different understanding of the English writing system and employ different tools to teach GPCs in very different ways. And whereas the goal of phonics is to teach children to decode words in isolation of the meaningful regularities in spellings before moving on to other forms of instruction (ideally, phonics is finished after grade 2 or 3 and is then discarded), SWI is a form of instruction that addresses phonology, spelling, vocabulary, and meaning in combination that can be used at any stage of reading instruction.

The distinction between teaching GPCs with and without phonics is often missed because researchers often use the term “phonics” in two different ways, namely, as a form of instruction as well as a form of knowledge (knowledge of letter-sound correspondences). At the moment of writing, the phrase “phonics knowledge” receives almost two-thousand hits in *Google Scholar*. Fletcher et al. appear to do this as well, writing: “Intervention research is needed on the “optimal” components of phonics, morphology, and other aspects of language for a range of learner groups.” Of course, if the term phonics is used to refer to knowledge of letter-sound correspondences, then by definition, phonics is necessary. But it does not follow from this that phonics instruction is necessary. And more important for present purposes, it obscures two

qualitatively different ways forward: initially teaching GPCs and other letter-sound correspondences independently of meaningful spelling regularities (as is the case with all existing phonics interventions) and teaching the phonological and semantic regularities of English spellings at the same time from the start (as is the case in SWI). Here, and in Bowers (2020), I use the term phonics to refer to a form of instruction.

## A Review of the Meta-analyses on Phonics

Given the widespread claim that phonics is a necessary component of reading instruction, there should be good evidence that phonics does indeed improve reading outcomes. Fletcher et al. agree that some researchers have exaggerated the strength for phonics, but at the same time, they claim that the meta-analyses do provide strong evidence that phonics is effective. I will briefly consider Fletcher et al.'s claims regarding the individual meta-analyses, but first let me make one general comment that applies across all the meta-analyses. Fletcher et al. criticize my focus on whether the meta-analyses show significant effects or not, writing:

Yet throughout his paper, Bowers presents conventions for the interpretation of effect sizes, sometimes drawing attention to their statistical significance as crucial and sometimes not, but never to the confidence intervals that surrounds effect sizes. Bowers does not consistently acknowledge that these conventions are arbitrary and must be contextualized. The real issue is their replicability, their practical significance given an estimated counterfactual, and their precise role in reading instruction.

I do not understand this point. In all cases, I have considered a main effect or an interaction as significant if it reaches the conventional value of  $p < 0.05$ . I do agree with Fletcher et al. that a small effect size can have an important practical effect when considering a large population of children, but that requires that the small effect is real (a nonsignificant effect size of  $d = 0.15$  should not be treated the same as an effect size of  $d = 0.15$ ,  $p < 0.001$ ). Fletcher et al.'s comment that findings should be interpreted in context is correct, and here is some of the context that should be considered: (1) many of the small effects reported in the meta-analyses are obtained in the context of publication bias and poor research practices (e.g., many studies are not randomized control trials, few studies are double blind, there are multiple mistakes regarding what studies should be included/excluded); (2) many of the strong conclusions taken from the meta-analyses are the product of the authors highlighting some findings and ignoring others in the abstracts and executive summaries of the meta-analyses; (3) there is a growing realization in psychology and other sciences that a significant effect of  $p < 0.05$  is an unreliable measure of whether an effect is likely to be replicated; (4) the observation that some small nonsignificant effects are repeated across multiple meta-analyses may reflect, at least in part, the fact that the same studies are often used across meta-analyses; and (5) the conclusion that there is little or no evidence for phonics does not rule out the possibility that phonics is necessary, but it does challenge the widespread view that there is currently strong evidence for this claim.

**National Reading Panel (2000)** Fletcher et al. responded to four points I made regarding the NPR results. First, based on the failure to observe significant short- or long-term benefits of phonics for “low achieving” readers ( $d = 0.15$ ), I concluded that the NRP provided no evidence that phonics helped the majority of struggling readers above grade 1. Fletcher rejected this

conclusion, writing “ $d=0.15$  is for a minority of children with lower IQ scores and low reading from grades 2 to 6, not for the effect of phonics instruction overall or even for older poor readers as a whole.” But this ignores the fact that the majority of children struggling to read in grades 2–6 will have lower than average IQs (due to the comorbidity of various learning difficulties) and struggling readers in grades 2–6 were classified as “low achieving” in the NRP even when their IQ was not assessed. It follows that the NRP provided no evidence that phonics helps the majority of struggling readers above grade 1. Second, I noted that the authors of the NRP were misleading in their executive summary where they wrote “Phonics instruction taught early proved much more effective than phonics instruction introduced after first grade” (pp. 2–93). Fletcher et al. agree that this summary was misleading, but then note that my criticism of the NRP is unwarranted because the authors of the NRP do acknowledge (elsewhere) that there was not enough research to make any strong conclusions on this point. I agree that the data do not support any strong conclusions regarding when phonics is most effective, but still, the misleading executive summary is problematic given that the NRP is frequently cited as evidence that it is important to start phonics early.

Third, I noted that the long-term effects of phonics were greatly reduced following a delay of 4–12 months, and further, that the long-term effects of spelling, reading texts, or reading comprehension were not even assessed. Fletcher et al. respond by contextualizing my comments, writing again that the data were limited, and concluded: “Perhaps the most important point is that the long-term effect sizes are positive and practically significant on the primary outcome.” But my point stands: the “primary outcome” was an amalgam of all reading outcomes, and accordingly, no evidence was reported for long-term effects on spelling, reading text, or comprehension.

Finally, in response to my claim that the evidence was even weaker when comparing phonics to whole language, Fletcher et al. note that the overall effects were still significant ( $d = 0.31$ ). But the fact remains that for this comparison, there was no evidence for short-term benefits for spelling, reading text, or comprehension (the effects were not even reported). And although we agree that the evidence for synthetic systematic phonics was “very modest” (only four relevant studies were included in the NRP meta-analysis, with effect sizes of  $d = 0.91$ ,  $d = 0.12$ ,  $d = 0.07$ , and  $d = -0.47$ ), Fletcher et al. minimize the significance of this, suggesting that this observation is largely relevant to “some consumers of it in some UK policy circles.” Apart from the fact that the NRP has been a key motivation for requiring the teaching of systematic synthetic phonics to millions of children in England (Rose, 2006), it is currently being used to support similar use of systematic synthetic phonics in Australia (Buckingham 2020a, b).

But even this weak evidence for systematic phonics from the NRP is further undermined by later work, most strikingly by the Camilli et al. (2006) and Torgerson et al. (2006) meta-analyses.

**Camilli et al. (2003, 2006)** Camilli et al. (2003, 2006) noted that the main analysis of the NRP compared systematic phonics interventions to a control condition that included two different types of interventions, namely, interventions that employed no phonics and interventions that included some phonics but taught unsystematically. As noted by Camilli et al., this design choice is problematic if researchers want to conclude from the NRP that phonics should replace alternative approaches common in schools given that these alternatives almost always include some degree of unsystematic phonics. In order to test this hypothesis, the control condition should exclude the intervention studies that included no phonics. When Camilli et al. (2006) carried out a new meta-analysis that included new covariates (including the degree to

which phonics was taught systematically), they failed to obtain evidence that systematic phonics was more effective than unsystematic phonics, although the short-term effect size (averaging over all measures) was positive ( $d = 0.123$ ,  $p > 0.05$ ). Fletcher et al. challenged the claim that unsystematic phonics is commonplace in schools, claiming I reached this conclusion based on a quotation from “one whole language scholar” and “one UK schools inspectors report.” I do not know who they are referring to with regard to the whole language scholar, but my main evidence comes from the NRP itself that I quoted in Bowers (2020). I will repeat part of the quote here (the first sentence in the following quote was not included in Bowers, 2020 but it is in the NRP report):

Whereas in the 1960s, it would have been easy to find a 1st grade reading program without any phonics instruction, in the 1980s and 1990s this would be rare. Whole language teachers typically provide some instruction in phonics, usually as part of invented spelling activities or through the use of graphophonemic prompts during reading... However, their approach is to teach it unsystematically and incidentally in context as the need arises.

Fletcher et al. also write that I have not added “any new substantive points to the discussion” compared to Camilli et al. (2006). However, a central point in my review is that similar control conditions were used for all subsequent meta-analyses, and accordingly, they do not even test the hypothesis that systematic phonics is more effective than the main alternatives common in schools. Nevertheless, this is the conclusion that is routinely drawn both by the authors of the meta-analyses and the researchers who cite this work.

**Hammill and Swanson (2006)** The key point of Hammill and Swanson (2006) was that the effect sizes reported in the NRP were often quite small. Again, Fletcher et al. highlight that small effects can be important when considering a large population, and I agree. The problem is not with the size of the effects, it is the fact that there is no statistical evidence that the small effects on spelling, reading texts, or reading comprehension extended beyond 4 months, or that phonics improved reading outcomes for the majority of struggling readers above grade 1.

**Torgerson et al. (2006)** Perhaps the most striking demonstration that the NRP provides little or no basis for concluding that phonics should be introduced in schools comes from the Torgerson et al. meta-analysis. This meta-analysis focused on the subset of studies in the NRP that used randomized control designs. Here, I agree with some of Fletcher et al.’s summary:

After assessing the evidence contextualized against rigorous inclusion criteria including randomization, Torgerson et al. (2006, p. 42) argue, “none of the findings of the current review were based on strong evidence because there simply were not enough trials (regardless of quality or size)” before drawing extremely cautious conclusions. The precise wording of the primary conclusion was that there was “No warrant for NOT using phonics” (p. 43).

In other words, when the studies with the best designs from the NRP are analyzed, Fletcher et al. and I agree that there is almost no evidence for systematic phonics. It is worth emphasizing that the NRP has been cited over 23,000 times, and the peer-reviewed article that specifically focused on the phonics chapter from NRP (Ehri et al., 2001) has been cited over 1000 times (almost 200 times since 2019).

However, Fletcher et al. are incorrect when they claim that Torgerson et al.'s primary conclusion was that there was "No warrant for NOT using phonics." This quote is taken from Table 4 on page 43. Their primary conclusion comes from the conclusion section of the executive summary on page 10 where they write:

Systematic phonics instruction within a broad literacy curriculum appears to have a greater effect on childrens' progress in reading than whole language or whole word approaches. The effect size is moderate but still important.

**McArthur et al. (2012)** The McArthur et al. meta-analysis was designed to assess the efficacy of systematic phonics with children, adolescents, and adults with reading difficulties. The authors reported significant effects of word reading accuracy and nonword reading accuracy, whereas no significant effects were obtained in word reading fluency, reading comprehension, spelling, and nonword reading fluency. My claim was that the significant word reading accuracy results depended on two studies by Levy and colleagues (Levy et al. 1999; Levy and Lysynchuk 1997) that were designed in such a way to inflate effect sizes, and once these two studies are removed, the only remaining significant effect was nonword reading accuracy. Fletcher et al. criticized the rejection of these two studies, writing:

Levy and colleagues studies are certainly not alone among the studies in this review in using bespoke researcher-developed outcome measures of grapheme-phoneme knowledge, so these two studies should not be excluded based on outcome measure used.

But it was the design of the studies, not the use of a bespoke outcome measure, that was the problem. Consider the Levy and Lysynchuk (1997) study in which children were taught to name a set of words with an "an" rime (e.g., <can>, <ban>, <man>, <pan>) and then tested on a new word that shared this rime (e.g., <fan>). The fact that study and test words were so similar means that the results should not be taken as evidence that systematic phonics improves word reading accuracy in general (the claim of the McArthur et al. meta-analysis). The same approach was taken in Levy et al. (1999).

To illustrate the problem, consider the Phonics Screening Check in England that assesses how well children have learned GPCs by asking children to read a set of words and nonwords aloud (discussed in more detail below). Teachers are not allowed to know the words (and nonwords) on the test so they cannot "cheat" by training children on the same or similar items. By contrast, in the two Levy et al. studies, the researchers were aware of the test words, and furthermore, the study-test words were designed to overlap as much as possible. To be fair to Levy et al., the authors were comparing the effectiveness of various forms of phonics instruction (and they failed to find an advantage of synthetic over onset-rime phonics as commonly claimed; e.g., Brady, 2020; Wheldall and Buckingham, 2020) and were not claiming that the effect sizes they observed in their studies would generalize to classroom settings. Fletcher et al. (correctly) note that I failed to report on the updated McArthur et al. (2018) meta-analysis that included three additional studies (increasing from 11 to 14 studies). But it is important to note that both meta-analyses compared small group phonics instruction to a control condition that included \*no\* instruction. So again, neither meta-analysis even tested the hypothesis that systematic phonics is more effective than standard alternative teaching methods.



**Galuschka et al. (2014)** Galuschka et al. carried out a meta-analysis that assessed the impact of various forms of instruction on children with reading difficulties and concluded that phonics was the most effective. I noted that the effect size for phonics ( $g' = 0.32$ ) was similar to the outcomes with phonemic awareness instruction ( $g' = 0.28$ ), reading fluency training ( $g' = 0.30$ ), auditory training ( $g' = 0.39$ ), and color overlays ( $g' = 0.32$ ), and that the only reason why their meta-analysis obtained a significant effect for phonics and not the other measures was that there were many more studies in the phonics condition. Hatcher et al. responded:

This is a genuinely startling conclusion. The combined evidence from a meta-analysis of 29 RCTs on phonics reported by Galuschka et al. is of a qualitatively different kind to the evidence from other trials (e.g., two on medical treatments, three on colored overlays, auditory training, and comprehension). Bowers conflates the gross size of an effect with the security (likely replicability) of the findings they represent.

But Fletcher et al. miss my point: In order to make the claim that phonics is more effective, Galuschka et al. should have tested for an interaction, with a greater effect for phonics compared to the other methods. This was not reported, and it would not have been significant (given the limited number of non-phonics studies included in the meta-analysis). Here is a fair characterization of what was found: Similar effect sizes were obtained across multiple different forms of instruction, with only phonics significant. However, there was no evidence that phonics was more effective than other methods.

**Suggate (2010, 2016)** Suggate (2010) reported evidence that the short-term impact of systematic phonics was most effective at the start of instruction, but I noted that the advantage of early phonics was small ( $d = \sim 0.1$ ), there was no indication that this advantage was significant, and the study that showed the largest benefit of early phonics ( $d = 1.37$ ) was carried out in Hebrew (a shallow orthography where GPCs are highly regular). Fletcher et al. responded: “Contextualizing this analysis again, it is important to look at the effects of phonics across languages to avoid pervasive Anglocentrism in our theorizing (Share 2008).” However, if the claim is that phonics instruction is the most effective in English, then it is appropriate to focus on English. Indeed, it is entirely possible that phonics is effective in languages with consistent GPCs but not English where GPCs are highly variable (due to the fact that English spellings are organized around both phonology and meaning). Suggate (2016) reported a meta-analysis that focused on the long-term impact of various teaching methods, and it showed that effects of systematic phonics were short-lived (reducing from  $d = 0.29$  immediately after training to  $d = 0.07$  following a mean delay of around 1 year) and indeed had the smallest effect size following a delay. Fletcher et al. responded that I only reported the “weighted” effect sizes, and that when unweighted measures are considered the long-term effects of phonics were greater (an overall effect of  $d = 0.25$  rather than  $d = 0.07$ ) and wrote: “The ‘weighted  $d$ ’ reported by Suggate (2016) and re-reported by Bowers adjusted for these large differences in sample size and thus reduces the observed effect sizes asymmetrically.” But it is not the case that larger sample sizes are selectively penalized; they are given more weight (as they should be). If studies with larger sample sizes produced larger effect sizes, then the weighted effect sizes would have been larger. In any case, the unweighted effect sizes for phonics were numerically smaller than all other forms of interventions. Fletcher et al. also criticized Suggate for excluding the Blachman et al. (2014) study as an outlier because it was a 10-year follow-up of their RCT intervention trial. However, Fletcher et al. failed to note that

the Suggate meta-analysis did include the 12-month follow-up of this study (where the effects were larger).

**Other Meta-analyses** I also briefly described another set of meta-analyses that focused on non-native speakers learning English (Adesope et al. 2011; Han 2010) and older children ranging from grades 5 to 12 (Sherman 2007). I did not consider these in any detail as these populations are not the focus of attention of most researchers claiming that there is strong evidence for systematic phonics. Fletcher et al. failed to note the main finding of the Sherman (2007) meta-analysis, namely, there was no overall significant effect of phonics (nor of phonological awareness). The two studies carried out on non-native speakers did report significant short-term effects but did not break down what outcome measures benefitted from phonics, and again did not distinguish between control conditions that included nonsystematic vs. no phonics. Another point I made regarding these latter two meta-analyses was that phonics was no more effective than alternative measures. For example, Adesope et al. found that collaborative reading and diary writing produced numerically larger effect sizes than systematic phonics. Given the strong emphasis on the importance of phonics, it does seem surprising that structured diary writing was as effective (is one to argue that diary writing is as critical as phonics based on these findings?). Fletcher et al. also note the fact that phonics was no more effective than other methods does not undermine the value of phonics, writing:

“We fully agree that there is no evidence to suggest “phonics-only” is optimal. We recognize that other practices such as “structured writing” might well be integrated within a reading approach to yield significant reading and writing outcomes.”

This is a fair point. This is why the natural experiment in England is so relevant as it assesses the impact of embedding systematic phonics in a broader literacy curriculum.

## The Natural Experiment in England

Proponents of systematic phonics claim that phonics is necessary but not sufficient, and indeed, a key point of Fletcher et al. is that the challenge for reading researchers is to combine phonics and various other types of instruction in an optimal manner. Still, the claim is that phonics does help on its own (as testified by the 1000s of citations to the above meta-analyses) and that phonics is a prerequisite for learning other skills. Even Fletcher et al., who at times reject the phonology first hypothesis, write:

Spellings can only be learned initially through phonological recoding because the child needs to link written word forms with spoken language (Seidenberg 2017). Shortly after initial instruction, explicit teaching of morphology can be part of a comprehensive reading program.

That is why the English experience is so relevant to this discussion. Since 2007, every state school in England is legally required to teach synthetic phonics, and since 2012, every child completes a Phonics Screening Check (PSC) in Year 1 that tests his or her ability to name aloud a set of regular words and nonwords in order to assess GPC knowledge. Of course, reading instruction in England includes more than phonics, and indeed, morphological instruction is introduced in later years. If indeed phonics instruction is necessary and most effective when combined in a broader reading curriculum, then there should be some

indication that reading outcomes have improved in England following over a decade of legally mandated phonics.

However, as I detail in my review, there is little or no evidence that the introduction of phonics has improved performance on any of the standardized reading outcomes, namely, the international standardized tests of PIRLS, PISA, nor the national reading tests in England called SATs. There is one standardized test that has improved, namely, performance on the PSC itself. The important point that needs to be emphasized is that contrary to the common claim of proponents of phonics, this has not yet translated into improved reading outcomes in general even when embedded in a wider reading curriculum.

It is also worth briefly noting that performance on the PSC has not improved as dramatically as commonly claimed. Many researchers have highlighted the finding that the pass rate on this test has increased from 58% pass rate in 2012 (correctly naming 80% of all items) to 82% in 2019. However, the mean improvement is somewhat less impressive. Because I could not find these results, I computed the means from the “phonics tables” from 2012 to 2019 from the GOV.UK website <https://www.gov.uk/government/collections/statistics-key-stage-1>. The mean scores improved from 29.5/40 to 33.8/40. Furthermore, this modest increase may simply reflect improvement due to training to the test. For example, there are reports that teachers in England are spending more time teaching children how to read aloud nonsense words that constitute half of the items on the test (Washburn, & Mulcahy, 2019). The PSC outcomes are not broken down by word and nonword (and the data are not available), but if this improvement was largely driven by improved performance in reading nonwords, it would help explain why the phonics instruction has been so ineffective in improving reading outcomes in general.

## Future Directions

Given that the evidence for systematic phonics is so weak, what should we do? One approach would be to continue to highlight the importance of phonics and figure out better ways to embed phonics in a richer reading program that combines the various aspects of reading instruction more effectively. Fletcher et al. advocate this position and describe in some detail an intervention of Morris et al. (2012) that they take to be a successful example of this. It is worth going into some detail into this study here as it provides a clear contrast with the alternative approach of explicitly teaching GPCs while rejecting phonics (Bowers & Bowers 2017, 2018a, b).

Morris et al. (2012) assessed three different reading interventions for struggling readers in grades 2–3. A phonological intervention taught children phonological awareness skills as well as phonics over 70 × 30-min lessons. The other two reading interventions included the same phonological intervention but included an additional 30 min of orthographic and semantic (including morphological) instruction to each lesson. The latter two conditions differed in how orthographic and semantic instruction was carried out and how they were combined with the phonological instruction. The authors found that the phonological intervention did not significantly improve any reading outcomes (including word reading accuracy, spelling, fluency, or comprehension) compared to a control condition that included no reading instruction, although the effect sizes were positive, ranging from  $g = 0.06$  to 0.32 (consistent with the meta-analyses reviewed above). By contrast, the two interventions that added orthographic and semantic instruction showed similar and significant benefits compared to the phonology condition across multiple outcome measures.

Although Fletcher et al. took this as evidence for the importance of including phonics in a broader instructional context, the study does not provide evidence for this conclusion. Not only did the phonology condition by itself fail to support significant effects, there is no evidence that it played a role in the significant results obtained in the two conditions that included orthographic and semantic instruction. In order to make this conclusion, it needs to be shown that the orthographic and semantic intervention conditions were less effective when the phonology intervention was absent. For example, the impact of combining the two different orthographic-semantic conditions (without phonology) could have been assessed and compared with the two interventions that included a combination of phonology, orthography, and semantic instruction. But this was not tested. In addition, it should be noted that the phonological intervention included instruction in phonological awareness and phonics, so it is not possible to conclude that the phonics was responsible for the small nonsignificant results. And given the control condition had no reading instruction, it is not even clear that the phonological condition would be more effective than 35 h of small group instruction with some version of nonsystematic phonics. In fact, Morris et al. (2012) did not describe their own study as providing evidence for the effectiveness of phonics—the word “phonics” only occurs once in their paper and not in reference to their intervention. I agree with Fletcher et al. that the two integrative interventions of Morris et al. are promising and should be followed up. But it is not yet clear what role phonics played in the successes.

In contrast with the Morris et al. approach, SWI rejects the “phonology first” hypothesis and explicitly teaches GPCs, morphology, spelling, and vocabulary together from the start (P. Bowers, 2021; Bowers and Bowers 2017, 2018a). There is preliminary evidence that SWI is effective in grade 1–2 students (Devonshire et al. 2013) that it is effective for a range of reading skills including vocabulary (Bowers and Kirby 2010) and spelling and naming (Devonshire and Fluck 2010). There is also preliminary evidence that the morphological matrices that are used in SWI to teach GPCs, spelling, and vocabulary together are an effective way to organize and learn information (Ng, Bowers, & Bowers 2020). But of course, the SWI approach needs more empirical support, and more work needs to be carried out in how to implement SWI at a larger scale given many of the concepts are unfamiliar to teachers (Colenbrander et al. 2021).

Yet another possibility is that some of the methods of SWI (e.g., morphological matrices, word sums, word investigations, cf. Bowers and Bowers, 2017) could be combined with phonics, with children first learning GPCs in isolation of the meaningful regularities of spellings, and only later teach children the interactions between GPCs and morphology and etymology. Currently, there is little research comparing different methods of teaching morphology and other orthographic regularities (Ng et al. 2020), and accordingly, more research on the SWI tools (and other tools, see Templeton & Bear 2017) is warranted when embedded within a phonics approach. Still, there are good arguments for rejecting phonics and teaching GPCs within a morphological context from the start, including the fact that learning in general is best when information is studied in a meaningful context (Bower et al. 1969), that spelling knowledge improves word naming (Ouellette et al. 2017), that morphological instruction improves phonological awareness and word decoding (Goodwin and Ahn 2013), and that morphological instruction is effective and sometimes most effective when introduced early (Bowers and Kirby 2010; Carlisle 2010; Galuschka et al. 2020; Goodwin and Ahn 2013).

In summary, there is little or no evidence that phonics by itself or in combination with other forms of instruction is more effective than whole language and other forms of instruction common in schools. This is not evidence in support of whole language; it is evidence that new

approaches are needed. There is no excuse for claiming that the science of reading strongly supports phonics without addressing the issues that Bowers (2020) has identified, or at least acknowledging that a systematic review of all the evidence has challenged this conclusion. Future research should consider not only new ways of combining phonics with other forms of instruction (as done by Morris et al. 2012) but also methods of explicitly teaching the phonological and semantic regularities of English spellings together, from the start.

**Acknowledgements** I would like to thank Patricia Bowers and Peter Bowers for their comments on earlier versions of this manuscript. For a blogpost associated with this article where you can comment see: <https://jeffbowers.blogs.bristol.ac.uk/blog/fletcher/>

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, M. J. (1990). *Beginning to read: thinking and learning about print*. Cambridge: MIT Press.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2011). Pedagogical strategies for teaching literacy to ESL immigrant students: a meta-analysis. *British Journal of Educational Psychology*, 81(Pt 4), 629–653.
- Anderson, L., Whiting, A., Bowers, P.N. & Venable, G. (2019). Learning to be literate: an orthographic journey with young students in R. Cox, S. Feez, L. Beveridge (Eds.). *The Alphabetic Principle and beyond... surveying the landscape*. Primary English Teaching Association Australia (PETAA).
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: are there effects a decade later? *Journal of Educational Psychology*, 106(1), 46–57.
- Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 8, 323–343. [https://doi.org/10.1016/S0022-5371\(69\)80124-6](https://doi.org/10.1016/S0022-5371(69)80124-6).
- 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. <https://doi.org/10.1007/s10648-019-09515-y>.
- Bowers, P.N. (2021). Structured Word Inquiry (SWI) Teaches Grapheme-Phoneme Correspondences More Explicitly Than Phonics Does: An open letter to Jennifer Buckingham and the reading research community. PsyArxiv. <https://psyarxiv.com/7qpyd/>
- 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. (2018a). Progress in reading instruction requires a better understanding of the English spelling system. *Current Directions in Psychological Science*, 27, 407–412. <https://doi.org/10.1177/0963721418773749>.
- Bowers, J. S. & Bowers, P. N. (2018b). There is no evidence to support the hypothesis that systematic phonics should precede morphological instruction: response to Rastle and colleagues. *PsyArXiv*. <https://psyarxiv.com/zg6wr/>.
- 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. *PsyArXiv* <https://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>.

- Brady, S. A. (2020). A 2020 Perspective on Research Findings on Alphabetics (Phoneme Awareness and Phonics): Implications for Instruction. *The Reading League Journal*, 1, 20–28.
- Buckingham, J. (2020). Systematic phonics instruction belongs in evidence-based reading programs: a response to Bowers. *The Educational and Developmental Psychologist*, 37, 105–113.
- Buckingham, J. (2020b). Evidence strongly favours systematic synthetic phonics. *Learning Difficulties Australia*, 52, 28–34.
- Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching children to read: the fragile link between science & federal education policy. *Education Policy Analysis Archives*, 11, 15.
- Camilli, G., Wolfe, M., & P., & Smith, M. L. . (2006). Meta-analysis and reading policy: perspectives on teaching children to read. *The Elementary School Journal*, 107, 27–36.
- Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, 45, 464–487. <https://doi.org/10.1598/RRQ.45.4.5>.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19, 5–51. <https://doi.org/10.1177/1529100618772271>.
- Colenbrander, D., Parsons, L., Murphy, S., Hon, Q., Bowers, J., & Davis, C. (2021). Morphological intervention for children with reading and spelling difficulties. (in press)
- Devonshire, V., & Fluck, M. (2010). Spelling development: fine-tuning strategy-use and capitalising on the connections between words. *Learning and Instruction*, 20, 361–371. <https://doi.org/10.1016/j.learninstruc.2009.02.025>.
- Devonshire, V., Morris, P., & Fluck, M. (2013). Spelling and reading development: the effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction*, 25, 85–94. <https://doi.org/10.1016/j.learninstruc.2012.11.007>.
- Ehri, L. C., & McCormick, S. (1998). Phases of word learning: implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly*, 14, 135–163. <https://doi.org/10.1080/1057356980140202>.
- 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, 393–447. <https://doi.org/10.3102/00346543071003393>.
- Fletcher, Savage, and Sharon (2020). A commentary on Bowers (2020) and the role of phonics instruction in reading. *Educational Psychology Review* (2020). <https://doi.org/10.1007/s10648-020-09580-8>
- Frith, U. (1985). Beneath the surface of dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: neuropsychological and cognitive studies of phonological reading* (pp. 301–330). Hillsdale, : Erlbaum.
- 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>.
- Galuschka, K., Görgen, R., Kalmar, J., Haberstroh, S., Schmalz, X., & Schulte-Körne, G. (2020). Effectiveness of spelling interventions for learners with dyslexia: a meta-analysis and systematic review. *Educational Psychologist*, 55(1), 1–20.
- Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, 17, 257–285. <https://doi.org/10.1080/10888438.2012.689791>.
- Hammill, D. D., & Swanson, H. L. (2006). The National Reading Panel's meta-analysis of phonics instruction: another point of view. *The Elementary School Journal*, 107(1), 17–26. <https://doi.org/10.1086/509524>.
- Han, I. (2010). 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>.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160.
- Larkin, R. F., & Snowling, M. J. (2008). Morphological spelling development. *Reading & Writing Quarterly*, 24, 363–376. <https://doi.org/10.1080/10573560802004449>.
- Levy, B., & Lysynchuk, L. (1997). Beginning word recognition: benefits of training by segmentation and whole word methods. *Scientific Studies of Reading*, 1, 359–387. [https://doi.org/10.1207/s1532799xssr0104\\_4](https://doi.org/10.1207/s1532799xssr0104_4).
- Levy, B., Bourassa, D., & Horn, C. (1999). Fast and slow namers: benefits of segmentation and whole word training. *Journal of Experimental Child Psychology*, 73, 115–138. <https://doi.org/10.1006/jecp.1999.2497>.
- McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., & et al. (2012). Phonics training for English speaking poor readers. *Cochrane Database of Systematic Reviews*, CD009115.
- McArthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H. C., Kohnen, S., & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD009115.pub3>.

- Morris, R. D., Lovett, M. W., Wolf, M. A., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities, 45*(2), 99–127.
- 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.
- Ng, M., Bowers, P. N., & Bowers, J. S. (2020). A promising new tool for literacy instruction: the morphological matrix. PsyArxiv, <https://doi.org/10.31234/osf.io/sgejh>
- Ouellette, G., Martin-Chang, S., & Rossi, M. (2017). Learning from our mistakes: improvements in spelling lead to gains in reading speed. *Scientific Studies of Reading, 21*(4), 350–357.
- Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex, 116*, 45–54.
- Rastle, K., & Taylor, J. S. H. (2018). Print-sound regularities are more important than print-meaning regularities in the initial stages of learning to read: response to Bowers & Bowers (2018). *Quarterly Journal of Experimental Psychology, 71*(7), 1501–1505.
- Seidenberg, M. (2017). *Language at the speed of sight: how we read, why so many Can't, and what can be done about it*. Basic Books.
- Shanahan, T. (2005). The national reading panel report: Practical advice for teachers. Learning Point Associates/North Central Regional Educational Laboratory (NCREL). Retrieved from: <https://files.eric.ed.gov/fulltext/ED489535.pdf>
- Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*(2), 151–218.
- Share, D. L. (2008). On the anglocentricities of current reading research and practice: the perils of overreliance on an “outlier” orthography. *Psychological Bulletin, 134*(4), 584–615.
- 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.
- Suggate, S. P. (2010). Why what we teach depends on when: grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*, 1556–1579.
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*, 77–96.
- Templeton, S., & Bear, D. (2017). *Word study research to practice: spelling phonics, meaning*. In *Handbook of research on teaching the English language arts* (pp. 206–231). London: Routledge.
- Torgerson, C. J., Brooks, G., & Hall, J. (2006). A systematic review of the research literature on the use of phonics in the teaching of reading and spelling (DfES research rep. 711). London: Department for Education and skills, University of Sheffield.
- Washburn, E. K., & Mulcahy, C. A. (2019). Morphology matters, but what do teacher candidates know about it? *Teacher Education and Special Education, 42*, 246–262.
- Wheldall, K. & Buckingham, J. (2020). Is systematic synthetic phonics effective? Nomanis Notes, 14, September. Download at: [https://57ebbb165-ef00-4738-9d6e-3933f283bdb1.filesusr.com/ugd/81f204\\_f8c0bc6c48f84da59fddb28e9b391282.pdf](https://57ebbb165-ef00-4738-9d6e-3933f283bdb1.filesusr.com/ugd/81f204_f8c0bc6c48f84da59fddb28e9b391282.pdf)