Reconsidering the evidence that systematic phonics is more effective than

alternative methods of reading instruction

Jeffrey S. Bowers

University of Bristol

Abstract

There is a widespread consensus in the research community that reading instruction in English should first focus on teaching letter (grapheme) to sound (phoneme) correspondences rather than adopt meaning-based reading approaches such as whole language instruction. That is, initial reading instruction should emphasize systematic phonics. In this systematic review I show this conclusion is not justified based on (a) an exhaustive review of 12 meta-analyses that have assessed the efficacy of systematic phonics, (b) summarizing the outcomes of teaching systematic phonics in all state schools in England since 2007. The failure to obtain evidence in support of systematic phonics should not be taken as an argument in support of whole language and related methods, but rather, it highlights the need to explore alternative approaches to reading instruction.

Reconsidering the Evidence that Systematic Phonics is More Effective than

Alternative Methods of Reading Instruction

There is a widespread consensus in the research community that early reading instruction in English should emphasize systematic phonics. That is, initial reading instruction should explicitly and systematically teach letter (grapheme) to sound (phoneme) correspondences. This contrasts with the main alternative method called whole language in which children are encouraged to focus on the meanings of words embedded in meaningful text, and where letter-sound correspondences are only taught incidentally when needed (Moats, 2000). Within the psychological research community, the "Reading Wars" (Pearson, 2004) that pitted whole language and phonics is largely settled – systematic phonics is claimed to be more effective. Indeed, it is widely claimed that systematic phonics is an essential part of initial reading instruction.

The evidence for this conclusion comes from various sources, including government panels that assessed the effectiveness of different approaches to reading instruction in the US (National Reading Panel, 2000), the UK (the Rose Review; Rose, 2006), and Australia (the Department of Education, Science and Training; Rowe, 2005), 12 meta-analyses of experimental research, as well as non-experimental studies that have tracked progress of students in England since the requirement to teach systematic phonics in state schools since 2007. The results are claimed to be clear-cut. For example, in his review for the English government, Sir Jim Rose writes:

> Having considered a wide range of evidence, the review has concluded that the case for systematic phonic work is overwhelming …" (Rose, 2006, p. 20).

Similarly, in a recent influential review of reading acquisition that calls for an end to the reading wars (in support of systematic phonics), Castles, Nation, and Rastle (2018) write:

> It will be clear from our review so far that there is strong scientific consensus on the effectiveness of systematic phonics instruction during the initial periods of reading instruction.

Countless quotes to this effect could be provided.

Importantly, this strong consensus has resulted in important policy changes in England and US. Based on Rose (2006), systematic phonics became a legal requirement in state-funded primary schools in England since 2007, and in to ensure compliance, all children (ages 5-6) complete a Phonics Screening Check (PSC) since 2012 that measures how well they can sound out a set of regular words and meaningless pseudowords. Similarly, based on the recommendations of the National Reading Panel (NRP, 2000), systematic phonics instruction was included in the Common Core State Standards Initiative in the US (http://www.corestandards.org/). The *Thomas Fordham Foundation* concluded that the NRP document is the third most influential policy work in US education history (Swanson & Barlage, 2006).

Nevertheless, despite this strong consensus, I will show that there is little or no evidence that systematic phonics is better than the main alternative method used in schools, including, whole language and balanced literacy. This should not be taken as an argument in support of these alternative methods, but rather, it should be taken as evidence that the current methods used in schools are far from ideal. Once this is understood, my hope is that researchers and politicians will be more motivated to consider alternative methods.

Structure of Paper

The remainder of the paper is organized in three main sections. First, I review the most common methods of reading instruction. There are some points of overlap between the alternative methods, but a commitment to systematic phonics entails some specific claims about what constitutes effective early reading instruction. Second, I explore the experimental evidence taken to support of systematic phonics by providing a detailed and exhaustive review of all meta-analyses that have assessed the efficacy of systematic phonics. The conclusion from this review is simple: There is little or no evidence that systematic phonics is better than the most common alternative methods used in schools. The problem is that: a) the findings are often mischaracterized by the authors of the reports and these mischaracterizations are passed on and exaggerated by many others citing the work, and b) that the designs of the meta-analyses often do not even test they hypothesis that systematic phonics is more effective than whole language and other common methods. Third, I review the outcomes a large naturalist experiment, namely, the impact of requiring systematic phonics in all English state schools since 2007. Again, the findings provide little or no evidence that systematic phonics has improved reading. Together, this should motivate researchers to consider alternative teaching methods.

What is Systematic Phonics and What are the Common Alternatives?

All forms of reading instruction are motivated by one or more of the following facts: (1) written words have pronunciations, (2) written words have a meaning, (3) words are composed of parts, including letters and morphemes (4) written words tend to occur in meaningful text, and (5) the ultimate goal of reading is to extract meaning from text. Different forms of instruction emphasize some of these points and down-play or ignore others, but there is nevertheless some overlap between different methods, and this complicates the task of comparing methods. For example, whole language instruction focuses on understanding words in the context of text, but it also includes some degree of phonics (e.g., Moats, 2000; NPR, 2000), and this has implications for how the meta-analyses described below can be interpreted. A further complication is that it is widely claimed that systematic phonics should be embedded in a broader literacy curriculum. For instance, the NRP (2000) emphasizes that systematic phonics should be integrated with other forms of instruction, including phonemic awareness, fluency, and comprehension strategies, and again, this makes it more difficult to make claims regarding systematic phonics per se. Because of these complexities, it is important to review systematic phonics and its relation to alternative methods in some detail so that the claims regarding the importance of systematic phonics can be evaluated.

As noted above, systematic phonics explicitly teaches children grapheme-phoneme correspondences prior to emphasizing the meanings of written words in text (as in whole language or balanced literacy instruction) or the meaning of written words in isolation (as in morphological instruction). That is, systematic phonics is committed to the "phonology first" hypothesis (Bowers & Bowers, 2018a). It is called systematic because it teaches grapheme-phoneme correspondences in an organized sequence as opposed to incidentally or on a 'when-needed' basis. Several versions of systematic phonics exist (most notably synthetic and analytic), but they all adopt the phonology first hypothesis.

The main alternative to phonics is *whole language* that primarily focuses on the meaning of words presented in text. Teachers are expected to provide a literacy rich environment for their students and to combine speaking, listening, reading, and writing. Students are taught to use critical thinking strategies and to use context to guess words that they do not recognize. Importantly, whole language typically includes some phonics, but the phonics instruction is not systematically taught (e.g., children are taught to sound out words when they

cannot guess the word from context). For example, the authors of the NRP (2000) report write:

> 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 (Routman, 1996). However, their approach is to teach it unsystematically and incidentally in context as the need arises. The whole language approach regards letter-sound correspondences, referred to as graphophonemics, as just one of three cueing systems (the others being semantic/meaning cues and syntactic/language cues) that are used to read and write text. Whole language teachers believe that phonics instruction should be integrated into meaningful reading, writing, listening, and speaking activities and taught incidentally when they perceive it is needed. As children attempt to use written language for communication, they will discover naturally that they need to know about letter-sound relationships and how letters function in reading and writing. When this need becomes evident, teachers are expected to respond by providing the instruction.

The fact that whole language (and related methods) includes non-systematic phonics turns out to be critical to the evaluations of the meta-analyses that follow.

Another approach to reading instruction called *balanced literacy* is designed to combine whole language with its focus on reading for meaning with systematic phonics. However, it is often claimed that balanced literacy is effectively just another name for whole language given that the phonics in balanced literacy is not taught first, not given enough emphasis, nor is it taught systematically (e.g., Moats, 2000).

Another teaching method is called *whole word or sight word training* in which children are taught to identify individual words (out of context) without breaking down the words into phonemes or other sub-lexical parts. For instance, in order to improve word naming, children might be given a list of written words and then one of the words is read aloud. The child's task is to select the corresponding written word, with the goal of improving their ability to read the word later (McArthur et al., 2013, 2015). Similarly, the *look-say-cover-write* method is commonly used in whole word instruction to teach children the spelling of words. In this method a child looks at a word, reads it aloud, covers the word up, and then attempts to spell the word (for review, see Browder, & Xin, 1998). Although whole word and whole language methods are different in many ways (most notably in whether some or no phonics is included), the two methods are often treated equivalently in the meta-analyses described below, and this has important implications for how the meta-analyses can be interpreted.

Morphological instruction, like whole language or balanced instruction, emphasizes the importance of attaching meaning to words, but it also teaches children to break down words into their meaningful parts (prefixes, bases, and suffixes). For review of this method see Carlisle (2010). Related to this, Structured Word Inquiry (SWI) teaches children the interrelation between all the sub-lexical components of written words (phonology, morphology, and etymology) in order to make sense of word spellings with the aim of improving all aspects of literacy, including reading, spelling, vocabulary, and comprehension (Bowers & Kirby, 2010). Like systematic phonics this approach explicitly teaches children the mappings between graphemes and phonemes, but children are taught how these mappings are organized within morphemes from the start (Bowers & Bowers, 2017, 2018a,b,c).

The overlap between methods and the claim that systematic phonics should be embedded with other methods makes the task of assessing the efficacy of systematic phonics per

se. more difficult. Nevertheless, proponents of systematic phonics are committed to two specific claims about what does and does not constitute good instruction, meaning that this approach can be evaluated.

First, it is claimed that systematic phonics should be taught before meaning-based approaches that focus on the meaning of written words in the context of sentences or focus on the meaningful sub-lexical structure of words (e.g., morphological instruction). For example, Castles, Rastle, and Nation (2018) write:

> …morphological instruction… may detract from vital time spent learning spelling-sound relationships. Instead, we would predict that the benefits of explicit morphological instruction are more likely to be observed somewhat later in reading development…

The claim that grapheme-phoneme correspondences should be taught prior to any morphological instruction is widespread (e.g., Adams, 1994; Ehri & McCormick, 1998; Henry, 1989; Larkin & Snowling, 2008; Taylor, Rastle, & Davis, 2017).

Second, it is claimed that grapheme-phoneme correspondences should be taught systematically (as the name suggests). That is, there should be a program of instruction in which all the relevant grapheme-phoneme mappings are taught explicitly in an ordered manner. This is not possible when teaching the grapheme-phoneme correspondences of words embedded in meaningful texts as typical with whole language (given that order of grapheme-phonemes in meaningful texts is too variable). The main justification for systematic phonics is empirical; namely, the widespread claim that studies support systematic phonics over alternative methods, as summarized in multiple meta-analyses detailed below.

To summarize, there are a number of different forms of reading instruction, some of which emphasize letter-sound mappings before other properties of words (e.g., systematic phonics), others that emphasize meaning from the start (e.g., whole language) and others that claim that the phonology and meaning of word spellings should be the focus of instruction from the beginning (structured word inquiry). There is no disagreement that reading instruction needs to ultimately incorporate both meaning and phonology, but the widespread consensus in the research community is that instruction needs to systematically teach children the grapheme-phoneme correspondences before meaning-based strategies are emphasized. Accordingly, almost all researchers today claim that systematic phonics is better than whole language, balanced literacy, and all forms of instruction that consider morphology from the beginning. The evidence for this claim is considered below and found wanting.

A Critical Examination of the Meta-Analyses Taken to Support Systematic Phonics

A total of twelve meta-analyses have assessed the efficacy of systematic phonics for individuals of different ages and abilities. In most cases (although not all) the meta-analysis are taken to support the conclusion that systematic phonics is an essential component of initial reading instruction and more effective than common alternatives such as whole language. As detailed below, this conclusion is not justified by any of the meta-analyses. The results have been mischaracterized by the authors themselves (summarizing the results in ways that mislead the reader), and in most cases, the design of the meta-analyses were not even designed to test the conclusions that were drawn by the authors.

National Reading Panel (2000) and Ehri et al. (2001) Meta-Analyses

The seminal report most often taken to support the efficacy of systematic phonics compared to alternative methods was a government document produced by the National Reading Panel (NRP, 2000), with the findings later published in peer review form (Ehri,

Nunes, Stahl, & Willows, 2001). The authors carried out the first meta-analysis evaluating the effects of systematic phonics compared to forms of instruction that include unsystematic or no-phonics across a range of reading measures, including word naming, nonword naming, and text comprehension tasks. The meta-analysis included 66 treatment-control comparisons taken from 38 experiments and the main findings can be seen in Table 1. Based on these findings Erhi et al. (2001) concluded in the abstract:

> "Systematic phonics instruction helped children learn to read better than all forms of control group instruction, including whole language. In sum, systematic phonics instruction proved effective and should be implemented as part of literacy programs to teach beginning reading as well as to prevent and remediate reading difficulties"

The NRP report has been cited over 24,000 times and continues to be used in support of systematic phonics, with over 1000 citations in 2019). In addition, the Ehri et al. (2001) article has been cited over 1000 times. However, a careful look at the results highlights undermines these strong conclusions.

The most important limitation is that systematic phonics did not help children labelled "low achieving" poor readers (d = .15, not significant). These were children above first grade who were below average readers and whose cognitive level was below average or not assessed. By contrast, children labelled "reading disabled" who were below grade level in reading but at least average cognitively and were above first grade in most cases did benefit (d = .32). Note, by definition, half the population of children above grade 1 will have an IQ below average, and it is likely that more than 50% of struggling readers above grade 1 will fall into this category given the comorbidity of developmental disorders (Gooch, Hulme, Nash, & Snowling, 2014). Of course, additional research may show that systematic phonics does benefit low achieving poor readers above grade 1 (the NRP only included 8 comparison groups in this condition), but there is no evidence for this from the NRP meta-analysis.

Second, based on the finding that effect sizes were greater when phonics instruction began by first grade (d = 0.55) rather than after first grade (d = 0.27), the authors of the NRP wrote in the executive summary "Phonics instruction taught early proved much more effective than phonics instruction introduced after first grade" (p 2-93). But in the body of the text it becomes clear that findings do not support this strong conclusion. One problem is that the majority of older students (78%) in the various studies included in the NRP analysis were either low achieving readers or students with reading disability, and as noted above, systematic phonics was less effective with both these populations (especially the former group). With regards to the normally developing older readers, the NRP meta-analysis only included seven comparison groups, and four of them used the Orton-Gillingham method that was developed for younger students. As noted by Ehri et al. (2001):

> "The conclusion that phonics instruction is less effective when introduced beyond first grade may be premature… Other types of phonics programs might prove more effective for older readers without any reading problems. (p. 428).

This is straightforwardly at odds with the above executive summary and explains why so many authors cite the NRP as providing evidence that early phonics instruction is important.

Third, although the authors of the NRP emphasized the systematic phonics had long-term impact, the effect size declined from d = .41 when children were tested immediately following the intervention to d = .27 following a 4 to12 month delay. However, the authors did not assess whether the long-term benefits extended to spelling, reading texts, or reading com-

prehension. Given that the short-term effects on spelling, reading texts, or reading comprehension was much reduced compared to the overall short-term effect (Table 1), there is no reason to assume these effects persisted.

Fourth, the evidence that that systematic phonics is more effective than whole language is weaker still. This claim is not based on the overall effect size of d = .41, but rather, on a sub-analysis that specifically compared systematic phonics to whole language. This analysis was based on 12 rather than 38 studies, and not one of these 12 studies used a randomized control trial (RCT) design. This analysis showed a reduced overall effect of d = .31 (still significant), with the largest effect obtained for decoding (mean of the reported effect sizes was d = .55) and smallest effect on comprehension (mean of the reported effect sizes was d = .19), with only two studies assessing performance following a delay. And although the NRP is often taken to support the efficacy of <u>synthetic</u> systematic phonics (the version of phonics legally mandated in the UK), the NRP meta-analysis only included four studies relevant for this comparison (of 12 studies that compared systematic phonics to whole language, only four assessed synthetic phonics). The effect sizes in order of magnitude were: d = .91 and d = .12 in two studies that assessed grade 1 and 2 students, respectively (Foorman, et al. 1998); d = .07 in a study that asses grade 1 students (Traweek & Berninger, 1997), and d = -.47 in a study carried out grade 2 students (Wilson & Norman, 1998).

In sum, rather than the strong conclusions emphasized the executive summary of the NRP (2000) and the abstract of Ehri et al. (2001), the appropriate conclusion from this meta-analysis should be something like this:

Systematic phonics provides a small short-term benefit to spelling, reading text, and comprehension, with no evidence that these effects persist following a delay of 4-12 months (the effects were not reported nor assessed). It is unclear whether there is an advantage of introducing phonics early, and there are no short- or long-term benefit for majority of struggling readers above grade 1 (children with below average intelligence). Systematic phonics did provide a moderate short-term benefit to regular word and pseudoword naming, with overall benefits significant but reduced by a third following 4-12 months.

And even these weak conclusions in support of systematic phonics are not justified given subsequent work by Camilli et al. (2003, 2006) and Torgerson et al. (2006) who re-analyzed the studies (or a subset of studies) included in the NRP, as described next.

Camilli et al. (2003, 2006)

Camilli et al. (2003) identified a number of flaws in the NRP meta-analysis, but here I emphasize one, namely, it was not designed to assess whether there is any benefit in teaching phonics systematically. Similar design choices were made by all subsequent meta-analyses taken to support systematic phonics, and this has led to unwarranted conclusions from these meta-analyses as I detail below.

As noted above, the headline figure from the NRP analysis is that systematic phonics showed an overall immediate effect size of d = .41. What needs to be emphasized is that this figure is the product of comparing systematic phonics to a heterogeneous control condition that included (1) intervention studies that used unsystematic phonics and (2) intervention studies that used no phonics. As elementary point of logic, if you compare systematic phonics to a mixture of different methods, some of which use unsystematic phonics and other that use no phonics, then it is not possible to conclude that systematic phonics is more effective than unsystematic phonics. In order to assess whether the "systematic" in systematic phonics

is important it is necessary to compare systematic phonics to studies that included unsystematic phonics, something that the NRP (2000) did not do.

The reason why this is important is that unsystematic phonics is standard in common alternatives to systematic phonics. Indeed, in addition to the widespread use of unsystematic phonics in the US prior to the NPR (2000) report (as shown above in a quote from NRP), Her Majesty's Inspectorate (1990) reported that unsystematic phonics was also common in the UK prior to the legal requirement to teach systematic synthetic phonics in England in 2007, writing:

> ...phonic skills were taught almost universally and usually to beneficial effect'' (p. 2) and that ''Successful teachers of reading and the majority of schools used a mix of methods each reinforcing the other as the children's reading developed (p. 15).

Accordingly, the important question is whether systematic phonics is more effective than the unsystematic phonics that is used in alternative teaching methods.

In order to assess the importance of teaching phonics systematically, Camilli et al. (2003, 2006) coded the studies included in the NRP as having no phonics, unsystematic phonics, or systematic phonics. In addition, the authors also noted that some moderator variables were ignored by the NRP analysis that may have contributed to the outcomes. Accordingly, the authors also coded whether or not the intervention studies included language-based reading activities such as shared writing, shared reading, or guided reading, whether treatments were carried out in the regular class or involved tutoring outside the class, and whether basal readers were used (if known). Both the experimental and control groups were coded with regards to these moderator variables. It should also be noted that the Camilli et al. (2003, 2006) analyses were carried out on a slightly modified dataset given problems with some of the studies and conditions included in the NRP report. For example, the authors dropped one study (Vickery et al., 1987) that did not include a control condition (an exclusion condition according to the NRP) and included three studies that were incorrectly excluded (the studies did fulfill the NRP inclusion criterion), resulting in a total of 40 rather than 38 studies. The interested reader can find out more details regarding the slightly modified dataset in Camilli et al. (2003), but in any case, the different datasets produce the same outcome as discussed below.

The Camilli et al. (2003) analysis showed that effect size of systematic phonics compared to non-systematic phonics was significant, but roughly half the size of the effect of systematic phonics reported in the NRP report (d = .24 vs. d = .41). Interesting, the analysis also found significant and numerically larger effects of systematic language activities (d = .29) and tutoring (d = .40). The subsequent analysis by Camilli et al., (2006) was carried out on the same dataset that but used a new method of analysis (a multilevel modelling approach) and included three rather than two levels of language-based reading activities as a moderator variable (none vs., some, vs. high levels of language-based activities). This analysis revealed an even smaller effect of systematic phonics (d = .12) that was no longer significant. Camilli et al. (2006) took these findings to challenge the strong conclusion drawn by the authors of the NRP.

These analyses were subsequently supported by Stuebing et al. (2008) who re-analyzed the Camilli et al. (2003, 2006) dataset and showed that the different outcomes were not the consequence of the slightly different studies included in the Camilli and the NPR meta-analyses. However, Stuebing et al. (2008) drew a different conclusion, writing:

The NRP question is analogous to asking about the value of receiving the intervention versus not receiving the intervention. The Camilli et al. (2003) report is analogous to asking what is the value of receiving a strong form of the intervention compared to a receiving weaker forms of the intervention and relative to factors that moderate the outcomes. From our view, both questions are reasonable for intervention studies.

But the two questions are not equally relevant to teaching policy. The relevant question is whether systematic phonics is better than pre-existing practices. Given that unsystematic phonics was standard practice, and given the Camilli et al. (2006) analysis failed to show an advantage of systematic over unsystematic phonics, this analysis challenges the main conclusion that schools should introduce systematic phonics.

To avoid any conclusion, it is important to highlight that the Camilli et al. (2006) reanalysis of the NRP dataset does <u>not</u> suggest that grapheme-phoneme knowledge is unimportant. Indeed, their reanalysis suggests that systematic phonics is significantly better than a non-phonics control condition. Rather, their key finding is that systematic phonics was no better than non-systematic phonics as commonly used in schools.

Torgerson et al. (2006)

The Torgerson et al. (2006) meta-analysis was primarily motivated by another key limitation of the NRP report not touched on thus far, namely, the fact that the NRP included studies that employed both randomized and non-randomized designs. Given the methodological problems with non-randomized studies, Torgerson et al. (2006) carried out a new meta-analysis that was limited to randomized control trials (RCTs). But it is worth noting two additional limitations of the NRP report that motivated this analysis.

First, the authors were concerned that bias played a role in 13 RCT studies included in the original NRP report given that the NRP report only considered published studies (studies that obtained null effects may have been more difficult to publish). Indeed, the authors carried out a funnel plot analysis on these 13 studies and concluded that the results provided: "…prima facie evidence for publication bias, since it seems highly unlikely that no RCT has ever returned a null or negative result in this field". Accordingly, Torgerson et al. (2006) searched for unpublished studies that met their inclusion criteria. They found one additional study that reported an effect size of -0.17 that they included in their analyses. Note, this bias would have inflated the small effects reported in the NRP (2000) and the Camilli (2003, 2006) meta-analyses. Second, Torgerson et al. removed two studies that should have been excluded from the NRP analyses (Gittelman & Feingold, 1983, because it did not include a phonics instruction intervention group; Mantzicopolous et al., 1992, because the children in the control condition did not receive a reading intervention, and the attrition rate of the studies was extreme, with 437 children randomized and only 168 children tested).

In total, the authors identified 12 studies that compared systematic phonics to a control condition that included unsystematic phonics or no phonics instruction control (same number as in the original NRP report, with one study dropped, one unpublished study added). The key positive result was with regards to word reading accuracy with an effect size estimated to between .27 and .38 (depending on assumptions built into the analyses). By contrast, no significant effects were obtained for comprehension (d estimates ranging between .24 and .35), or spelling (d = .09).

There are, however, reasons to question the significant word reading accuracy results. This result was largely due to one outlier study (Umbach et al., 1989) that obtained a massive

effect on word reading accuracy (d = 2.69)[1]. In this study, the control group was taught by two regular teachers with help from two university supervised practicum students, whereas the experimental group was taught by four Masters' degree students who were participating in a practicum at a nearby university. Accordingly, there is a clear confound in the design of the study. Torgerson et al. themselves re-analyzed the results when this study was excluded and found that the word reading accuracy result was reduced (d estimates between .20 and .21) with the effect just reaching significance one analysis (p = .03) and non-significant on another (p = .09). For summary of findings see Table 1. And even these findings likely over-estimate the efficacy of systematic phonics given the evidence that bias may have inflated the estimate of effect sizes in this study. As Torgerson et al. wrote:

> In addition, the strong possibility of publication bias affecting the results cannot be excluded. This is based on results of the funnel plot... It seems clear that a cautious approach is justified (p. 48).

The conclusions one can draw are further weakened by the quality of the studies included in the meta-analysis, with the authors writing:

> …none of the 14 trials reported method of random allocation or sample size justification, and only two reported blinded assessment of outcome… all were lacking in their reporting of some issues that are important for methodological rigour. Quality of reporting is a good but not perfect indicator of design quality. Therefore due to the limitations in the quality of reporting the overall quality of the trials was judged to be 'variable' but limited.

Nevertheless, despite all the above issues, the authors concluded:

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

This quote not only greatly exaggerates the strength of the findings (which helps explain why the meta-analysis has been cited over 250 times in support of systematic phonics), but it again reveals a misunderstanding regarding the conclusions one can draw from the design of the meta-analysis. The study continued to use the design of the NRP (2000) meta-analysis that compared systematic phonics to a control condition that combined (1) non-systematic phonics and (2) no phonics. Accordingly, it is not possible to conclude that systematic phonics is more effective than whole word instruction that uses unsystematic phonics. That would require a direct comparison between conditions that was not carried out.

To summarize thus far, a careful review of the NPR (2000) findings show that that the benefits of systematic phonics for reading text, spelling, and comprehension are weak and short-lived, with reduced or no benefits for low achieving poor readers beyond grade 1. The subsequent Camilli et al. (2003, 2006) and Torgerson et al. (2006) reanalyses further weakens these conclusions. Indeed, Camilli et al. (2006) found no overall benefit of systematic phonics over non-systematic phonics, and Torgerson et al. (2006) did not find any benefit of systematic phonics in the

---

[1] For some reason the NRP report estimated the word reading accuracy effect size to be 1.3 whereas Torgerson et al. reported it to be 2.69. Umbach and Darch did not report standardized effect sizes themselves, but reported that the word identification scores from the Woodcock Reading Master subtest were 30.43 and 10.36 in the experimental and control conditions, respectively.

subset of RCT studies included in the NRP for word reading accuracy, comprehension or spelling (when one outlier study was excluded). The null effects in the Torgerson et al. (2006) meta-analysis were obtained despite evidence for publication bias and flawed design that combined unsystematic and no phonics studies into a control condition (with both of these factors serving to inflate the benefits of systematic phonics). A summary of all results can be found in Table 1.

McArthur et al. (2012)

This meta-analysis was designed to assess the efficacy of systematic phonics with children, adolescents, and adults with reading difficulties. The authors included studies that use randomization, quasi-randomization, or minimization (that minimizes differences between groups for one or more factors) to assign participants to either a systematic phonics intervention group, or a control group that received no training or alternative training that did not involve any reading activity (e.g., math training). That is, the control group received no phonics at all. Based on these criteria the authors identified 11 studies that assessed a range of reading outcomes, although some outcome measures were only assessed in a few studies. Critically, the authors found a significant effect of word reading accuracy (d = .47, p = 0.03) and nonword reading accuracy (d = 0.76, p < 0.01), whereas no significant effects were obtained in word reading fluency (d = -0.51; expected direction), reading comprehension (d = 0.14), spelling (d = 0.36), and nonword reading fluency (d = 0.38, the unexpected direction). Based on the results the authors concluded that systematic phonics improved performance, but they were also cautious in their conclusion, writing:

> …there is a widely held belief that phonics training is the best way to treat poor reading. Given this belief, we were surprised to find that of 6632 records, we found only 11 studies that examined the effect of a relatively pure phonics training programme in poor readers. While the outcomes of these studies generally support the belief in phonics, many more randomised controlled trials (RCTs) are needed before we can be confident about the strength and extent of the effects of phonics training per se in English-speaking poor word readers.

But there are reasons to question even these modest conclusions. One notable feature of the word reading accuracy results is that they were largely driven by two studies (Levy et al., 1997; Levy et al. 1999) with effect sizes of d = 1.12 and d= 1.80, respectively. The remaining 8 studies that assessed reading word accuracy reported a mean effect size of .16 (see appendix 1.1, page 63). This is problematic given that the children in the Levy studies were trained on one set of words and then reading accuracy was assessed on another set of words that shared either onsets or rhymes with the trained items (e.g., a child might have been trained on the word *beak* and later be tested on the word *peak*; the stimuli were not presented in either paper). Accordingly, the large benefits observed in the phonics conditions compared to a non-trained control group only shows that training generalized to highly similar words rather than word reading accuracy more generally (the claim of the meta-analysis). In addition, both Levy et al. studies taught systematic phonics using one-on-one tutoring. Although McArthur et al. reported that group size did not have an overall impact on performance, one-on-one training studies with a tutor showed an average effect size of d= .93 (over three studies). Accordingly, the large effects size for word reading accuracy may be more the product of one-on-one training with a tutor rather than any benefits of phonics per se, consistent with the findings of Camilli et al. (2003). Absent the two Levy et al. studies, there is no evidence from the McArthur et al. (2012) meta-analysis that systematic phonics condition improved

word reading accuracy, word reading fluency, reading comprehension, spelling, or nonword reading fluency, leaving only a benefit for nonword reading accuracy.

But even putting these concerns aside, the most important point to note is that meta-compared systematic phonics to no extra training at all, or to training on non-reading tasks. Accordingly, it is not appropriate to attribute any benefits to systematic phonics. Any form of extra instruction may have mediated the (extremely limited) gains. So once again, this analysis should not be used to make any claims that systematic phonics is better than standard alternative methods, such as whole language that do include unsystematic phonics.

Galuschka et al. (2014)

Galuschka et al. carried out a meta-analysis of randomized controlled studies that focused on children and adolescents with reading difficulties. The authors identified twenty-two trials with a total of 49 comparisons of experimental and control groups that tested a wide range of interventions, including five trials evaluating reading fluency trainings, three phonemic awareness instructions, three reading comprehension trainings, 29 phonics instructions, three auditory trainings, two medical treatments, and four interventions with colored overlays or lenses. Outcomes were divided into reading and spelling measures.

The authors noted that only phonics produced a significant effect, with an overall effect size of g' =. 32, and concluded:

> This finding is consistent with those reported in previous meta-analyses...
> At the current state of knowledge, it is adequate to conclude that the system-
> atic instruction of letter-sound-correspondences and decoding strategies, and
> the application of these skills in reading and writing activities, is the most
> effective method for improving literacy skills of children and adolescents
> with reading disabilities

However, there are serious problems with this conclusion. Most notably, the overall effect sizes observed for phonics (g' =.32) was similar to the outcomes with phonemic awareness instruction (g' = .28), reading fluency training (g' = .30), auditory training (g' = .39), and colour overlays (g'= .32), with only reading comprehension training (g' = .18) and medical treatment (g' = .12) producing numerically reduced effects. The only reason significant results were only obtained for phonics is that there were many more phonics interventions. In order to support their conclusion, the authors need to show an interaction between the phonics results and the alternative methods. They did not report this analysis, and given the similar effect sizes across conditions (with small sample sizes), this analysis would not be significant. Of course, future research might support the authors conclusion, but this meta-analysis does not support it.

To further compromise the authors' conclusion, Galuschka et al. reported evidence that the published phonics studies were biased using a funnel plot analysis. Using a method called Duval and Tweedie's trim and fill they measured the extent of publication bias and estimated an unbiased effect size for systematic phonics to be greatly reduced, although still significant, g'= 0.198. And yet again, the design of the meta-analysis did not assess whether systematic phonics was more effective than unsystematic phonics (let alone show that systematic phonics is more effective than the alternative methods they did investigate). Nevertheless, the meta-analysis is frequently cited as evidence in support of systematic phonics over whole language (e.g., Lim & Oei, 2015; Treiman, 2018; van der Kleij, Segers, Groen, & Verhoeven, 2017).

Suggate (2010, 2016)

Suggate (2010) carried out a meta-analysis to investigate the relative advantages of systematic phonics, phonological awareness, and comprehension-based interventions with children at-risk of reading problems. The central question was whether different forms of interventions were more effective with different age groups of children who varied from preschool to Grade 7.

The meta-analysis included peer-reviewed randomized and quasi-experimental studies, with control groups receiving either typical instruction or an alternative "in-house" school reading intervention. They identified 85 studies with 116 interventions: 13 were classified as phonological awareness, 36 as phonics, 37 as comprehension based, and 30 as mixed. Twelve studies were conducted with participants who did not speak English. A range of dependent measures were assessed, from pre-reading (e.g., letter knowledge, phonemic/sound awareness), reading, and comprehension measures.

Averaging over age, similar overall effects were for phonological awareness (d = .47), phonics (d = .50), meaning based (d = .58), and mixed (d = .43). The critical novel finding, however, was that there was a significant interaction between method of instruction and age of child, such that phonics was most useful in Kindergarten for reading measures, but alternative interventions were more effective for older children. As Suggate (2010) writes:

> If reading skills per se are targeted, then there is a clear advantage for phonics interventions early and—taking into account sample sizes and available data— comprehension or mixed interventions later.

However, this is not a safe conclusion. First, the difference in effect size in phonics compared to alternative methods was approximately d = .10 in Kindergarten and .05 in Grade 1 (as estimated from Figure 1 in Suggate, 2010). This is not a strong basis for arguing the importance of early systematic phonics. It is also important to note that 10% of the studies included in the meta-analysis were carried out on non-English children. Although the overall difference between non-English (d = 0.61) and English (d = .48) studies was reported as non-significant, the difference approached significance (p = .06). Indeed, the phonics intervention that reported the very largest effect size (d = 1.37) was carried out in Hebrew speakers (Aram & Biron, 2004), and this study contributed to the estimate of the phonics effect size in pre-kindergarten. Accordingly, the small advantage of phonics for early English instruction (the main novel finding in this report) is inflated. And once again, the treatments were compared to a control condition that combined a range of teaching conditions, and accordingly, it is again unclear whether there was a difference between systematic vs. unsystematic phonics during early instruction.

But the most critical limitation is that Suggate's (2010) conclusion regarding the benefits of early phonics instruction is contradicted in a subsequent Suggate (2016) meta-analysis. This meta-analysis included 71 experimental and quasi-experimental reading interventions that assessed the short- and long-term impacts of phonemic awareness, phonics, fluency, and comprehension interventions on pre-reading, reading, reading comprehension, and spelling measures. The analysis revealed an overall short-term effect ($d = 0.37$) that decreased in a follow-up test ($d = 0.22$; with mean delay of 11.17 months) with phonics producing the most short-lived benefits. Specifically, the long-term effects were: phonics, d = .07; fluency, d = .28; comprehension d = 46; and phonemic awareness, d = .36.

As with the other meta-analyses there are additional issues that should be raised. For example, a funnel plot observed evidence for publication bias, especially in the long-term condition, and once again, the study does not compare systematic to unsystematic phonics. It

is striking that long-term benefits of systematic phonics are so small despite these factors that should be expected to inflate effect sizes.

Other Meta-Analyses and a Systematic Review of Meta-Analyses

There are a number of additional relevant meta-analyses and reviews of meta-analyses that should be mentioned briefly as well.

Hammill and Swanson (2006).   These authors took a different approach to Camilli et al. (2003, 2006, 2008) in criticizing the NRP (2000) report. Rather than challenging the logic and analyses themselves, they noted that the effects sizes reported in the NRP were small and questioned their significance.

The NRP reported that systematic phonics instruction was effective across a variety of conditions, with 94% of the *d*'s supporting the superiority of phonics instruction over other approaches. However, as noted by Hammill and Swanson, the standard convention in evaluating the magnitude of d' sizes (d = 0.2 is small, d = .5 medium, and .9 large) reveals that 65% of the significant *d*'s were small. In order to get a better intuitive understanding of the practical significance of the results, the authors converted all these d's values to r-type statistics. They noted that the overall effect of .44 corresponds to an r^2 value of .04. That is, 96% of the variance in reading achievement can be attributed to factors other than the systematic phonics instruction.[2] The r^2 value for the follow-up analysis (4-12 months later) was .02.

What Hammill and Swanson do not acknowledge, however, is that these small effect sizes translate into real benefits when considering an entire population of children. The real problem is not with the size of the effects, it is that many of the critical contrasts were not significant or not assessed, that the small effects that were significant were inflated for the reasons noted above, and perhaps most importantly, the main meta-analysis did not even test the critical hypothesis of whether systematic phonics is better than unsystematic phonics that is used in alternative methods such as whole language.

Han (2010) and Adesope, Lavin, Tompson, and Ungerleider (2011). These authors reported meta-analyses that assessed the efficacy of phonics for non-native English speakers learning English. Hans (2010) included 5 different intervention conditions and dependent measures and reported the overall effect sizes as 0.33 for phonics, 0.41 for phonemic awareness, 0.38 for fluency, 0.34 for vocabulary, and 0.32 for comprehension. In the case of Adesope et al. (2011), the authors found that systematic phonics instruction improved performance (g = +0.40), but they also found that an intervention they called Collaborative Reading produced a larger effect (g = +.48) as did a condition called Writing (Structured & Diary) that produced an effect of g = +.54. Accordingly, ignoring all other potential issues discussed above, these studies do not provide any evidence that phonics is the most effective strategy for reading acquisition.

Sherman (2007). Sherman compared phonemic awareness and phonics instruction with students in grades 5 through 12 who read significantly below grade level expectations. Neither method was found to provide a significant benefit.

Torgerson et al. (2018). Finally, Torgerson et al. carried out a systematic review of all meta-analyses that assessed the efficacy of systematic phonics (unlike the papers discussed above this is not a meta-analysis itself). They identified 12 meta-analyses, all of which were considered above. The authors raised several concerns regarding design and

---

[2] Note, the overall effect size reported in the NRP are estimated as .44 or .41, depending on the specific tests that are included in the analyses.

publication bias of studies included in these meta-analyses and argued that more data (in the form of large randomized controlled studies) are needed before strong conclusions can be made. Nevertheless, the authors still conclude the evidence support systematic phonics, writing:

> Given the evidence from this tertiary review, what are the implications for teaching, policy and research? It would seem sensible for teaching to include systematic phonics instruction for younger readers – but the evidence is not clear enough to decide which phonics approach is best.

Despite their modest conclusions the authors are still far too positive regarding the benefits of systematic phonics. In part, this is due to the way the authors summarize the findings they do report. But more importantly it is the consequence of ignoring many of key limitations of the meta-analyses discussed above.

With regards to their own summary of the meta-analyses, they stated that 10 the 12 meta-analyses showed that there were significant benefits of systematic phonics on at least one reading measure, with effect sizes ranging from small to moderate effects (Ehri et al. 2001; Camilli, Vargas, and Yurecko 2003; Torgerson, Brooks, and Hall 2006; Sherman 2007; Han 2010; Suggate 2010; Adesope et al. 2011; McArthur et al. 2012; Galuschka et al. 2014; Suggate 2016). Furthermore, they note that positive effects were found in the remaining non-significant meta-analyses (Camilli, Wolfe, and Smith 2006; Hammill and Swanson 2006). They take this to support the conclusion teaching should include systematic phonics.

One problem with this description of the results is that it does not indicate which measures tended to be significant over the meta-analyses. In fact, as discussed above, most meta-analyses failed to obtain significant effects for the measure we should care about most. For example, only 1 of 12 studies reported significant effects in comprehension, and there is no evidence that this effect survived a delay (NRP, 2000). And this characterization of the findings obscures the fact that the benefits did not always extend to the children who are below average in their cognitive capacities (NRP, 2000).

This summary also does not highlight the fact that many of 12 meta-analyses observed larger effect sizes for non-phonics interventions. For example, from Table 3 of Torgerson et al. (2018), you find out that systematic phonics did not produce the largest effect in 5 of the 12 meta-analyses (Adesope et al. 2011; Camilli, Vargas, and Yurecko 2003; Camilli, Wolfe, and Smith 2006; Han 2010; Suggate 2016). And this table does not include the Galuschka et al. (2014) meta-analysis that reported similar sized effect sizes for phonics, phonemic awareness instruction, reading fluency training, and auditory training, with the largest numerical effect obtained with colour overlays.

In addition, when claiming that 10 of the 12 meta-analyses reported significant benefits of systematic phonics, this included the Suggate (2010) meta-analysis that was challenged by a subsequent Suggate (2016) meta-analysis that failed to obtain long-term benefits of systematic phonics. Furthermore, the claim that 10 of the 12 meta-analyses reported a significant benefit for systematic phonics does not incorporate a key point highlighted by Torgerson et al. (2018) elsewhere in their review, namely, the evidence that publication and method bias has inflated these effect sizes in at least some of these meta-analyses.

The conclusion that systematic phonics is better than alternative methods is further compromised by additional factors not considered by Torgerson et al. (2018). As detailed above, there were multiple examples of methodological errors in the meta-analyses (e.g., excluding studies that should have been included given the inclusion criteria; Camilli et al.,

2003; and including studies that should have been excluded given the exclusion criteria, Camilli et al., 2003; Torgerson et al., 2006), examples of including flawed studies that strongly biased the findings in support of systematic phonics (e.g., the Umbach et al., 1989; Levy, 1997, 1999), including non-English studies that biased the results in support of systematic phonics (Suggate, 2010), amongst others. These errors consistently biased the estimates of systematic phonics upwards.

Most importantly, however, Torgerson et al., 2006 did not address the key point identified by Camili et al. (2003, 2006, 2008) that compromises all meta-analyses used in support of systematic phonics, namely, systematic phonics was compared to a control condition that included both non-systematic phonics and non-phonics conditions (or only included a non-phonics condition in the case of McArthur et al., 2012). Accordingly, these meta-analyses did not even test the hypothesis that systematic phonics is more effective than unsystematic phonics as used in whole language and other methods. For all these reasons, Torgerson et al.'s are unwarranted in their conclusion that systematic phonics is effective for young children.

Summary of Meta-Analyses

In sum, the above research provides little or no evidence that systematic phonics is better than standard alternatives methods used in schools. The findings do not challenge the importance of learning grapheme-phoneme correspondences, but they do undermine the claim that systematic phonics is more effective than alternative methods that include unsystematic phonics (such as whole language) or that teach grapheme-phoneme correspondences along with meaning-based constraints on spellings (morphological instruction or structured word inquiry). There can be few areas in psychology in which the research community so consistently reaches a conclusion that is so at odds with available evidence.

The Systematic Phonics Experiment in England

One possible response to the many null results is to note that many of the studies included in these meta-analyses were flawed. On this view, the null results can be attributed to a limitation of the studies rather than any problems with systematic phonics per se. This is hard to reconcile with the fact that these meta-analyses have been cited thousands of times in support of systematic phonics. Nevertheless, the quality of the studies included in the meta-analyses have been repeatedly question (e.g., McArthur et al., 2012; Torgerson et al., 2006, 2018), and accordingly, it is possible that systematic phonics is effective, but the meta-analyses are simply not picking his up. Another possible response is to note that systematic phonics needs to be taught in combination with many other skills, and the fact that phonics by itself does not improve reading outcomes is not surprising. Again, this is hard to reconcile with the claims that are drawn from the meta-analyses, but these concerns raise the question as to whether there are other sources of data that can be used to assess the benefits of systematic phonics when embedded in a broader literacy environment?

There is. In 2006 Sir Jim Rose wrote a UK government report concerned with the teaching of reading in primary schools in England where he concluded that "... the case for systematic phonic work is overwhelming …" (Rose, 2006, p. 20). Although this conclusion is unwarranted (see above), the report led to the legal requirement to teach synthetic systematic phonics in English state schools since 2007. And since 2012, a phonics screen check (PSC) was introduced in order to encourage better teaching of systematic phonics and to assess how well children decode regular words and pseudowords. Over 650,000 children took the PSC in 2018 alone. This constitutes a massive naturalistic experiment that can be used to assess the efficacy of systematic phonics, and indeed, it is widely claimed that the experiment

has been a success which systematic phonics improving literacy. But once again, a careful look into the findings shows that the data do not support this conclusion. I summarize the findings next.

Machin, McNally, and Viarengo's (2018) Analysis of Standard Assessment Tests (SATs) Results in England Provides Little or No Evidence in Support of Systematic Phonics

The authors took advantage of the fact that systematic phonics instruction was phased in slowly in different Local Authorities in England, and accordingly, it was possible to compare how children who were part of the systematic phonics trial compared with children who received standard instruction on various standardized language measures. In 2005 the "*Early Reading Development Pilot*" (ERDp) that involved 18 Local Authorities and 172 schools began, with each school receiving funding for a dedicated learning consultant who trained teachers in systematic phonics (typically for 1 year). Then in 2006, the "*Communication, Language and Literacy Development Programme*" (CLLD) that included a further 32 Local Authorities began, again with each school receiving 1-year funding for a dedicated learning consultant.

In order to assess the immediate efficacy of introducing systematic phonics, scores from the communication, language, and literacy components of Foundation Stage assessment were collected (when children completed Year 1 at age 5). And in order to assess the long-term effects of this intervention, reading scores from SATs Key Stage 1 (when children were 7 years of age), and reading scores from Stage 2 test (when children were 11) were collected. These are standardized tests given to all students in state schools, with teachers providing the assessment in the Foundation Stage and Key Stage 1, and the tests externally marked in Key Stage 2. Various statistical methods were used to control for the differences between the schools included in the trials and those not-included, and moderator variables included the impact of language background (Native English or not) and economic background (operationalized as children receiving or not receiving a free-school lunch).

For the ERDp sample the authors reported highly significant effect of systematic phonics on the Foundation Stage assessment immediately after the intervention (.298), but the effect dissipated on Key Stage 1 tests (.075), and was eliminated on the Key Stage 2 tests (-.018). Similarly, with the CLLD treatment, an initially robust effect (.217) was reduced on the Key Stage 1 tests (.017), and then was lost on the Key Stage 2 tests (.019). So much like the Suggate (2016) meta-analyses, the overall systematic phonics intervention effect did not persist. However, Machin et al. (2018) highlighted that the effects did persist in the Key Stage 2 tests in the CLLD treatment condition for non-native speakers (.068) and economically disadvantaged children as measured by their receipt of free school meals (.062), with both effects significant at the $p < .05$ levels. They took these small effects to show that phonics does provide long-term benefits for children who are in the most need for literacy interventions, writing:

> Without a doubt it is high enough to justify the fixed cost of a year's intensive training support to teachers. Furthermore, it contributes to closing gaps based on disadvantage and (initial) language proficiency by family background.

However, there are both statistical and methodological problems with using these findings to support the efficacy of systematic phonics. With regards to the statistics, apart from the fact that there were no overall long-term effects in either sample, it is important to note that the ERDp sample of children did not show significant advantage for non-native speakers (.045) or for economically disadvantaged children (.050) on the Key Stage 2 tests.

Indeed, for the ERDp sample, there was a tendency for more economically advantaged native English children (not in receipt of free school meals) to read more poorly in the phonics condition in the Key Stage 2 test (-.061), p < .1. As the authors write: "It is difficult to know what to make of this estimate" (p. 22). Note, the long-term negative outcome economically advantaged native English children in the ERDp sample was of a similar magnitude to the long-term benefits enjoyed non-native speakers (.068) and economically disadvantaged children (.062) in the CLLD treatment condition, and accordingly, is difficult to brush this finding aside.

More importantly, this study did not include the appropriate control condition. The advantages in Foundation and Key Stage 1 were the product intensive training support in systematic phonics to teachers in Year 1, but it is possible that similar outcomes would result if intensive training support was given to teachers in whole language instruction, or any other method. As was the case with most of the above meta-analyses, the conclusion the authors made was not even tested.

The Recent Success of English Children on PIRLS Provides Little or No Evidence for Systematic Phonics

A great deal of attention in the mainstream and social media has been given to the recent success of English children in the "Progress in International Reading Literacy Study" (PIRLS) carried out in 2016. PIRLS assesses reading comprehension in fourth graders across a wide range of countries every five years: 35 countries participated in 2001, 38 in 2006, 48 in 2011, and 50 in 2016. Many supporters of systematic phonics have noted how far up the league table England has moved since 2006 given that systematic phonics was mandated in English state schools in 2007, and phonics check was introduced in 2012. Specifically, England was in 15th position in 2006 (with a score of 539), joint 11th position in 2011 (score 552), and joint 8th in 2016 (score 559).

In response to the most recent results, Mr. Gibbs, the Minister of State at the Department for Education said:

> The details of these findings are particularly interesting. I hope they ring in the ears of opponents of phonics whose alternative proposals would do so much to damage reading instruction in this country and around the world.

A Department for Education report for the UK (December, 2016) reported:

> The present PIRLS findings provide additional support for the efficacy of phonics approaches, and in particular, the utility of the phonics check for flagging pupils' potential for lower reading performance in their future schooling.

Sir Jim Rose, author of the Rose (2006) report, used "the spectacular success of England shown in the latest PIRLS data" as further evidence in support of systematic synthetic phonics (Rose, 2017).

However, once again, these conclusions are unjustified. One important fact ignored in the above story is that English children did well in 2001, ranking third (scoring 553). Of the six countries that completed all the PIRLS tests from the beginning (England, New Zealand, Russian Federation, Singapore, Sweden, and USA), England has gone from second to third position. If the introduction of systematic phonics is used to explain the improved performance from 2006-2016, how is the excellent performance in 2001 explained? In addition, the results of the 2016 PIRLS were based on combing the performance of state and private

schools (private schools were not required to implement systematic phonics or use the phonics check). When only state schools are considered, performance dropped to 11th (rather than joint 8th), same as the 2011 PIRLS rating (Solity, 2018). Note, one of the common criticisms of systematic phonics is that the focus on phonology makes instruction less engaging. The PIRLS 2016 also ranked English children's enjoyment of reading at 34th, the lowest of any English-speaking country (Solity, 2018).

It is also interesting to note that Northern Ireland participated in the last two PIRLS, and they did better than England, ranking 5th and 6th in 2011 and 2016 respectively. This is relevant as the Reading Guidance for Key Stage 1 published by the "Northern Ireland Education & Library Boards" does not include the words "systematic phonics", nor do children complete a phonics screening check that was introduced in the UK to improve the administration of phonics in English schools. Of course, reading instruction in Northern Ireland does teach children letter-sound correspondences, but this is carried out along with a range of methods that encourage children to encode the meaning of words and passages. For instance, according to the Reading Guidance for Key Stage 1, when children encounter an unknown word, various strategies for naming the word are encouraged, including: phonics, using knowledge of context (semantics), using knowledge of grammar (syntax). This is similar to National Literacy Strategy in place in England from 1998 to 2006 that recommended phonics as one of four 'searchlights' for learning to read, along with knowledge of context, grammatical knowledge, word recognition and graphic knowledge. If the introduction of systematic phonics is used to explain the strong performance of England in 2016, how is the even better performance of Northern Ireland explained?

A final point worth emphasizing is that the PIRLS test assesses reading comprehension, and as noted above only 1 of the 12 meta-analyses reported a benefit for comprehension (NRP, 2000), and only at a short delay (ignoring the problems of this meta-analysis that question robustness of the short-term effect as well). Attributing any PIRLS gains to phonics is hard to reconcile with existing experimental research.

The Improving Performance on the Phonics Screening Check (PSC) in England Provides Little or No Evidence that Systematic Phonics Improves Literacy

Since 2012 the UK government has required all children in state schools in England to complete a phonics screening check (PSC) in Year1 in order "to confirm that all children have learned phonic decoding to an age-appropriate standard" (Department for Education, 2012; p. 4). The phonics screening check is composed of one- and two-syllable real words (e.g., day, grit, shin) and 20 pseudowords that can only be read on the basis of learned grapheme-phoneme correspondences (e.g., fape, blan, geck). Children near the end of Year 1 are asked to read the words and pseudowords aloud, with each item marked correct or incorrect. A child who correctly names aloud 32 items (80% of all items) is said the 'meet the standard', whereas a child who misses the standard is to be given further support to improve their phonics knowledge (and complete the phonics check again in Year 2).

Strikingly, the performance on the task has improved from 58% students meeting the standard in 2012 to 82% in 2018. This is taken to show that the PSC has improved the teaching of systematic phonics, and this in turn has improved decoding skills. The critical question is whether this has translated into better reading.

The obvious way to test whether the improved decoding skills translate to better reading is to compare the PSC results to the SATs carried out at Key Stage 1 and 2 during the years 2012-2017. These are the same tests analyzed by Machin et al. (2018) above (although

they analyzed data from before 2012).   And in fact, there have been some claims that improved performance on the phonics screening check are associated with improved performance on the SATS. For instance, Buckingham (2016) writes:

> There has also been an improvement in Key Stage 1 (Year 2) reading and writing results since the introduction of the Phonics Screening Check. The proportion of students achieving at or above the target reading level hovered around 85% from 2005 to 2011 but steadily increased to 90% in 2015. There was an even greater improvement in writing in the same period — a seven percentage point increase. (p. 16).

The results of the phonics check and the Key Stage 1 SAT scores are displayed in Figure 1.

But this characterization of the findings is inconsistent with a report from the Department for Education (Walker, Sainsbury,Worth, Bamforth, & Betts, 2015). The authors analyzed the reading and writing scores for the KS1 for the two years preceding and following the in introduction of the screening check and concluded:

> The evidence offered by these analyses is therefore inconclusive in identifying any impact of the [phonics screening check] on literacy performance at KS1 or on progress in literacy between ages five and seven.

Why the different conclusions? One key point to note is that although the SAT scores did start slowly increasing in 2012 (consistent with Buckingham, 2016), it is not possible to attribute these gains to the phonics screening check because these children completed Year 1 in 2011, and accordingly, were never given the PSC. As noted by Walker et al. (2015):

> These analyses of national data therefore indicate small improvements in attainment at KS1, which were a feature before the introduction of the check and continued at a similar pace following the introduction of the check.

In addition, as can be seen from Figure 2, there is little evidence that SAT scores for reading and writing improved more than SAT scores for maths or science between 2013-2015.[3]

Another important question that can be asked is whether the introduction of the phonics check was associated with improved reading skills at Key Stage 2 when children were in Year 6 (age 11). That is, did the improved teaching of phonics in Year 1 (in response to the PSC) have any long-lasting effect on reading outcomes? The results from 2017 provide the first relevant data given that children who completed these Key Stage 2 SATs were the first to complete the phonics check in 2012 in Year 1. As can be seen in Figure 2, the reading results when slightly down between 2016-2017 (while writing results went slightly up), and despite the large gains in the phonics check between 2012-2013 (an increase of 11%), there was no corresponding benefit in the reading and writing performance between the 2017-2018 SAT (e.g., the reading results improved 1% whereas the math results improved 2%). These findings show that the decoding skills of children (as measured by the PSC) did not support sort- or long-lasting reading improvement as measured by the Key Sage 1 and 2 SAT results. Note, the fact that the PSC scores have improved so dramatically demonstrates that systematic phonics instruction has improved in recent years in England. The fact that there is such a stark disconnect between PSC and SAT scores undermines the common claim that systematic phonics has improved literacy in England.

---

[3] Note, the tests changed in 2016, and this accounts for the large dip in performance.

The Program for International Student Assessment (PISA) results provide Little or No Evidence that Systematic Phonics Improves Literacy

PISA assesses 15-year-old school pupils' scholastic performance on mathematics, science, and reading. It has been carried out every three years since 2000, and critically, the 2018 version is the first to assess the reading outcomes of children in England who were taught systematic phonics in Year 1 as legally required post 2007. There is little or no evidence that this cohort of children have benefitted from this new policy as summarized in the PISA 2018 executive summary: "The mean scores in reading and science in England have not changed significantly over successive PISA cycles, but in mathematics, England's overall mean score showed a statistically significant increase compared with PISA 2015". It is also worth comparing the English results to other countries that use a range of teaching methods, with none requiring systematic phonics: Canada, Singapore, and Northern Ireland did better, whereas the USA, Australia, and New Zealand did equally well. So again, this challenges the claim that systematic phonics has improved reading outcomes in England.

In summary, despite the widespread claim that children are reading better in England since the mandatory inclusion of systematic phonics in state schools in 2007 and the introduction of the PSC in 2012, there is little or no evidence to support this conclusion. Indeed, the only noticeable change in performance is on the PSC itself, with no discernable effects on reading more generally. This is despite the fact that children in England have received early systematic phonics in the context of a broader literacy environment as recommended by the NRP (2000).

Conclusion

Despite the widespread support for systematic phonics within the research literature there is little or no evidence that this approach is more effective than many of the most common alternative methods used in school, including whole language. This does not mean that learning grapheme-phoneme correspondences is unimportant, but it does mean that there is little or no empirical evidence that systematic phonics leads to better reading outcomes. The "reading wars" that pitted systematic phonics against whole language is best characterized as a draw. The conclusion should not be that we should be satisfied with either systematic phonics or whole language, but rather, teachers and researchers should consider alternative methods of reading instruction. For example, one possibility is that reading instruction in English should focus more on the role that meaning plays in organizing spellings (via morphology), and that English spelling system makes sense once the interrelation between phonology, morphology, and etymology are considered (Bowers & Bowers, 2017, 2018c). Of course, other possibilities need to be considered as well, but the first step in motivating more research into alternative forms of instruction is to realize there is a problem with the current approach.

References

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print.* Cambridge, MA: The 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*, 629-653.

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. DOI: 10.1080/00461520.2017.1288571

Bowers, J.S. & Bowers, P.N. (2018a). There is no evidence to support the hypothesis that systematic phonics should precede morphological instruction: Response to Rastle and colleagues. PsyArXiv. DOI:https://psyarxiv.com/zg6wr/

Bowers, J.S., and Bowers, P.N. (2018b).   The importance of correctly characterizing the English spelling system when devising and evaluating methods of reading instruction: Comment on Taylor, Davis, and Rastle (2017). *Quarterly Journal of Experimental Psychology, 71*, 1497-1500. https://doi.org/10.1177/1747021818759477

Bowers, J.S., and Bowers, P.N. (2018c). 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, P. N., & Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing: An Interdisciplinary Journal, 23*, 515–537. http://dx.doi.org/10.1007/s11145-009-9172-z

Browder, D. M., & Xin, Y. P. (1998). A meta-analysis and review of sight word re-search and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education, 32(3)*, 130-153. https://doi.org/10.1177/002246699803200301

Buckingham, J. (2016). Focus on Phonics: Why Australia should adopt the Year 1 Phonics Screening Check. *Centre for Independent Studies*. Research Report 22.

DOI: 10.13140/RG.2.2.35472.61440

Camilli, G., Kim, S., & Vargas, S. (2008). A Response to Steubing et al., "Effects of systematic phonics instruction are practically significant": The origin of the national reading panel. *Education Policy Analysis Archives, 16(16)*, 1–17. http://dx.doi.org/10.14507/epaa.v16n16.2008

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., M. Wolfe, P., & 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

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, 12, 169–190. http://dx.doi.org/10.1023/A:1008131926604

Department of Education, Science and Training (2005). Teaching Reading. National Inquiry into the Teaching of Literacy. Canberra: Commonwealth of Australia.

Ehri, L. C. (1997). Phases of development in learning to read by sight. *Journal of Research in Reading, 18*, 116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

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. http://dx.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. http://dx.doi.org/10.3102/00346543071003393

Foorman, B., Francis, D., Novy, D., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology, 83*, 456-469.

Frith, U. (1985). Beneath the surface of surface dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). London: Routledge and Kegan Paul.

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

Gittelman, R., & Feingold, I. (1983). Children with reading disorders—I. Efficacy of reading remediation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 24*, 167-191.

Gooch, D., Hulme, C., Nash, H. M., & Snowling, M. J. (2014). Comorbidities in preschool children at family risk of dyslexia. Journal of Child Psychology and Psychiatry, 55(3), 237-246.

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.

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

Henry, M. K. (1989). Children's word structure knowledge: Implications for decoding and spelling instruction. *Reading and Writing: An Interdisciplinary Journal, 2,* 135–152. http://dx.doi.org/10.1007/BF00377467

Her Majesty's Inspectorate (HMI) (1990) *The Teaching and Learning of Reading in Primary Schools*. London: Department of Education and Science (DES).

Larkin, R. F., & Snowling, M. J. (2008). Morphological spelling development. *Reading & Writing Quarterly, 24,* 363–376. http://dx.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

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

Lim, L., & Oei, A. C. (2015). Reading and spelling gains following one year of Orton-Gillingham intervention in Singaporean students with dyslexia. British Journal of Special Education, 42, 374-389. https://doi.org/10.1111/1467-8578.12104

Machin, S., McNally, S., & Viarengo, M. (2018). Changing how literacy is taught: evidence on synthetic phonics. *American Economic Journal: Economic Policy*, *10*, 217-241. DOI: 10.1257/pol.20160514

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2013). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities, 48*, 391–407. http://dx.doi.org/10.1177/0022219413504996

McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., &, et al. (2012, December 12). Phonics training for English speaking poor readers. *Cochrane Database of Systematic Reviews,* CD009115

McArthur, G., Kohnen, S., Jones, K., Eve, P., Banales, E., Larsen, L., & Castles, A. (2015). Replicability of sight word training and phonics training in poor readers: A randomised controlled trial. *PeerJ, 3*, e922. https://doi.org/10.7717/peerj.922

Moats, L. C. (2000). Whole language lives on: The illusion of "balanced" reading instruction. Retrieved from http://www.ldonline.org/article/6394/

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.

Nunes, T., Bryant, P. E., & Bindman, M. (1997). Morphological spelling strategies: Developmental stages and processes. *Developmental Psychology, 33*, 637–649. http://dx.doi.org/10.1037/0012-1649.33.4.637

Pearson, P. D. (2004). The reading wars. *Educational policy, 18*, 216-252. https://doi.org/10.1177/0895904803260041

Rose, J. (2006). *Independent Review of the Teaching of Early Reading.* Nottingham: DfES Publications

Rose, J. (2017). War and Peace in Reading – Tie for a Truce? Downloaded: https://iferi.org/war-and-peace-in-reading-time-for-a-truce-by-sir-jim-rose/

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

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

Solity, J. (2018). Commercial Systematic Synthetic Phonics Programmes: A Possible Cause of Pupils' Literacy Difficulties. Poster presented at *Society for the Scientific Study of Reading*, Brighton.

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

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*, 1556-1579. DOI: 10.1037/a0020612

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(1)*, 77-96. https://doi.org/10.1177/0022219414528540

Swanson, C. B., & Barlage, J. (2006). *Influence: A study of the factors shaping education policy.* Bethesda, MD: Editorial Projects in Education Research Center.

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General*, *146*(6), 826. http://dx.doi.org/10.1037/xge0000301

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2018). Phonics: reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education,* 1-31. https://doi.org/10.1080/02671522.2017.1420816

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.

Treiman, R. (2018). Teaching and Learning Spelling. Child Development Perspectives, 4, 235-239. https://doi.org/10.1111/cdep.12292

Umbach, B., Darch, C., & Halpin, G. (1989). Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. *Journal of Instructional Psychology, 16*, 23-30.

van der Kleij, S. W., Segers, E., Groen, M. A., & Verhoeven, L. (2017). Response to Intervention as a Predictor of Long-Term Reading Outcomes in Children with Dyslexia. Dyslexia, 23, 268-282. https://doi.org/10.1002/dys.1562

Figure Captions

Figure 1. Results on Keys Stage 1 SAT tests in reading, writing, maths, and science from 2006-2018 as well as the results of the phonics screening check from 2012-2018.   SAT scores to the left of vertical dashed line were achieved without having completed the phonics screening check in Year 1, and SAT scores to the right of the vertical dashed lined were achieved after having completed the phonics check in Year 1. According, the improved SAT results on reading and writing between 2011-2012 cannot be attributed to the improved administration of phonics.

Figure 2. Results on Keys Stage 2 SAT tests in reading, writing, maths, and science from 2007-2018 as well as the results of the phonics screening check from 2012-2018. SAT scores to the left of vertical dashed line were achieved without having completed the phonics screening check in Year 1, and SAT scores to the right of the vertical dashed lined were achieved after having completed the phonics check in Year 1. The finding that SAT reading results did not improve between 2016-2018 indicates that the improved administration of phonics in Year 1 (starting in 2012) did not have a long-term impact on the SAT scores.
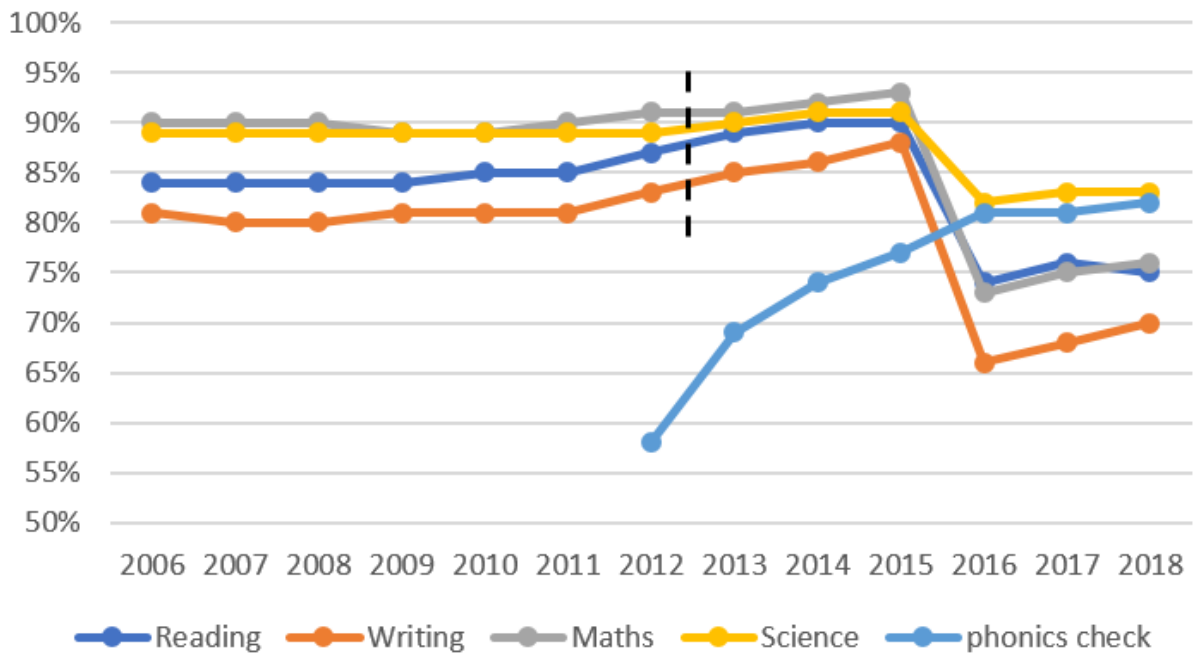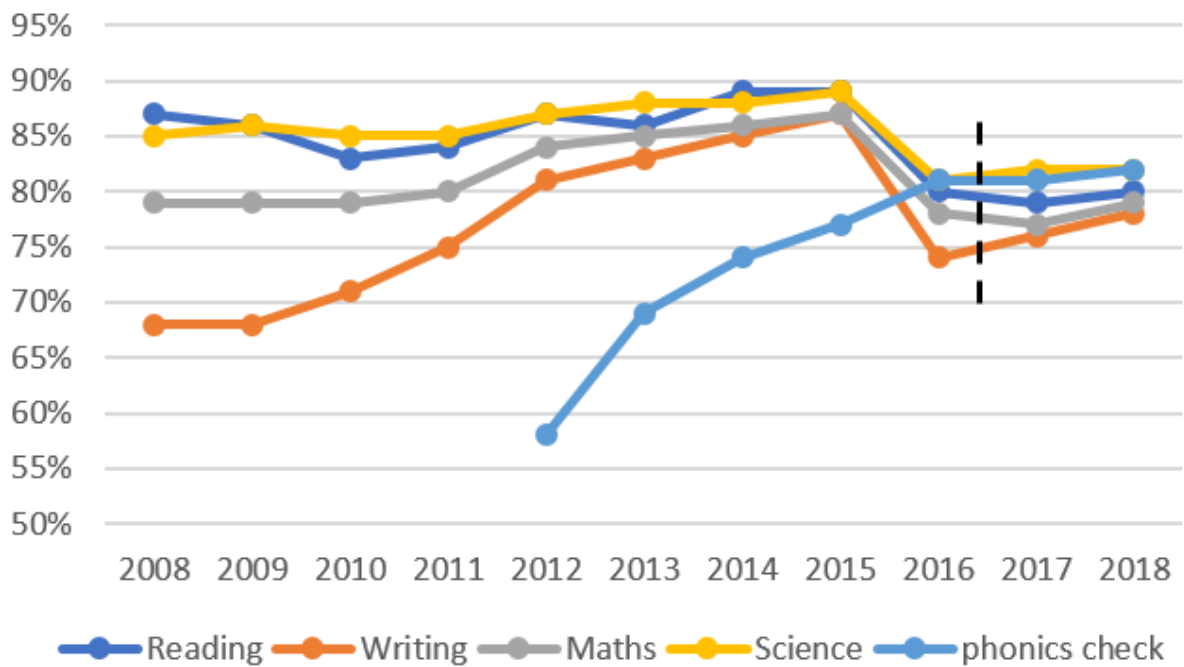
Figure 1

Figure 2

Table 1. Reading outcomes across a range of measures in the NPR (2000), Camilli et

al. (2003, 2006), and Torgerson et al. (2006) datasets

	NPR (2000) Immediate	NPR (2000) Delay	Camilli et al. (2003)	Camilli er al. (2006)	Torgerson et al. (2006)	Torgerson et al. (2006) updated estimate
All measures	0.41	0.27	0.31	0.15; NS	0.27	NR
Naming regular words	0.67	NR	NR	NR	NR	NR
Psueudoword naming	0.6	NR	NR	NR	NR	NR
Reading miscellaneous words	0.4	NR	NR	NR	0.27-.38	.20-.21; NS
Spelling words	0.35	NR	NR	NR	.09; NS	NR
Reading text	0.25	NR	NR	NR	NR	NR
Comprehension	0.27	NR	NR	NR	.24-.35; NS	NR
Low Achieving readers	0.15	NR	NR	NR	NR	NR
Reading disabled	0.32	NR	NR	NR	NR	NR