

A commentary on Bowers

**Note: Educational Psychology Review, in press. Author accepted manuscript.
Prepublication version. Not for reproduction, copying, or distribution without permission**

A commentary on Bowers (2020) and the role of phonics instruction in reading

Jack M. Fletcher

Department of Psychology

University of Houston

Robert Savage

Department of Psychology and Human Development, Institute for Education

University College- London

Sharon Vaughn

Department of Special Education

University of Texas-Austin

Corresponding author: Jack M. Fletcher, Ph.D., jackfletcher@uh.edu

Acknowledgements

We would like to thank Professor Genevieve McArthur for discussion of ideas in this paper and Dr. Yusra Ahmed for assistance with the computation of effect sizes.

Declarations

Funding: Grant P50 HD052117, Texas Center for Learning Disabilities, from the Eunice Kennedy Shriver National Institute of Child Health & Human Development, supported Fletcher and Vaughn. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Conflicts of Interest: None

Ethics approval, consent to participate, availability of data, code availability: Not applicable

Abstract

Bowers (2020) reviewed 12 meta-analytic syntheses addressing the effects of phonics instruction, concluding that the evidence is weak to nonexistent in supporting the superiority of systematic phonics to alternative reading methods. We identify five issues that limit Bowers' conclusions: 1. *Definition issues*; 2. *What is the right question?* 3. *The assumption of "phonics first"*; and 4. *Simplification of issues around systematic versus explicit phonics*. We then go on to consider 5. *Empirical issues in the data from meta-analyses*, where Bowers misconstrues the positive effects of explicit phonics instruction. We conclude that there is consistent evidence in support of explicitly teaching phonics as part of a comprehensive approach to reading instruction that should be differentiated to individual learner needs. The appropriate question to ask of a 21st century science of teaching reading is not the superiority of phonics versus alternative reading methods, including whole language and balanced literacy, but how best to combine different components of evidence-based reading instruction into an integrated and customized approach that addresses the learning needs of each child.

Keywords: phonics, whole language, reading instruction, morphology, National Reading Panel, dyslexia

A commentary on Bowers (2020) and the role of phonics instruction in reading

Introduction

This paper is a commentary on the analysis provided by Bowers (2020) of the effects of systematic phonics instruction. Bowers concludes that the evidence supporting the superiority of systematic phonics to what he describes as alternative reading methods is weak to nonexistent. He argues that any presumed scientific consensus on how to teach reading is premature and based on unsettled science devoted to rhetoric on the superiority of systematic phonics. He concludes:

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.

In this paper, we suggest that in drawing this conclusion, Bowers (2020) is not asking the correct question, makes a number of unwarranted definitional assumptions, and partly because of these problems, misconstrues the available evidence. The title of Bowers paper and his narrative above implies that he will compare phonics approaches with *all other* available methods. However, Bowers does not present data on all other methods, but instead revisits and attempts to re-analyze existing reviews that have in the main, but not exclusively, contrasted phonics with wider language arts interventions. Most broadly, we thus argue that the question of systematic versus incidental or no phonics instruction is outdated. Furthermore, this question does not build on the knowledge and evidence of the other methods Bowers briefly mentions. These views do

not help legislators, educational leaders, researchers, or teachers make the decisions needed to implement effective reading practices.

While we seek to surface some of the problems we identify with Bowers' analysis, wherever we can, we identify points of agreement. We argue, for example, that Bowers is correct that the value of systematic phonics over morphological training or meaning-based training *in isolation* is often exaggerated. However, as the difficulties both Bowers and Camilli et al. (2006) observed in coding experimental studies to isolate the effects of phonics instruction, most reading interventions are rarely restricted to a monolithic view of systematic phonics. We conclude that the correct question to ask with the evidence that we now have is: How do we best combine different components of evidence-based reading instruction into an integrated, comprehensive approach that can be customized according to the needs of each learner? This question recognizes the 21st century science of teaching reading that will involve explicit phonics, other forms of sublexical instruction, and all aspects of language and knowledge.

We first provide an overview of Bowers (2020), and then identify five issues that limit his conclusions. Three of these are conceptual issues related to definition, the question of interest, and the assumption of "phonics first." We also discuss two challenges involving simplification of issues concerning systematic versus explicit phonics instruction and empirical issues related to interpreting the meta-analyses. We provide data that supports our contention that contemporary instructional research has moved beyond the comparisons made by Bowers.

Overview and Background

Bowers (2020) constructed the paper in three sections. First, he reviewed different methods of reading instruction, defining systematic phonics as an approach that *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. 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.*

Second, Bowers provided an extended review of 12 meta-analyses of “experimental” reading research addressing the role of systematic phonics instruction. Third, he interpreted the results of what he described as a large, naturalistic experiment emanating from a mandate in England in the United Kingdom (UK) to teach systematic phonics. We do not address findings arising from US or UK government policy. While important, policy issues raise distinct questions that deserve separate treatment. Instead, we focus on the scientific research and the meta-analyses that potentially inform evidence-based practice in teaching of reading worldwide.

Commentary

Definition Issues

We interpret Bowers (2020) statement about phonics approaches to reading instruction as being driven by phonology first and meaning second – whereas, whole language and balanced literacy approaches are meaning first – as being an inaccurate portrayal of reading instruction

research in the last several decades. Most approaches to reading instruction that include explicit phonics are also focused on meaning and understanding the words and texts used. The misunderstanding of “meaning first” in whole language/balanced literacy instruction is the over emphasis on teaching students to guess rather than using what is known about the alphabetic principle to read the word(s). The assumption that explicit phonics is based on phonics first and meaning second and whole language/balanced literacy is meaning first and phonics second does not accurately reflect our interpretation of intervention research nor of educational practice. Bowers also seems to mix his concerns about phonics first versus meaning first with a different issue, which is whether phonics instruction should precede instruction in morphology, which is suffused with meaning (Bowers & Bowers 2017).

Whole language and its current iteration, balanced literacy, mean different things to different people and are difficult to define. These approaches often stem from the misunderstanding that skilled adult readers do not explicitly use sublexical strategies to identify words. Therefore, beginning readers should not break words apart. Instead, they should focus on the whole word and its meaning. Instead of “teaching” children to read, the teacher facilitates reading development by providing rich and authentic reading experiences through immersion in age-appropriate literature. Phonics instruction, if provided at all, should be minimal and incidental depending on children’s needs as they encounter text (Fountas & Pinnel 2012-2013). Although some balanced literacy programs are shifting, they do not clearly embrace explicit phonics instruction (Student Achievement Partners 2020)

Although Bowers defines whole language as an inclusive approach that incorporates some phonics among other approaches, whole language is a broad label for a range of pedagogical philosophies, often associated with an understanding of learning to read that does

not routinely require application of the alphabetic principle. It is not hard to find influential advocates who eschew the teaching of phonics altogether (e.g., Smith 2004). If one really were set on coding the phonics delivered within whole language interventions (and we are not), it might justifiably span from none at all to unsystematic, incidental, and infrequently used. The extent to which teachers followed the extreme version of whole language advocated for example by Smith behind their classroom doors is ultimately unknowable.

Bowers' starting point is to question the findings of Chapter 2 Part II of the 2000 US National Reading Panel report (NRP; NICHD 2000) that addresses the role of phonics instruction. As part of multiple meta-analytic and narrative summaries of different domains of reading instruction, the NRP report found through meta-analysis that systematic phonics is an effective component of early reading instruction compared to some/no phonics instruction. The comparisons in the NRP report were studies that examined the relative effects of approaches to reading that emphasized systematic phonics compared to conditions in which incidental or no explicit phonics was utilized. As Bowers acknowledges, the conclusions reached by the NRP were much broader than those involving the single chapter on phonics. The NRP recognized that learning to read required multiple competencies in phonics, phonological-awareness, fluency, vocabulary, and comprehension, and considering the role of teacher professional development and the role of technology in six equally detailed chapters.

What is the Right Question?

Like many contemporary accounts of the best methods for reading instruction, we interpret many of Bowers' (2020) arguments as establishing, perhaps unintentionally, a dichotomy of reading methods (i.e. phonics versus whole language). Bowers states in his main conclusion that, "*The "reading wars" that pitted systematic phonics against whole language is*

best characterized as a draw.” Although Bowers also uses broader terms such as “alternative reading methods,” simple comparisons of phonics versus meaning versus morphology for teaching reading to determine which one is best is an outmoded comparison (Snow & Juel 2005). Most empirically minded reading researchers, seemingly including Bowers, would argue that none of these approaches is adequate in isolation. In fact, as Bowers (2020) and Camilli et al. (2006) demonstrated, it is hard to find pure comparisons of approaches that could be called “phonics-only,” “morphological-only,” and “meaning –only.” Even in these comparisons, the interventions are more complex than simply teaching phonics or meaning. We believe current reading intervention research generally embraces this complexity, recognizing the value of both explicit phonics instruction integrated with fluency, language, and comprehension practices that reflect the necessary complexity of reading instruction at both the sublexical and lexical level.

Beyond this broad position, Bowers first correctly claims that in some systematic phonics programs, all of the relevant grapheme-phoneme mappings are first taught explicitly in a prescribed, ordered manner. He then claims, *“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).”* However, there is no shared position among all advocates of systematic phonics requiring that practices such as using authentic “real” books be eschewed. Indeed, it is quite possible to find many examples of studies carried out in England, Canada, New Zealand, and Hong Kong that show that highly systematic prescribed phonics instruction is more effective when integrated with real books compared to more de-contextualized phonics (Chen & Savage, 2014; Hatcher, Hulme, & Ellis, 1994; Shapiro & Solity, 2008; Tse & Nicholson, 2014; Yeung & Savage, 2020).

There are very good scientific reasons for asserting that the whole language/balanced literacy argument that learning to read is a “natural” process is misguided. This misguided idea promotes the misunderstanding that all teachers and parents need to do is provide youngsters with supporting materials and environments and reading will develop. Unlike language, reading is not an evolutionary process that is simply activated by immersion in literature or exposure to words (Liberman 1996; Seidenberg 2017). Thus, learning to read is not a developmental process that occurs for all individuals as long as the environment is conducive. Rather, reading development is an acquired skill that for most students requires carefully organized instruction and for some students, very explicit and customized instruction.

As noted above, contrasting systematic phonics with reading approaches that teach phonics incidentally or not at all cannot be adequately addressed through empirical syntheses and inadequately addresses the range of learners and their instructional needs. We expect that Bowers would agree that few studies provide adequate information so that they can be coded well enough to characterize all the components of reading instruction to determine the contrasting conditions. One issue that reading researchers struggle with is incorporating sufficient detail within the ever-restricted academic journal word count restrictions. As McArthur et al. (2019) noted, this lack of detail impedes meta-analysis generally and needs to be improved. It is particularly hazardous to try to code papers post hoc into new categories based on such limited information (Camilli et al. 2006). Furthermore, many intervention studies utilize phonics as one of several components in the intervention. The phonics component varies and may vary in systematicity, but the approaches that show the strongest effects are explicit and intentional (Stuebing et al. 2008). The reading program may include instruction at the sublexical level that includes work on morphology, syllables, and larger units of words. However, phonics instruction

typically includes word work that teaches common rules regarding phoneme-grapheme correspondence. There are additional moderators to consider that affect outcomes, such as group size, other language and literacy activities, and the amount of time in general education and supplemental intervention (Camilli et al. 2006).

To illustrate some of these complexities, we summarize findings from a study that examined the relative effects of three treatment conditions with various emphases of systematic phonics with and without an emphasis on deeper analysis of words, including morphology and meaning, beyond teaching phoneme-grapheme rules for decoding (Morris et al. 2012). We selected this study because the conditions approximate the effect of systematic phonics versus a no phonics control and shows the advantages of more comprehensive approaches to sublexical and meaning-based instruction.

Utilizing a randomized controlled trial (RCT), 279 children in Grades 2-3 with significant word reading problems were randomly assigned to one of three small group treatment conditions or a comparison condition in which the researchers provided no reading instruction (see Morris et al. 2012, Figure 1, p. 103 for a flow diagram of the assignments in the study). Morris et al. used a program called Phonological Analysis and Blending/Direct Instruction (PHAB) that teaches grapheme-phoneme relations using Direct Instruction methods (Engelmann & Bruner 1988). PHAB was compared with Phonological and Strategy Instruction (PHAST; Lovett et al. 2000), which includes PHAB and strategy-based sublexical instruction. Children are taught five strategies for word recognition: phonological letter-sound decoding through PHAB (Sounding Out); word identification by analogy (Rhyming), how to separate affixes in multisyllabic words (Peeling Off), how to seek familiar parts of unfamiliar words (I Spy), and how to attempt variable vowel pronunciations (Vowel Alert). A third treatment condition was the Retrieval,

Automaticity, Vocabulary, Engagement with language, and Orthography program (RAVE-O; Wolf et al. 2000). The RAVE-O program included systematic decoding instruction through PHAB, but also targeted four linguistic systems essential to reading proficiency: orthography, semantics, syntax, and morphology. Both PHAST and RAVE-O are based on empirical research and theory showing that the more a child knows about the structures of words, including meaning, the faster the child can decode and understand the word (Lovett et al. 2000; Seidenberg, 2017; Wolf & Katzir-Cohen 2001). The study also included a comparison group that received no reading instruction from the researchers, instead receiving 70 hours of math instruction and “classroom survival skills” (CSS).

Altogether, there were four groups in the study: PHAST (+ PHAB), RAVE-O (+ PHAB), PHAB+ CSS, and Math + CSS. The relevant comparisons for this paper involve PHAB + CSS with + Math + CSS to estimate the effects of systematic phonics instruction and PHAST (+PHAB) with PHAB + CSS to estimate the effects of the more complex approaches to word reading. We did not include comparisons of PHAST and RAVE-O or PHAST and RAVE-O because PHAST and RAVE-O showed similar outcomes after 70 hours of instruction.

Table 1 shows effect sizes (Hedge’s g) for outcomes involving word attack, word identification, word reading fluency, spelling, a cloze-based passage comprehension measure, and measures of reading accuracy, fluency, and comprehension. These effect sizes were computed from Table 3 in Morris et al. (2012), which provided raw scores for each of these measures at baseline, 35 and 70 hours of intervention, and a one-year follow-up. We focused on outcomes at the end of treatment (i.e., 70 hours) as the most reasonable estimate of effectiveness.

For the PHAB+CSS compared with Math+ CSS comparison, Table 1 shows effect sizes that are consistently positive, with the largest for word attack skills ($g = 0.32$), word

identification ($g = 0.22$), spelling ($g = .25$), and text reading fluency ($g = 0.24$) and comprehension ($g = 0.25$). These effect sizes meet conventional levels for practical significance, but the confidence intervals include 0 and would not meet conventional standards for statistical significance. Although these are not robust differences in a single study for a single component of reading instruction, the study is underpowered to detect significant effect sizes in this range, especially in older poor readers. Bowers would interpret these findings as consistent with his argument that there is no advantage of systematic phonics. However, the effect sizes are consistent with the meta-analytic evidence of a positive effect size for systematic phonics and large enough to be practically significant, cautioning against acceptance of the null hypothesis. These effect sizes should not be construed as effect sizes for Direct Instruction, which is a more complex reading program with good evidence of efficacy in general education and as a remedial intervention (Stockard & Wood 2017).

When comparing PHAST (+PHAB) and PHAB+CSS, Table 2 shows much larger effect size advantages for PHAST (+ PHAB) across multiple outcomes involving decoding, fluency, and comprehension, with confidence intervals that do not include 0. The effect for text-based comprehension are negligible, although the effect size for a cloze-based comprehension measure was larger. Note that comparisons of PHAST (+ PHAB) to MATH+CSS would be even larger.

These results show the benefits of more in-depth instruction in word work for poor readers compared to systematic phonics instruction. However, both PHAST (+PHAB) and RAVEO (+ PHAB) programs spent half the lesson time on systematic phonics (PHAB), controlling the amount of phoneme-grapheme training in the three groups. Despite major differences in PHAST and RAVE-O, the latter providing much more explicit instruction in morphology and vocabulary, outcomes were generally similar for the two programs, with both

clearly superior to PHAB+CSS. In evaluating the conditions, would Bowers characterize all three interventions as examples of systematic phonics? How would he deal with the fact that the extended activities occur *simultaneously* and are interactive, in contrast with the phonics-first idea? Is it plausible to conclude that poor readers selected for inadequate word reading do not benefit from explicit instruction in phonics? This study is an example of the complexity of coding the *next generation* intervention approaches that recognize the importance of more complex approaches to sublexical instruction that incorporate meaning based instruction with explicit phonics instruction.

The Assumption of “Phonics First”

Morris et al. (2012) blended systematic phonics instruction with other approaches to accessing the sublexical structure of the word and its meaning. These activities are combined in many reading programs. For example, Savage et al. (2018) taught at-risk readers both ‘direct mapping’ of grapheme-phoneme rules to authentic texts and a meaning-based “set for variability” strategy for inconsistent rules. However, Bowers argues that proponents of systematic phonics assert that phonics instruction should (and indeed, must) occur before any focus on larger units of words (e.g., morphology) or on meaning. Bowers argues:

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.

We agree that grapheme-phoneme correspondence as a mechanism for learning about the form of a word is valuable and that knowing word meaning while learning to read words is valuable. We also recognize that the units in which phonics approaches are used vary. For example, many instructional practices use the “rime” as the building block for phonics applying consonants, blends, and digraphs to extend decoding and generalize phonics rules, e.g., /at/ extended through /c/ to cat or /b/ to bat or /ch/ to chat. Co-teaching mechanisms for better capitalizing on word meaning and at the same time teaching students to acquire word reading practices are valuable. However, we do not agree that the phonics first approach Bowers alludes to represents the widespread view of reading scholars – even those who recognize the impact of phonics instruction. If Bowers is questioning the idea that phonics instruction is limited without attention to other forms of sublexical instruction and word meaning, including morphology, we agree. If he is questioning the idea that all phoneme-grapheme correspondence rules must be learned before exposure to spellings and meanings of words, we agree. If he is arguing that initial reading instruction can proceed based solely on morphology, we disagree. 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 and does not need to emphasize phonics before morphological training or exposure to word meanings (Morris et al., 2012). After all, youngsters come to school with an abundance of knowledge of word meanings that serve as valuable capital for teaching words and understanding text.

The role of morphology may be somewhat different and involves how and when exposure to increasingly large units of words should occur. Both the PHAST (+ PHAB) and RAVE-O (+PHAB) programs included morphological (and other strategy) instruction

simultaneously with systematic phonics, varying the emphases as the child proceeded through instruction. This issue may also have to do with the transparency of the language, its orthographic representation in the written word, and the size of the sublexical unit. In more transparent languages, access to larger sublexical units may occur more quickly in beginning reading because there are fewer exceptions to letter-sound correspondence rules compared to a less transparent language, such as English. Even in English, and even if phonics is emphasized, children are taught and exposed to vocabulary; they engage with grammatical forms including inflectional morphology routinely; they learn new things about the world, improving their background knowledge as they increase their word knowledge. This symbiotic relation between word knowledge building and world knowledge building is an essential feature of schooling and a necessary foundation for enriching text comprehension. If children are exposed to “real books,” they will necessarily engage with both morphology and phonology. Inspection of the children’s printed word database (Masterson et al. 2008) shows that some 32% of words in texts written for children in kindergarten are bi-syllables (including common morphemic inflections: “-ing,” “-ed,” and “-s”), 5% are trisyllables, and 0.7% have four or five syllables.

There should be a focus on both language development and strategies for understanding what is heard and read in any evidenced and truly “balanced” model of reading, such as in the Simple View of Reading (e.g., Gough & Tunmer 1986) and its expression in reading instruction (Savage 2020). Similarly, children learn to read print by accessing sublexical parts of words, linking what words sound and look like. This is true in a beginning reader and in an illiterate adult (Seidenberg, 2017). From a theoretical view, the Simple View is a theory of reading comprehension that shows that children must learn about the form and meaning of the word. These occur in parallel, although children vary in their capacity for learning decoding and

language comprehension and the developmental process of dealing with print emphasizes word learning more in early grades and comprehension in later grades. The Simple View is consistent with Perfetti's (2007) lexical quality hypothesis, where acquiring and integrating information about both word form and meaning are necessary for on-line reading comprehension. In computational models of reading (See Seidenberg, Borkenhagen, & Kearns 2020), the well-known triangle model specifies pathways involving orthography, phonology, and semantics to compute word meanings. These pathways are trained using computational modeling and statistical learning theory. The models show that in order to deal with a written input (spelling of words in an alphabetic language), the models initially use an orthographic to phonologic to semantic pathway. With training, the direct pathway from orthography to semantics becomes more prominent, but the orthographic to phonology pathway continues to contribute. Models that train the orthography to semantics pathway take much longer than models that involve phonology. At some point, as the brain begins to pick up on the statistical regularities of the spellings, learning becomes more implicit than explicit, consistent with Share's (1995) self-teaching hypothesis. According to Share, orthographic representations are acquired by self-teaching through phonological recoding of new letter strings that becomes increasingly implicit. In all these theoretical perspectives, orthography and phonology serve to access meaning and are interactive.

Empirical evidence on the relative role of morphological, phonetic, and other teaching strategies comes from a recent systematic review (Galuschka et al. 2020). The authors directly investigated what we know about the "phonics first" question in spelling interventions for children with dyslexia, where similar arguments have been made. Galuschka et al. report an exploratory analysis of the limited number of available studies. This analysis suggested no

significant advantage for early phonics over other programs, where RCTs exist. Interestingly, non-significant trends towards declining effectiveness of phonics and rising effectiveness of morphological training were evident with both age and severity of disability. Galuschka et al. concluded that we need more studies to resolve this question definitively. It is important to recognize that Galuschka et al. also reported robust main effects of spelling approaches based on phonics on both reading ($g = .62$) and spelling ($g = .68$), with morphological and orthographic approaches having moderate effects on spelling, but much smaller effects on reading. This work both further confirms but also contextualizes the role of phonics instruction in struggling readers.

Bowers is correct if he equates “phonology first” with a requirement to learn the entire corpus of phoneme-grapheme correspondence rules before dealing with morphology or meaning, but wrong in denying that phonics is an effective way of facilitating the implicit learning that must occur for the child to become an efficient reader (Seidenberg et al. 2020). It is also not obvious how children would learn spelling-sound correspondence rules initially through “morphology first,” which is likely why Bowers indicates in different places that learning phoneme-grapheme correspondence rules is important. We encourage Bowers to design and conduct the studies needed to address his hypotheses but not to advance his hypotheses as facts until his findings support this view. Phonological awareness is likely necessary in early reading, but not sufficient, alongside morphological and orthographic awareness, but the phonological component made explicit by phonics is likely to be essential for children with poor initial phonological awareness to adequately progress in reading (Berninger et al. 2010).

Systematic versus Explicit Phonics

Although the term systematic phonics is widely used, it is instructive to think about what systematic really means and whether the NRP question of systematic phonics versus less

systematic phonics is outmoded. In general, systematic refers to an organized structure – in this case, the organized structure for teaching the grapheme-phoneme correspondence rules. It is presumed that these rules are prescribed and often taught in an accepted sequence. However, the necessary sequence is not well established through research but rather derived from practices that provide ready access to reading words. Consonants such as /m/ and /s/ are taught early as they make sounds that sound like their letter names, are readily learned by young readers, and provide ready access to word reading- as do short vowels such as /a/ and /e/. These approaches are epitomized by the different methods under the umbrella of structured literacy (Moats, 2019).

We agree with Bowers' recognition of the limited evidence in support of a specific scope and sequence in which the child has to learn grapheme-phoneme rules as a prerequisite for decoding. As Morris et al. (2012) demonstrated, there are many ways to teach decoding, and the level of systematicity is related to the needs of the learner. Some students inferentially learn the pattern and rules through exposure to common word types; other students need more explicit instruction that is facilitated by an organized set of lessons. However, this organized sequence of lessons does not require that particular sound units be taught in any particular order as long as they are taught in ways that readily allow children to access print, words, and text.

We also agree with Bowers (2020) that phonics instruction can be effectively accomplished with approaches that would not meet the definition of systematic phonics that he provided, but not with the idea that any approach to phonics instruction will be effective. There is research that shows little difference in reading outcomes for methods based on a rigid scope and sequence versus methods in which the phonics instruction is explicit and mandatory, but embedded in reading and writing (Mathes et al. 2005; Torgesen et al. 2001), at least for many learners. Effective methods share explicit instruction in phonics, other sublexical approaches,

and language-rich activities, that through differentiated instruction responds to students' needs (Stuebing et al. 2008). Effective approaches to phonics instruction provide word work that is explicitly lead, intentional, and demonstrated by the teacher using a lesson that can be scripted or embedded, but which is explicit. It provides opportunities for students to respond, teacher feedback, and examples that both extend the principle as well as challenge it. It differs from what is undertaken for word work in many balanced literacy programs because of the amount of time spent on word work and the intentional nature. Instruction is likely to vary across individual children. The seven randomized trials by Carol Connor and colleagues (summarized in Connor & Morrison 2016) clearly demonstrate the value of *differentiating* the relative emphasis of code-based and meaning-based instruction for the individual child using materials in place in the school as opposed to a specific curriculum. In these studies, the amount of time devoted to each broad component was leveraged in the general education classroom such that less capable decoders spend more time on teacher directed code-related activity and more capable decoders spend more time on student directed meaning-related activity). As Stuebing et al. (2008, p. 132) stated,

...the explicitness of instruction may be more important than systematic, scripted lessons in accounting for the effect of systematic phonics. Creating a scope and sequence, using decodable text, and engaging in other ways of systematizing instruction make instruction explicit, but explicitness can be achieved in other ways. Where a teacher operates on the instructional continuum may depend on factors like preparation, experience, the base rate of struggling readers, the school context, and related factors. However, teachers need to be intentionally clear about how the alphabet relates conventionally to sound

segments in speech. The supporting materials that are used may vary depending on teacher and student knowledge and skills.

Rather than minimizing the effects of phonics instruction, we should be thinking more about how individual learners are responding to the methods that are used and be prepared to change instructional approaches based on their response to instruction (Connor & Morrison, 2016).

Empirical issues in the Data from the Meta-Analyses

At the heart of Bowers' analysis is his critique of the evidence from 12 meta-analyses of reading interventions. We argue that the Bowers analysis of these meta-analyses does not accurately represent the findings. We treat each of the 12 analyses in the order Bowers considers them describing his substantive analysis and with reasoned responses to each of his points.

The National Reading Panel (NRP; NICHD, 2000) and Ehri et al. (2001) meta-analyses. The NRP phonics chapter considered RCT, matched control trials of typical and atypical development, and interventions for word reading, fluency and comprehension outcomes. Ehri et al. (2001) published the chapter as a peer-reviewed paper. We will refer to both as NRP hereafter. Bowers makes four points in his criticism:

(1) The NRP analysis shows that phonics does not help low achieving poor readers - those whose cognitive level was below average ($d = .15, n.s.$), and compared to those with typical cognitive levels ($d = .32$). Bowers later concludes, "*there are no short or long term benefits for the majority of struggling readers above grade 1.*"

In making sense of this critique, the first thing to note is that a non-zero effect size was still evident even in students with demonstrated multiple risk factors and who struggle with any learning. The effect size of $d = .15$ 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. One also has to contextualize this effect to make sense of it. This effect size may be large for these children in relative terms if they made modest or no progress before then and sets the bar for contrasts with alternative pedagogies for this distinct multi-need group. A key issue in evaluating any intervention against a counterfactual is: What might the expected progress be for this sample? Even without such considerations, an effect size of $d = .15$ places about 6% more of the intervention group clearly above the control group mean at post-test. Thus, even a small effect size of .15 can still be very practically important if played out at national scale, as Bowers himself acknowledges elsewhere in his article. The authors of the NRP noted (Part II, p. 117) that among a range of plausible explanations, many of the studies contributing to this effect involved regular whole class teaching. This delivery method may simply have not been sufficient to meet the literacy needs of this multiply at-risk group (Connor & Morrison, 2016). Finally, it is also important to recognize that later reviews (including meta-analysis of responsiveness to intervention studies) show that IQ is not a strong predictor of the responsiveness of poor readers to reading intervention, including experimental studies that explicitly controlled for IQ (e.g., Morris et al. 2012) and in a meta-analysis (Stuebing et al. 2009). As is expected in all professions (e.g., medicine), we cannot minimize visible, consistent effects of any size in education, particularly for those students with multiple learning needs.

2) The NRP analysis showed that the effects of explicit phonics are larger in grade 1 than after grade 1, where programs focused on phonics are often targeted for older struggling readers. We start by noting that the NRP report executive summary does, as Bowers observes, exaggerate the more cautious interpretations found in the chapter on phonics. The authors also acknowledged that the NRP report could not say anything strongly about the effect of grade *per se* on intervention because there were insufficient number of articles across each of the

appropriate grade levels. Here Bowers uses the NRP acknowledgement of not being able to differentiate by grade level as a criticism of the NRP findings, when failing to interpret grade level effects because of a limited number of studies is the only empirically defensible approach. Phonics is an approach to teaching beginning readers and children who struggle to access sublexical components of words by learning the alphabetic principle, i.e., the relation of the shared phonemic structure of oral and written language. Once a child has mastered the alphabetic principle, what rationale is there to offer ever more phonics to capable readers in the later grades: Age and ability will *always* be confounded except in very rare cases.

3) Bowers points to a reduced effect size of $d = .27$ on word reading in the studies available of the effects of phonics between 4 months and a year after the interventions closed. He also notes that the authors of the NRP did not assess whether the long-term benefits extended to spelling, reading texts, or reading comprehension. In contextualizing these comments, meta-analyses always reflect the available independent and dependent variables in the underlying studies. Perhaps the most important point is that the long-term effect sizes are positive and practically significant on the primary outcome – which is not a common finding in studies that conduct follow-up assessments. Rather than criticize the long-term effects, most researchers would find this long-term effect noteworthy. We would thus agree with Bowers only in the specific sense that more research is needed on a range of longer-term outcome measures potentially affected by phonics instruction.

4) Bowers' fourth point is a cluster of criticisms we consider together. Bowers concludes, "*the evidence that systematic phonics is more effective than whole language is weaker still.*" In his analysis Bowers appears to be looking at the subset of 12 studies from the NRP that compared systematic phonics to so-called whole language approaches when he reports a $d = 0.31$

effect size. Again, we note the positive effect is still evident. Bowers also correctly notes that there is very modest evidence for synthetic phonics versus alternative methods. However, we would also note that his comment here appears less to be aimed at the authors of the NRP (who explicitly acknowledge the fragility of the evidence base available on this specific point), but at the actions of some consumers of it in some UK policy circles.

Camilli et al. (2006); Hammill and Swanson (2006). Bowers re-describes the views from Camilli et al. (2006) about the contrasts within the NRP and the possible role of other features of pedagogy such as wider language and tutoring in driving reported effects. It is not clear to us that Bowers adds any new substantive points to the discussion. Bowers uses Camilli et al. (2006) and Hammill and Swanson (2006) to rebut the NRP conclusions about phonics (though it is here in the context of the Hammill and Swanson (2006) study that he correctly notes that small effect sizes can be practically important if played out across whole populations). Bowers cites Stuebing et al. (2008) as a replication of Camilli et al. (2006), but does not report that Stuebing et al. raised questions about the Camilli et al. (2006) analysis of their recoded NRP database and that the purpose of the paper was to empirically defend and contextualize the NRP report – not to agree with Camilli et al. Stuebing et al. concluded that the effect size for systematic and unsystematic phonics ($d = .123$) could not be compared to the NRP $d = .41$ for systematic phonics versus unsystematic/no phonics instruction because of Camilli's use of moderator variables. As Bowers noted, Stuebing interpreted comparisons of no phonics, unsystematic phonics, and systematic phonics as a dosage effect, supporting this conclusion in their Table 2 where the effects of systematic phonics ($d=.49$) is larger than the effect of some phonics ($d=.31$) when the moderators coded by Camilli et al. are excluded from the comparisons.

Stuebing et al. (2008) showed that the Camilli et al. (2006) dataset of NRP findings actually yielded effect sizes similar to the NRP depending on the question that was asked of the data. In the re-analysis, the effect size for reading interventions that included systematic phonics interventions versus programs that include unsystematic or no phonics was $d = .39$, close to the NRP report ($d = .41$). Although Camilli et al. (2008) raised questions about the re-analyses, any interpretation depends on how the NRP studies are recoded and whether moderators are used. Stuebing et al. (2008, p. 131) concluded,

The comparisons by Camilli et al. (2006) ask questions that are different from the primary question asked by the NRP, but the results of the two sets of analyses can be reconstructed to yield comparable effect sizes for the effects of systematic phonics versus either unsystematic phonics or no phonics controls when the same study parameters are estimated.

Importantly, Stuebing et al. concurred with Camilli et al. (2006) on the value-added effects of additional literacy-related activities as well as tutoring. The largest effect sizes in Camilli et al. (2006) recoding of the NRP database occurred when systematic phonics is combined with additional language and literacy activities and delivered through tutoring ($d = 1.35$, see Table 2 of Stuebing et al. 2008). Again, none of this shows that phonics does not in and of itself make a distinct and important contribution to early literacy. Nor does it suggest that students only need phonics instruction. To reiterate, Stuebing et al. showed in their Table 2 (line 14) that if one only considers the relatively pure cases of interventions involving what the NRP defined as systematic phonics and compares these against conditions where Camilli et al. had coded the absence of both tutoring and wider language activities (85 contrasts in 17 studies), the effect size is $d = .49$.

In response to Hammill and Swanson (2006), Stuebing et al. (2008) also emphasized the contextual nature of effect sizes, showing that even smaller effect sizes were practically significant and could improve reading outcomes for many children depending on the base rate of reading failure, a point on which Bowers seems to agree. 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. We do not understand the motivation for discounting the consistently positive effects in favor of “systematic” phonics approaches. Arguing about whether the effects are small, medium, or large is not the relevant issue when making educational decisions about whether some level of explicit phonics instruction is beneficial to learning to read. . Encouraging educators to discount positive effects of explicit phonics instruction is simply *not* helpful, but is potentially harmful to many children struggling to access appropriate reading instruction (Seidenberg et al., 2020).

Reflecting their wider finding of the combinatorial effect of phonics with language and tutoring, Stuebing et al. (2008, p. 133) thus concluded,

Although it seems difficult to move beyond the historic dichotomy of reading instructional approaches, it is time to embrace comprehensive approaches to reading instruction and work toward determining how to integrate different components of reading instruction into classroom practice so that the diversity of students and their individual needs can be addressed.

As the NRP (NICHD 2000, p. 2-97) stated, “*Phonics instruction is never a total reading program,*” and it “*should be integrated with other reading instruction.*” The monolithic view of phonics versus meaning-based instruction current when the NRP report was published is not a contemporary platform for discussing reading instruction.

Torgerson et al (2006). Torgerson et al. considered only RCTs on phonics and explicitly considered a range of other indicators of methodological quality in the phonics literature. Bowers’ primary concern is his issue with the contrasts of systematic versus a combined unsystematic or no phonics control, which we have discussed. Beyond this, he simply draws attention only to the limitations in the research base identified by the authors themselves (possible effects of outlier study removal, publication bias, lack of detail on randomization processes for example). 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). The authors also observed that the effect sizes from the selected RCTs were generally small and that the evidence quality for all phonics related questions was “moderate” (e.g., for the overall effects of phonics on typical and atypical learner’s word reading) or “weak” (e.g., the effects of analytic or synthetic phonics on outcomes). It is not clear what Bowers adds in the absence of these nuances.

McArthur et al. (2012). These authors sought to explore the effects of phonics on outcomes for struggling readers. In particular, they sought RCTs and matched control studies that allowed an analysis of the unique effects of phonics independent of content often taught with phonics such as sight words. Bowers criticizes this review based on his observation that the

overall effects reported may have been driven by the impact of two atypical interventions (Levy & Lysynchuk 1997; Levy et al. 1999). There are several points to be noted. First, Bowers reports from the 2012 McArthur et al. review rather than the most recent McArthur et al. (2019) updated review that adds further studies to the analysis from the most recent 6 years of studies. The two papers by Levy and colleagues are included among the 14 studies. The rubric of the McArthur et al. (2012) review makes it clear that reading was assessed through “*various outcomes.*” The outcomes used in the Levy studies included measures of both novel word and pseudoword reading, so they do provide measures of generalizable phonic skills. Indeed, they assess exactly the mechanism of transfer expected from an analytic phonics approach. The Levy and colleagues studies are certainly not alone amongst 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.

All good meta-analyses statistically test for the homogeneity of effects within the included studies. McArthur et al. (2012) show that there was statistically significant heterogeneity in the reported studies. However, McArthur et al. also noted that this heterogeneity could equally have been due to another study with an untypically *low* score on primary outcomes. The study could possibly have been, but was not, excluded. We note that the purpose of meta-analysis as an empirical synthesis is potentially undermined if studies are included or excluded post hoc. As a further insurance against effects of heterogeneity, random effects and fixed effects models were both run within the meta-analysis and both gave very similar results, suggesting the findings are robust.

Finally, given the weight Bowers (2020) attaches to the influence of studies by Levy and colleagues, it is important to understand that the overall effect sizes for outcomes involving word

and pseudoword reading accuracy were not the largest effects reported. McArthur et al. (2012) concluded the evidence base on these two specific outcomes across all studies was weak. Because the outcomes used by Levy and colleagues were specific to word and pseudoword accuracy, these two studies did not feature *at all* in the five other outcome analyses reported by McArthur et al. (2012): exception word reading, where the largest effect size of the entire study was .84 standard deviation (SD) for phonics treatment over controls; regular and exception word reading fluency (the second and third largest effect sizes), spelling; and reading comprehension. The strongest of the modest claims made by McArthur et al. (2012) are thus of the “moderate-quality” evidence of impact of phonics on exception word reading accuracy and word and pseudoword reading fluency. The Levy studies cannot explain these effects.

Bowers (2020) goes on to state that his most important point here is that the McArthur meta-analysis compared systematic phonics to no extra training at all, or to training on nonreading tasks, such that it is not appropriate to attribute any benefits to systematic phonics. He argues that any form of extra instruction may have “mediated” the gains observed. In making sense of this analysis, it might first be noted that Bowers appears to have moved from *his* purported original question about systematics versus less systematic or no phonics instruction to a more general critique of all aspects of phonics interventions. We should also note that the stated purpose of the McArthur et al. (2012) review was not to compare phonics against other interventions but to establish the role of phonics as it stands alone, and independent of other possible intervention teaching. Levy and colleagues featured tutoring in both treatment and control conditions and thus allow an analysis of the unique effect of phonics. Of course, the interventions did not occur in a vacuum in that all of the children were in school receiving business as usual teaching. In general, business as usual comparisons are not as intensive or

structured as the interventions in most research studies and is heterogeneous across schools and classrooms, making it hard to code, but controlled in a randomized trial.

Finally, just doing *something* in an RCT rarely works in and of itself. Lortie-Forgues and Inglis (2019) recently reviewed 144 principled, well-executed, large-scale RCT trials funded by the Education Endowment Foundation in England and the National Center for Educational Evaluation and Regional Assistance in the United States. An average effect size of 0.06 SD was found across all trials. The trials involved a much broader performance range of children, not the lowest performing tail of the distribution addressed by McArthur et al. (2012). The *smallest* effect reported in McArthur on reading comprehension is 0.28 SD, and all six other effects reported are above 0.45 of an SD. The critical issue here is that effect sizes such as 0.28 SD for comprehension are not common in well-conducted interventions and they are meaningful.

Galuschka (2014). Galuschka et al. explored the widest base of evidence of RCT interventions of any kind that impact reading in poor readers. Bowers (2020) contention with this study is that the reported effect size for phonics interventions ($g = 0.32$) was no bigger than for other intervention components such as reading fluency training ($g = 0.30$), auditory training ($g = 0.39$), and colour overlays ($g = 0.32$). Bowers summarizes, “*The reason significant results were only obtained for phonics is that there were many more phonics interventions.*” 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. This security of the findings for phonics across multiple trials is the reason why it is the only treatment that is statistically significant for reading

outcomes. Bowers also notes that the estimated effect size for phonics adjusting for possible publication bias ($g = 0.198$) was smaller than the overall estimate. Again, this effect is both positive in favour of phonics intervention (as coded) and remains statistically significant and practically important. Finally, Bowers repeats his earlier views, criticizing Galuschka et al. for not exploring direct contrasts of interventions as rivals and of not exploring contrasts of systematic versus unsystematic phonics. This critique is made despite the difficulties of such comparisons given the limited number of trials of other approaches beyond phonics that Bowers has already noted. Further, neither of these goals was the purpose of the Galuschka et al. review.

In the spirit of the Stuebing et al. (2008) analysis discussed above and based on well-established theory (e.g., the Simple View of Reading; Gough & Tunmer 1986), it seems likely that phonics interventions will both have a discernible unique effect and interact with fluency and comprehension interventions to support reading comprehension. Again, it is troubling to consider why Bowers discounts the consistently positive effects in favour of explicit phonics-type approaches. When treatments in any domain (medicine, counselling, and education) are associated with consistently significant effects, it is wise for professionals in those fields to consider ways to integrate these treatments into their protocols. We are not suggesting that asking empirical questions and further investigating these treatments is unwarranted. However, empirical investigations and naysaying are two completely different pathways. The question that Bowers would be required to answer is whether the bulk of the evidence on the impact of systematic or explicit phonics instruction (not first or solely) is associated with greater impact on student's learning to read. According to the evidence the answer is yes.

Suggate (2010, 2016). Bowers (2020) considers two reviews by Suggate. The 2010 study explored the effects of student grade severity of impairment and intervention modality on

outcomes for at-risk and disabled readers. Suggate (2010) provides evidence of an interaction between modality of intervention and student grade of students, concluding that there is (a) an appreciable effect of phonics in the early elementary years only and (b) that a comprehension focus yields large effects in later elementary years. Bowers contests even this finding about the potential limits of phonics by arguing that the effect size difference between phonics and other approaches is modest in size in kindergarten and grade 1. Bowers focuses on a minority of studies (10%) as being carried out in non-English language contexts as problematic, emphasizes “near-significant” contrasts of English ($d = .48$) and non-English language learning context outcomes ($d = .61$), and identifies a single study amongst these with potential outlier effects.

Contextualising this analysis again, it is important to look at the effects of phonics across languages to avoid pervasive Anglocentrism in our theorizing (Share, 2008). The differences across languages (Suggate, 2010) are not great in magnitude. Indeed, an “outlier,” if accepted as such, further reduces this difference (note there were 85 studies in this meta-analysis). Bowers draws attention to the modest differences between preliminary effect sizes for overall contrasts of comprehension and phonics in Suggate’s Table 1. Bowers does not mention that the absolute effect of phonics or comprehension interventions was mediated by an interaction – the final model in Suggate’s Table 2 and Figure 1, with no significant main effect of phonics. While emphasizing the early role of phonics, Suggate (2010, p. 1596) is careful to conclude that: *“This does not suggest that other interventions, such as language interventions or dialogic book reading, are superseded in importance by phonics interventions in kindergarten or earlier.”* There is no obvious reason to set up the “alternative methods” as rivals.

Suggate (2016) examined longer-term outcomes of the impact of 71 phonics and other reading interventions. Overall, the analysis revealed an overall short-term effect ($d = 0.37$) that

decreased in a follow-up test around a year later ($d = 0.22$) with phonics producing the weakest overall effects (phonics, $d = 0.07$; fluency, $d = 0.28$; comprehension $d = 0.46$; and phonemic awareness, $d = 0.36$). Suggate also noted design features and treatment dosage affect outcomes.

A number of important points need to be born in mind here. First, the absolute number of contrasts, and participants available in each case varied widely. For phonics, compare the $n = 4045$ and $k = 22$ contrasts with fluency, which has an $n = 736$ and $k = 2$ contrasts. 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. The *unadjusted* observed effect for phonics at the just under one-year follow-up from Suggate’s Table 3 is $d = 0.30$ on word reading ($k = 20$ studies, $n = 3895$ students) and $d = 0.25$ over all measures ($k = 22$ studies, $n = 4045$ students). Ultimately, we simply have much more data on phonics interventions and word reading outcomes at delayed post-test in the studies within the meta-analysis than we have on pretty much anything else in reading intervention research: It is a highly replicated pattern.

Finally, Suggate (2016) excluded both designs with nested analyses and very long-term follow up studies from his review, the latter on the somewhat surprising grounds that they were unrepresentative. However, such studies are highly informative. Suggate specifically excluded Blachman et al. (2014) as an outlier because it was a 10-year follow up of their RCT intervention trial (Blachman et al. 2004) for struggling 2nd and 3rd grade readers. The children had received a 1:1 tutoring intervention for 50 minutes daily each day for 8 months, focusing on unscripted, organized lesson plans for phonics, phoneme blending, fluency, oral reading practice, and spelling. In comparison with a business as usual condition, the intervention yielded large effect sizes at immediate and 1-year post-test on measures of reading accuracy, speed and comprehension. Fully 10 years later, Blachman et al. (2014) located 33 treatment and 25

comparison participants from the original sample of 37 treatment and 32 comparison children. Blachman et al. found that of 12 measures of reading the intervention children significantly outperformed the controls on two: Word identification ($d = .53$) and a basic skills cluster scores made up of the word identification and a word attack measure ($d = .62$). Across seven reading and spelling measures, the mean effect size was $d = .24$, which is positive and practically significant. We know of no other paper reporting long-term effects of an RCT in English (although see Kjeldsen et al. (2014), a kindergarten intervention to Grade 9 in Swedish learners).

It is not strictly necessary to prove such long-term effects to show that phonics is useful. A medical metaphor has sometimes been used here (e.g. Coyne et al. 2004). Do we construe phonics (or indeed, any intervention) as akin to “insulin” or “inoculation” models? We would not say insulin ‘does not work’ for diabetic patients. For *some* children though, phonics might be akin to inoculation or at least sustained ‘symptom’ improvement. While we have limited evidence of long-term effects beyond fairly good evidence at the one-year post-test, there is no need to accept that somehow phonics is not efficacious.

Other meta-analyses and a systematic review of meta-analyses. Bowers (2020) clusters five other studies, so we will consider these as he does, as a group, excluding Hammill and Swanson (2006), which we discussed above. Bowers reports the overall effect sizes from two theses (Han 2010; Sherman 2007), and a published paper (Adesope et al. 2011). Sherman reports an effect size of $d = 0.53$ for word reading in older struggling readers. Han and Adesope et al. both looked at phonics among non-English speakers. In both cases these analyses showed positive effects of phonics ($d = 0.41$ and $d = 0.40$ respectively), but Bowers notes that there were other interventions such as structured writing interventions with slightly larger overall effects ($d = 0.48$). We are not sure what the point is here, unless it is to suggest that other practices such

as structured writing are also associated with improved outcomes. We fully agree that there is no evidence to suggest “phonics-only” is optimal. We recognize that other practices such as “structure writing” might well be integrated within a reading approach to yield significant reading and writing outcomes. We do not think the issue is which one approach should be used but rather how do we integrate and leverage evidence-based practices (e.g., phonics, fluency) to meet the range of learning needs in the classroom.

In Adesope et al. (2011) the “phonics” interventions included guided reading so are not a pure comparison. As in his consideration of some previous reviews, Bowers (2020) does not report that there are many more studies in the case of phonics ($k = 14$ studies, $n = 1647$ students) versus all other interventions ($k = 2 - 5$, largest $n = 648$). Unlike Bowers, Adesope et al. (2011, p. 648) do not place these different interventions in opposition to each other, concluding quite reasonably that *“The evidence presented in this meta-analysis suggests that cooperative reading, systematic phonics instruction, and diary and structured writing interventions have the potential to enhance the teaching of English literacy to ESL immigrant students.”* Finally, Adesope et al. also wisely conclude that contextual factors such as school contexts as well as minority language learner characteristics may influence the effectiveness of any of these strategies. Again, Bowers ignores this entire important authorial nuance.

The final section of Bowers (2020) is devoted to a critique of Torgerson et al. (2018), a tertiary review that concludes cautiously that phonics has received support, but that more work is still needed to be certain about its effects. Here Bowers repeats his arguments about the absolute size of non-phonetic effects, the inclusion of non-English studies, and the purported effects of studies such as those of Levy et al. (1999) on outcomes that we have already considered.

Discussion

Interpretation of Empirical Research

Measured in conventional terms, there is consistent evidence of positive effects for explicit phonics-based intervention on reading outcomes. These positive effects are persuasive because of the large sample size across studies and the range of investigators, settings, and participants. Bowers (2020) perceives that this effect has been exaggerated and we agree there are examples of this being the case. We are also aware of professionals who deny that phonics has a role in improving learning to read (Smith 2004). While we appreciate many of the significant points that Bowers makes, we are aware of the potential dangers of denying the impact of phonics as well as the dangers of overemphasizing its effect. *We do not agree that the impact of systematic or explicit phonics instruction is modest and unimportant.* This advantage is consistently reported, spanning the 14 meta-analyses reviewed above executed across many countries worldwide over at least eight decades (e.g., Scammacca et al. 2016). Effects are reliably still present up to, on average, one year after the interventions have ended (Suggate, 2016). We even have *some* evidence of the very long-term impacts of reading programs that include phonics on word reading up to 10 years after interventions in children with word-level reading disabilities (Blachman et al. 2014). The effect of phonics is evident for typical and atypical readers and across alphabetic languages in the 14 meta-analyses; for second language learners (Vaughn et al. 2006) and for immigrant children (Adesope et al. 2011), and across RCT trials and other controlled studies. There is evidence that instruction that includes explicit phonics affects fluency and comprehension, even in the most struggling readers, but the latter are more inconsistent (Morris et al., 2012). At a certain level, however, we must ask how comprehension proceeds in struggling readers if they cannot access the print. Table 1 shows a

modest effect of PHAB on reading comprehension, which is even larger for comparisons involving RAVE-O. In Torgesen et al. (2001), one of the interventions was an intense 80-hour sublexical tutoring program with children severely impaired in word reading. It was associated with major gains in decoding and text-based comprehension, but little gains in fluency because these upper elementary children had little access to print before the intervention because of their severe decoding problems.

While sometimes presented otherwise by Bowers (2020), the reality of the research base is a serious effort from international scientists who, mindful of the strengths and limitations of the outcomes evident in their own data, cautiously explain and contextualize their findings. At best, they know that effect sizes are *estimates* of effects. Despite the 14 meta-analytic studies supporting explicit phonics, like many of the other researchers involved in this work, we recognize that the relative effects of phonics require conditioned application including such terms such as “may” and “probably” to any causal claims about the role of phonics. On the other hand, when Bowers suggests in the concluding paragraph of his review that the effects of systematic phonics instruction versus alternative reading methods including whole language are a “draw,” we think this conclusion is tantamount to acceptance of the null hypothesis and is not helpful to educators or their students. Not only is this statement not supported by the evidence from which Bowers claims to derive his judgements, it unnecessarily arouses controversy in a field that needs to focus on the best practices available.

Areas of Agreement

Throughout the paper we have highlighted areas of agreement with Bowers (2002). We agree that design and content aspects of many studies is inconsistent and hard to code in meta-analysis. This is an issue across intervention studies and not specific to phonics interventions.

We agree that publication bias may be evident: Torgerson et al. (2018) show that of 12 meta-analyses of phonics, six explore publication bias and three report evidence consistent with there being some bias (see also Galuschka et al. 2020). While we agree that there remain some significant weaknesses in the empirical data, we note some of these issues span the social and natural sciences broadly – not specifically to this topic. These issues also apply to candidate “alternative treatments.” Furthermore, the bar is set high here for *any* credible intervention: RCTs, with clear contrast with alternatives, clear (replicable) contrasts of content, clear impacts on comprehension, no publication bias (we assume to this end, widespread trial registration and CONSORT-quality reporting of all trials), then later, clear evidence from meta-analysis of all such RCTs, possibly even 10 year follow ups! We eagerly await such platinum standard data more generally in reading intervention: There is much work to be done! Finally, we also suspect we would share common cause with Bowers against overly strong uses of evidence by some advocates of phonics and some policy makers, and wherever this leads to denuded and impoverished curricula. We prefer a focus on “explicit” instruction and agree with Bowers that word work does not necessarily require a particular scope and sequence or that any particular method for delivering phonics is strongly supported, although we think effective phonics instruction is facilitated by an organized lesson plan that is intentional and not incidental.

What Are the Right Questions?

What we do not have is a sufficiently strong science of *teaching* children to read (Seidenberg 2017). Although it would be nice to be able to break down the teaching of reading into a very precise package outlining the role of each component and the best way to teach it, unpacking these components is difficult to do and would potentially yield small effects, as in the PHAB condition in Morris et al. (2012). In the teaching world, there are many factors affecting

children and learning to read. What is clear is that the questions posed by Bowers (2020), as well as extreme proponents of phonics, is an artificial one that does not support progress in the science of teaching children to read. In the US and UK, legislation is passed mandating certain approaches to reading instruction, and even worse, prescribing specific commercial programs for children identified with dyslexia (Petscher et al. 2019). However, we would do well to ask the right questions about how to teach effectively the range of learners in our classrooms to read. This requires embracing the complexity and multiple competency view of reading. Even in the area of phonics, there are many ways to help children access sublexical components of words and one approach does not work with all children, even those identified with dyslexia. We also recognize that while phonics instruction is associated with improved reading outcomes, the amount of phonics instruction required, the level of explicitness, and how phonics instruction is integrated into other important elements of reading instruction requires further investigation.

We agree with Bowers (2020) that alternative approaches to teaching reading need to be studied and suggest that the current research base will be enhanced by comprehensive approaches that integrate- not balance- different components of reading instruction with ample consideration of the range of learners and their instructional needs. For many children, these components need to be taught explicitly. In particular, children who have trouble accessing words because of problems with phonological processing need explicit code-based instruction. They also need to practice to build automaticity. Children learning to read English will need to learn how to navigate both the morphemic and phonemic aspect of the spelling system of English. Comprehension instruction recognizing the critical roles of word meaning and world knowledge also requires a more explicit approach because many children do not develop vocabulary, background knowledge, and comprehension strategies through passive reading.

Focusing on the integration of these activities in a comprehensive approach to reading instruction would be beneficial. Reading instruction should not occur in the absence of opportunities to read and write and oral language development. These opportunities are usually present in reading instruction, making it hard to isolate the effects of systematic phonics instruction. However, these successful integrated approaches rely on facilitating students' access to word reading and meaning through effective instructional practices that demonstrate the ways in which phonemes map to print in regular and irregular ways providing many opportunities to read words so that the structure of language is acquired both explicitly and implicitly. How to integrate comprehensive programs and use them to differentiate and customize instruction for individual children is a much better question. Programs should package components with known efficacy and evaluate intervention outcomes when the package permits differentiation (e.g., Connor & Morrison, 2016).

The dichotomy of systematic phonics versus less systematic/no phonics instruction is artificial because these activities exist on a continuum, representing an outmoded question that is hard to address from the available studies (Stuebing et al. 2008). The issue is how to consider the reading development of the learner and to integrate these components into a comprehensive reading program that permits differentiation for the individual learner. As Seidenberg (2017) pointed out, many children come to school primed to learn to read. However, because of environmental factors as well as biological factors that make it harder for the brain to mediate reading, many children struggle to learn to decode and therefore are less able to access print. Much of what Bowers (2020) calls exaggeration is a reaction to the need of these children for explicit phonics instruction. Many children do not get the word work they need, partly because it is not intentional, explicit, and well organized.

In the spirit of seeking clarity and maximal consensus, we map out seven major claims made by Bowers underpinning his original review and the roughly equal number of points of agreement and disagreement in Table 3 along with relevant evidence. Table 3 acts only as a summary overview of the multiple issues discussed here to aid a reader in mapping the broad arguments, and should only be interpreted in light of the nuances we elaborate in the main text throughout this article. We finish with three claims in points 8-10 that are not explicitly made by Bowers (2020) but which might potentially serve as unifying statements and underpin the most productive future work on reading interventions to meet the needs of all learners.

The 21st century challenge is to meet the distinct needs of students who are “primed” for reading as well as those requiring much more explicit access to the alphabetic principle, while also engaging all children with the complexity of their language and their literature. To this end, a suitably nuanced and evidenced science of teaching reading is a work in progress. This ambitious enterprise may involve healthy friendly professional disagreement, but it will also need a mindset among all research leaders that acknowledges this complexity over old binary modes of the 20th century, the importance of this goal to the wide community, and also the importance of communicating it accurately and effectively to all of the users of our science. To stimulate this type of discussion was the goal of this paper.

References

- 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.
- Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of Psycholinguistic Research, 39*, 141-163.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology, 96*(3), 444.
- 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*, 46–57.
- Bowers, J. S. (2020). Reconsidering the evidence that systematic [phonics](#) is more effective than alternative methods of reading instruction. *Educational Psychology Review*.
<https://link.springer.com/article/10.1007/s10648-019-09515-y>
- 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.
- Camilli, G., Kim, S. H., & Vargas, S. (2008). A response to Stuebing et al., " Effects of systematic phonics instruction are practically significant": The origin of the National Reading Panel. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, 16*, 1-20.
- Camilli, G., Wolfe, M., & Smith, M. L. (2006). Meta-analysis and reading policy: perspectives

- on teaching children to read. *The Elementary School Journal*, 107(1), 27–36.
- Chen, V. & Savage, R. S. (2014). Evidence for a simplicity principle: Teaching common complex grapheme-phonemes improves reading and motivation in at-risk readers. *Journal of Research in Reading*, 37, 196–214,
- Connor, C. M., & Morrison, F. J. (2016). Individualizing student instruction in reading: Implications for policy and practice. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 54-61.
- Coyne, M.D.L., Kame'enui, E.J., Simmons, D.C., & Harn, B.A. (2004). Beginning reading intervention as inoculation or insulin: first-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, 37, 90-104.
- 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.
- Engelmann, S., & Bruner, E. C. (1988). *Reading Mastery I/II Fast Cycle: Teacher's guide*. Chicago: Science Research Associates.
- Fountas, I. C., & Pinnell, G. S. (2012–13). Guided reading: The romance and the reality. *Reading Teacher*, 66, 268–284.
- 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),
- 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, DOI: 10.1080/00461520.2019.1659794
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial & Special Education*, 7(1), 6–10.
- 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, 17–26.
- 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>.
- Hatcher, P. J., Hulme, C., & Ellis, A. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The Phonological Linkage Hypothesis. *Child Development*, 65, 41–57.
- Her Majesty’s Inspectorate (HMI). (1990). *The teaching and learning of reading in primary schools*. London: Department of Education and Science (DES).
- Kjeldsen, A.-C., Kärnä, A., Niemi, P., Olofsson, A., & Witting, K. (2014). Gains from training in phonological awareness in kindergarten predict reading comprehension in grade 9. *Scientific Studies of Reading*, 18, 452-467. DOI:10.1080/10888438.2014.940080.
- Levy, B., & Lysynchuk, L. (1997). Beginning word recognition: benefits of training by segmentation and whole word methods. *Scientific Studies of Reading*, 1, 359–387.
- Levy, B., Bourassa, D., & Horn, C. (1999). Fast and slow namers: Benefits of segmentation and whole word training. *Journal of Experimental Child Psychology*, 73(2), 115–138.
- Liberman, A.L. (1996). *Speech: A special code*. MIT Press.

- Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000). Putting struggling readers on the PHAST track: A program to integrate phonological and strategy-based remedial reading instruction and maximize outcomes. *Journal of Learning Disabilities, 33*(5), 458–476.
- Lortie-Forgues, H., & Inglis, M. (2019). Most rigorous large-scale educational RCTs are uninformative: Should we be concerned? *Educational Researcher, 48* (3), 158–166.
- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2008). Children’s printed word database: Continuities and changes over time in children’s early reading vocabulary. *British Journal of Psychology, 101*, 221–242.
- Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). An evaluation of two reading interventions derived from diverse models. *Reading Research Quarterly, 40*, 148–183.
- McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., Larsen, L., Marinus, M., Wang, E., & Castles, A. (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., Banales, E., Anandakumar, T., Marinus, E. & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*, (11).
- Moats, L. (2019). *Structured literacy: Effective instruction for students with dyslexia and related reading difficulties*. <https://dyslexiaida.org/structured-literacy-effective-instruction-for-students-with-dyslexia-and-related-reading-difficulties/>
- 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*, 99–127.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction*. (NIH Publication No. 00–4769). U.S. Government Printing Office.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.
- Petscher, Y., Fien, H., Stanley, C., Gearin, B., Gaab, N., Fletcher, J.M., & Johnson, E. (2019). *Screening for dyslexia*. Washington, DC: Office of Special Education Programs, National Center on Improving Literacy. Retrieved from improvingliteracy.org.
- Savage, R.S. (2020). The Simple View of Reading: A scientific framework for effective teaching. *The Reading League Journal, 1*, 41-45.
- Savage, R.S., Georgiou, G., A. Parrila, R., & Maiorino, K. (2018). Preventative reading interventions teaching direct mapping of graphemes in texts and set-for-variability aid at-risk learners. *Scientific Studies of Reading, 22*, 225-247.
- Scammacca, N., Roberts, G. J., Cho, E., Williams, K., Roberts, G., Vaughn, S., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4-12, 1914-2014. *Review of Educational Research, 86*, 756-800.
- Seidenberg, M. (2017). *Language at the speed of sight: How we read, why so many cannot, and what can be done about it*. Basic Books.
- Seidenberg, M. S., Cooper Borkenhagen, M., & Kearns, D. M. (2020). Lost in translation?

- Challenges in connecting reading science and educational practice. *Reading Research Quarterly*, 55, S119-S130.
- Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole class teaching. *British Journal of Educational Psychology*, 78, 597–620.
- Share, D. L. (1995) Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55, 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, 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.
- Smith, F. (2004). *Understanding reading*. (6th Ed.). Erlbaum.
- Snow, C. E., & Juel, C. (2005). Teaching children to read: What do we know about how to do it? In M. J. Snowling & C. Hulme (Eds.), *Science of reading: A handbook* (pp. 501–520). Blackwell.
- Stockard, J., & Wood, T. W. (2017). The threshold and inclusive approaches to determining “best available evidence” An empirical analysis. *American Journal of Evaluation*, 38, 471-492.
- Student Achievement Partners (2020). *Comparing reading research to program design: An examination of Teachers College Units of Study*.
<https://achievethecore.org/page/3240/comparing-reading-research-to-program-design-an-examination-of-teachers-college-units-of-study>
- Stuebing, K. K., Barth, A., Cirino, P., Francis, D., & 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.
- Stuebing, K.K., Barth, A.E., Molfese, P.J., Weiss, B., & Fletcher, J.M. (2009). IQ is *not* strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children*, *76*, 31-51.
- Suggate, S. P. (2010). Why what we teach depends on when: grade and reading intervention modality moderate effect size. *Developmental Psychology*, *46* (6), 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* (1), 77–96.
- 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*, *34*, 1-31.
- 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.
- Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33–58.
- Tse, L., & Nicholson, T. (2014). The effect of phonics-enhanced big book reading on the language and literacy skills of 6-yearold pupils of different reading ability attending lower SES schools. *Frontiers in Psychology*, *5*, Article 1222.

- Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., ... & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*(3), 449-487.
- Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, Automaticity, Vocabulary Elaboration, Orthography (RAVE-O): A comprehensive, fluency-based reading intervention program. *Journal of Learning Disabilities*, *33*(4), 375–386.
- Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*(3), 211–239.
- Yeung, S., & Savage, R.S. (2020). Teaching grapheme-phoneme correspondances using a Direct Mapping approach for at risk second language learners: A randomized control trial. *Journal of Learning Disabilities*, *53*, 131-144.

Table 1

Effect Sizes And Confidence Intervals for Comparisons of Interventions for Poor Readers that Taught Systematic Phonics and Classroom Survival Skills (PHAB + CSS) Versus no Reading Instruction and Classroom Survival Skills (MATH + CSS)

Measure	PHAB + CSS (<i>n</i> = 69)		MATH + CSS (<i>n</i> = 68)		<i>g</i>	Variance of <i>g</i>	SE	95% CI (lower)	95% CI (upper)	<i>z</i>
	Mean	SD	Mean	SD						
WRMT-R Word Attack	7.49	5.36	5.77	5.47	0.32	0.03	0.17	-0.02	0.65	1.85
WRMT-Word Identification	32.21	13.24	29.15	14.73	0.22	0.03	0.17	-0.12	0.55	1.28
TOWRE Word Reading Efficiency	20.72	12.44	19.18	11.95	0.13	0.03	0.17	-0.21	0.46	0.74
WRMT-R Passage Comprehension	15.87	8.44	14.71	8.85	0.13	0.03	0.17	-0.20	0.47	0.78
WRMT-R Spelling	19.10	2.85	18.37	3.13	0.24	0.03	0.17	-0.09	0.58	1.42
GORT-3 Accuracy	2.54	3.94	2.31	3.50	0.06	0.03	0.17	-0.27	0.39	0.36
GORT-3 Fluency	1.33	2.01	0.91	1.37	0.24	0.03	0.17	-0.09	0.58	1.42
GORT-3 Comprehension	9.66	6.38	8.04	6.44	0.25	0.03	0.17	-0.08	0.59	1.47

Note. WRMT-R = Woodcock Reading Mastery test- Revised; TOWRE = Test of Word Reading Efficiency; GORT-3 = Gray Oral Reading Test (3rd Ed.)

Table 2

Effect Sizes and Confidence Intervals for Comparisons of Interventions for Poor Readers that Taught Multiple Reading Strategies and Systematic Phonics (PHAST+PHAB) Versus Systematic Phonics and Classroom Survival Skills (PHAB+ CSS)

Measure	PHAST (n = 73)		PHAB+CSS (n = 69)		g	Variance of g	SE	95% CI (lower)	95% CI (upper)	z
	Mean	SD	Mean	SD						
WRMT-R Word Attack	11.68	6.67	7.49	5.36	0.69	0.03	0.17	0.35	1.02	3.99
WRMT-Word Identification	38.51	12.41	32.21	13.24	0.49	0.03	0.17	0.16	0.82	2.88
TOWRE Word Reading Fluency	25.80	12.31	20.72	12.44	0.41	0.03	0.17	0.08	0.74	2.42
WRMT-R Passage Comprehension	19.68	8.12	15.87	8.44	0.46	0.03	0.17	0.13	0.79	2.71
WRMT-R Spelling	20.33	2.64	19.1	2.85	0.45	0.03	0.17	0.11	0.78	2.64
GORT-3 Accuracy	3.88	4.68	2.54	3.94	0.31	0.03	0.17	-0.02	0.64	1.83
GORT-3 Fluency	2.37	3.03	1.33	2.01	0.40	0.03	0.17	0.07	0.73	2.37
GORT-3 Comprehension	9.77	6.58	9.66	6.38	0.02	0.03	0.17	-0.31	0.34	0.10

Note. WRMT-R = Woodcock Reading Mastery test- Revised; TOWRE = Test of Word Reading Efficiency; GORT-3 = Gray Oral Reading Test (3rd Ed.)

Table 3

Possible Points of Agreement and Disagreement with Tenets of Bowers (2020) Argument

Claim	Evidence for claim from Bowers (2020)	Example evidence against claim from authors (2020)?	Agreement achievable?
1. Reading wars are a ‘draw’ with no evidence favouring systematic phonics	Overall conclusion	Evidence is consistently positive and replicable for effects of explicit phonics	No
2. Whole language <i>typically</i> involves at least some ‘unsystematic’ phonics	Quotation from one whole language scholar One UK school inspector’s report from 1991	Comparable citations from whole language scholar teams No other evidence for his strong claim	No
3. Systematic phonics <i>always</i> requires eschewing real books	Opinion. No data presented.	Intervention findings from several researchers (e.g. Tse & Nicholson, 2014; Yeung & Savage, 2020)	No
4. Systematic phonics <i>always</i> requires commitment to ‘phonics first’ (to the exclusion of for example morphology or semantics)	Citations from some researchers (e.g. Adams 1994; Ehri & McCormick 1998)	Intervention findings from several researchers (e.g. Morris et al. 2012) Savage et al., 2018)	No

5. There exists limited strong evidence on specifics of phonics –

i). required numbers of GPCs,	Absence of evidence of relevant literature on GPCs	Agree	Yes
ii). optimal method (synthetic / analytic),	Synthetic phonics not <i>clearly</i> favoured in meta-analyses	Agree	Yes
iii). optimal targeting /duration	Absence of relevant evidence	Agree	Yes

6. Evidence from 12 meta-analyses does not show advantages for phonics

Explicitly argued by Bowers

Does not bear up to close scrutiny; Evidence is consistently positive, replicable, and meaningful

No

7. There exist some notable gaps in the research base on ‘what works’ including non-phonetic counterfactuals, longer term outcomes beyond word reading and for some ‘non-responders’, issues in coding of studies and publication bias

Argued for NRP and other reviews

Agreed

Yes

8. Evidence for phonics alone *is* evident but sometimes overstated
Mixed (multicomponent) approach likely to be more successful

Argued by Bowers

Agreed: suggested by several data streams (e.g. Morris et al., 2012a)

Potentially

9. “Explicitness” of teaching phonics over “systematicity” per se may be important

Not explicitly made by Bowers

Suggested by authors

Potentially

10. Phonics alone is sub-optimal

Intervention research on the “optimal” components of phonics and other aspects of language for a range of learner groups

Not explicitly made by Bowers

Suggested by authors

Potentially