How the Science of Reading Informs 21st-Century Education

Yaacov Petscher  
Sonia Q. Cabell  
Hugh W. Catts  
Donald L. Compton  
Barbara R. Foorman  
Sara A. Hart  
Christopher J. Lonigan  
Beth M. Phillips  
Christopher Schatschneider  
Laura M. Steacy  
Nicole Patton Terry  
Richard K. Wagner  

Florida State University, Tallahassee, USA

ABSTRACT

The science of reading should be informed by an evolving evidence base built on the scientific method. Decades of basic research and randomized controlled trials of interventions and instructional routines have formed a substantial evidence base to guide best practices in reading instruction, reading intervention, and the early identification of at-risk readers. The recent resurfacing of questions about what constitutes the science of reading is leading to misinformation in the public space that may be viewed by educational stakeholders as merely differences of opinion among scientists. The authors’ goals in this article were to revisit the science of reading through an epistemological lens to clarify what constitutes evidence in the science of reading, and to offer a critical evaluation of the evidence provided by the science of reading. To this end, the authors summarize those things that they believe have compelling evidence, promising evidence, or a lack of compelling evidence. The authors conclude with a discussion of areas of focus that they believe will advance the science of reading to meet the needs of all students in the 21st century.

For more than 100 years, the question of how best to teach students to read has been debated in what has been called the reading wars. The debate cyclically fades into the background only to reemerge, often with the same points of conflict. We believe that this cycle is not helpful for promoting the best outcomes for students’ educational success. Our goal in this article is to make an honest and critical appraisal of the science of reading, defining what it is, how we build a case for evidence, summarizing those things for which the science of reading has provided unequivocal answers, providing a discussion of things we do not know but that may have been oversold, identifying areas for which evidence is promising but not yet compelling, and thinking ahead about how the science of reading can better serve all stakeholders in students’ educational achievements.

At its core, scientific inquiry is the same in all fields. Scientific research, whether in education, physics, anthropology, molecular biology, or economics, is a continual process of rigorous reasoning supported by a dynamic interplay among methods, theories, and findings. It builds understandings in the form of models or theories that can be tested. Advances in scientific knowledge are achieved by the self-regulating norms of the scientific community over time, not, as sometimes believed, by the mechanistic application of a particular scientific method to a static set of questions. (National Research Council, 2002, p. 2)
What Is the Science of Reading, and Why Are We Still Debating It?

“The science of reading” is a phrase representing the accumulated knowledge about reading, reading development, and best practices for reading instruction obtained by the use of the scientific method. We recognize that the accrual of scientific knowledge related to reading is ever evolving, at times circuitous, and not without controversy. Nonetheless, the knowledge base on the science of reading is vast. In the last decade alone, over 14,000 peer-reviewed journal articles have been published that included the keyword reading based on a PsycINFO search. Although many of these studies likely focused on a sliver of the reading process individually, collectively, research studies with a focus on reading have yielded a substantial knowledge base of stable findings based on the science of reading. Taken together, the science of reading helps a diverse set of educational stakeholders across institutions (e.g., preschools, schools, universities), communities, and families to make informed choices about how to effectively promote literacy skills that foster healthy and productive lives (DeWalt & Hink, 2009; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001).

An interesting question concerning the science of reading is, Why is there a debate surrounding the science of reading? Although there are certainly disputes within the scientific community regarding best practices and new areas of research inquiry, most of the current debate seems to settle on what constitutes scientific evidence, how much value we should place on scientific evidence as opposed to other forms of knowledge, and how preservice teachers should be instructed to teach reading (Brady, 2020). The current disagreement in what constitutes the scientific evidence of reading (e.g., Calkins, 2020) is not new. During the last round of the reading wars, in the late 1990s and early 2000s, these same issues were discussed and debated. Much of the debate focused on conflicting views in epistemology between constructivists and positivists on the basic mechanisms associated with reading development. Constructivists, such as Goodman (1967) and Smith (1971), believed that reading was a natural act akin to learning language and thus emphasized giving students the opportunity to discover meaning through experiences in a literacy-rich environment. In contrast, positivists, such as Chall (1967) and Flesch (1955), made strong distinctions between innate language learning and the effortful learning required to acquire reading skills. Positivists argued for explicit instruction to help foster understanding of how the written code mapped onto language, whereas constructivists encouraged students to engage in “a psycholinguistic guessing game” (Goodman, 1967, p. 126) in which readers use their graphic, semantic, and syntactic knowledge (known as the three-cueing system) to guess the meaning of a printed word.

Research has clearly indicated that skilled reading involves the consolidation of orthographic and phonological word forms (Dehaene, 2011). Work in cognitive neuroscience has indicated that a small region of the left ventral visual cortex becomes specialized for this purpose. As students learn to read, they recruit neurons from a small region of the left ventral visual cortex within the left occipitotemporal cortex region (i.e., visual word form area) that are tuned to language-dependent parameters through connectivity to perisylvian language areas (Dehaene-Lambertz, Monzlavo, & Dehaene, 2018). This process provides an efficient circuit for grapheme-phoneme conversion and lexical access allowing efficient word-reading skills to develop. These studies have provided direct evidence for how teaching alters the human brain by repurposing some visual regions toward the shapes of letters, suggesting that cultural inventions, such as written language, modify evolutionarily older brain regions. Furthermore, studies have suggested that instruction focusing on the link between orthography and phonology promotes this brain reorganization (e.g., Dehaene, 2011).

Yet, arguments between philosophical constructivists and philosophical positivists on what constitutes the science of reading and how it informs instruction remain active today (e.g., Castles, Rastle, & Nation, 2018). In an interview with Hanford (2019), Goodman defended his advocacy for the three-cueing system by saying that the three-cueing theory is based on years of observational research: “In his view, three cueing is perfectly valid, drawn from a different kind of evidence than what scientists collect in their labs. ‘My science is different,’ Goodman said” (‘My Science Is Different’ section, paras. 13–14). Without question, observational research maintains an important place in science (see our How We Build a Case for Compelling Evidence section), but observational research devoid of rigorous methodology, testing, and replication produces spurious results and leads to biased inferences (Guyatt et al., 2011).

As scientists at the Florida Center for Reading Research, we are often frustrated when what we view to be the empirically supported evidence base about the reading process is distorted or denied in communications directed to the public and to teachers. However, Stanovich (2003) posited,

In many cases, the facts are secondary—what is being denied are the styles of reasoning that gave rise to the facts; what is being denied is closer to a worldview than an empirical finding. Many of these styles are implicit; we are not conscious of them as explicit rules of behavior. (pp. 106–107)

Stanovich proposed five different dimensions that represent styles of generating knowledge about reading. For our purposes here, we focus on the first dimension: the correspondence theory of truth versus the coherence theory of truth. This dimension hits at the heart of how
people believe something to be true. People who believe that a real world exists independent of their beliefs and that interrogating this world using rigorous principles to gain knowledge is a fruitful activity are said to subscribe to the correspondence theory of truth. In contrast, those who subscribe to the coherence theory of truth believe that something is true if the beliefs about something fit together in a logical way. In essence, something is true if it makes sense.

Stanovich believed these differing truth systems might lie at the heart of the disagreements surrounding the science of reading to explain one side stating, "Look at this mountain of evidence! How can you not believe it?" and the other side stating, "It doesn't make sense! It doesn't match up with our experiences! Why should we value your knowledge above our own?" By approaching the science of reading from the perspective of the correspondence theory of truth, we first consider how a body of compelling evidence can be generated. We then summarize issues related to the development and instruction of reading in alphabetic languages for which we believe the science of reading either has or has not yielded compelling evidence, identify what we believe are promising areas for which sufficient evidence has not yet accumulated, and suggest a number of areas that we believe will help move the science of reading forward, increasing knowledge and enhancing its positive impacts for a variety of stakeholders.

How We Build a Case for Compelling Evidence

Research is the means by which we acquire and understand knowledge about the world (Dane, 1990) to create scientific principles. Relatively few scientists would argue with the importance of using research evidence to support a principle or to make claims about reading development and the quality of reading instruction. Where significant divergence often occurs among scientists is in response to policy statements, organizations’ position stands, and the like that categorize research claims and instructional strategies into those with greater or lesser levels of evidence. This divergence among scientists is typically rooted in their applied epistemology, which can be understood as the study of whether the means by which we study evidence are themselves well designed to lead to valid conclusions. Researchers often frame the science of reading from contrasting applied epistemological perspectives. Thus, two scientists who approach the science of reading with different epistemologies will both suggest that they have principled understandings and explanations for how students learn to read; yet, the means by which those understandings and explanations were derived are often distinct.

The correspondence and coherence theories of truth described earlier are examples of explanations from contrasting epistemological perspectives. Consistent with these perspectives, researchers approaching the science of reading using a correspondence theory typically prioritize deductive methods that embed hypothesis testing, precise operationalization of constructs, and efforts to decouple the researchers’ beliefs from their interpretation and generalization of empirical evidence. Researchers approaching the science of reading using a coherence theory of truth typically prioritize more inductive methods, such as phenomenological, ethnographic, and grounded theory approaches that embed focus on the meaning and understanding that comes through a person’s lived experience and where the scientist’s own observations shape meaning and principles (e.g., Israel & Duffy, 2014).

When the National Research Council (2002) published Scientific Research in Education, a significant amount of criticism levied against the report boiled down to differences in epistemological perspectives. Yet, these genuine contrasts can often obscure contributions to the science of reading that derive from multiple applied epistemologies. Observational research, using both inductive (e.g., case studies) and deductive (e.g., correlational studies) approaches, substantively informs the development of theories and of novel instructional approaches (e.g., Scruggs, Mastropieri, & McDuffie, 2007). Public health research offers a useful parallel. As it would be unethical to establish a causal link from smoking cigarettes to lung cancer through a randomized controlled trial, that field instead used well-designed observational studies to derive claims and principles. These findings then informed later stages in the broader program of research, including randomized controlled trials of interventions for smoking cessation.

In the science of reading, principles and instructional strategies should indeed capitalize on a program of research inclusive of multiple methodologies. Yet, as the public health domain ultimately takes direction from the efficacy of smoking cessation programs, so too must the science of reading take direction from theoretically informed and well-designed experimental and quasi-experimental studies of promising strategies when the intention is to evaluate instructional practices. The use of experimental (i.e., randomized trials) and quasi-experimental (e.g., regression discontinuity, propensity score matching, interrupted time series) designs, in which an intervention is competed against counterfactual conditions, such as typical practice or alternative interventions, provides the strongest causal credibility regarding which instructional strategies are effective. The What Works Clearinghouse (WWC; e.g., 2020) of the Institute of Education Sciences and the Every Student Succeeds Act of (ESSA; 2015) are efforts by the U.S. Department of Education to hierarchically characterize the
levels of evidence currently available for instructional practices in education. The WWC uses a review framework, developed by methodological and statistical experts, for evaluating the quality and scope of evidence for specific instructional practices based on features of the design, implementation, and analysis of studies. Similarly, the ESSA uses four tiers that focus on both the design and the results of the study in which the tiers differ based on the quantity and quality of evidence supporting an approach. For both the WWC and the ESSA, quantity of evidence refers to the number of well-designed and well-implemented studies, and quality of evidence is defined by the ability of a study’s methods to allow for alternative explanations of a finding to be ruled out, for which the randomized controlled trial provides the strongest method.

As outlined earlier, the science of reading utilizes multiple research approaches to generate ideas about reading. Ultimately, we contend that the highest priority in the science of reading should be the replicable and generalizable knowledge from observational and experimental methods, rooted in a deductive research approach to knowledge generation that is framed in a correspondence theory of truth. In this manner, the accumulated evidence is built on a research foundation by which theories, principles, and hypotheses have been subjected to rigorous empirical scrutiny to determine the degree to which they hold up across variations in samples, measures, and contexts.

Compelling Evidence in the Science of Reading

In this section, we focus on a number of findings centrally important for understanding the development and teaching of reading in alphabetic languages. The evidence base provides answers varying across orthographic regularity (e.g., English vs. Spanish), reading subskill (i.e., decoding vs. comprehension), grade range or developmental level (e.g., early childhood, elementary, adolescence), and linguistic diversity (e.g., English learners, dialect speakers).

There are large differences among alphabetic languages in the rules for how graphemes represent sounds in words (i.e., a language’s orthography). In languages such as Spanish and Finnish, there is a near one-to-one relation between letters and sounds. The letter–sound coding in these languages is transparent, and they have shallow orthographies. In other languages, most notably English, there is often not a one-to-one relation between letters and sounds. The letter–sound coding in these languages is opaque, and they have deep orthographies. Students must learn which words cannot be decoded based solely on letter–sound correspondence (e.g., two, knight, laugh) and learn to match these irregular spellings to the words they represent. Where a language’s orthography falls on the shallow–deep dimension affects how quickly students develop accurate and fluent word-reading skills (Ellis et al., 2004; Ziegler & Goswami, 2005) and how much instruction on foundational reading skills is likely needed. Studies have indicated that students learning to read in English are slower to acquire decoding skills (e.g., Caravolas, Lervåg, Defior, Málkova, & Hulme, 2013). Ziegler, Stone, and Jacobs (1997) reported that 69% of monosyllabic words in English are consistent in spelling-to-phonology mappings and that 31% of the phonology-to-spelling mappings are consistent. Thus, in teaching students to read in English, the grain size of phoneme, onset-rime, and whole word matters (Ziegler & Goswami, 2005), and the preservation of morphological regularities in English spelling matters (e.g., vine vs. vineyard).

Gough and Tunmer’s (1986) “simple view of reading” model, which has been supported by a significant amount of research, provides a useful framework for conceptualizing the development of reading skills across time. It also frames the elements for which it is necessary to provide instructional support. The ultimate goal of reading is to extract and construct meaning from text for a purpose. For this task to be successful, however, the reader needs skills in both word decoding and linguistic comprehension. Weaknesses in either area will reduce the capacity to achieve the goal of reading. Decoding skills and linguistic comprehension make independent contributions to the prediction of reading comprehension across diverse populations of readers (Kershaw & Schatschneider, 2012; Sabatini, Sawaki, Shore, & Scarborough, 2010; Vellutino, Tunnier, Jaccard, & Chen, 2007). Results of several studies employing measurement strategies that allow modeling of each component as a latent variable indicate that decoding and linguistic comprehension account for almost all of the variance in reading comprehension (e.g., Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015; Lonigan, Burgess, & Schatschneider, 2018). The relative influence of these skill domains, however, changes across development. The importance of decoding skill in explaining variance in reading comprehension decreases across grades, whereas the importance of linguistic comprehension increases (e.g., Catts, Hogan, & Adlof, 2005; Foorman, Petscher, & Herrera, 2018; Garcia & Cain, 2014; Lonigan et al., 2018). By the time students are in high school, linguistic comprehension and reading comprehension essentially form a single dimension (e.g., Foorman et al., 2018).

Students’ knowledge of the alphabetic principle (i.e., how letters and sounds connect) and knowledge of the morphophonemic nature of English are necessary to create the high-quality lexical representations essential to accurate and efficient decoding (Ehri, 2005; Perfetti, 2007). Acquiring the alphabetic principle is dependent on understanding that words are composed of smaller sounds (i.e., phonological awareness) and alphabet knowledge. Both phonological awareness and alphabet knowledge are
substantial correlates and predictors of decoding skills (e.g., Wagner & Torgesen, 1987; Wagner, Torgesen, & Rashotte, 1994). Prior to formal reading instruction, young students are developing phonological awareness and alphabet knowledge, as well as other early literacy skills that are related to later decoding skills following formal reading instruction (Lonigan, Burgess, Anthony, & Barker, 1998; Lonigan et al., 2009; National Institute for Literacy, 2008; Whitehurst & Lonigan, 1998). Reading comprehension takes advantage of the reader’s ability to understand language. In most languages, written language and spoken language have high levels of overlap in their basic structure. Longitudinal studies have indicated that linguistic comprehension skills from early childhood predict reading comprehension at the end of elementary school (Catts et al., 2015; Language and Reading Research Consortium & Chiu, 2018; Mancilla-Martinez & Lesaux, 2010; Storch & Whitehurst, 2002; Verhoeven & van Leeuwe, 2008). The developmental precursors to skilled reading are present prior to school entry. Consequently, differences between children in the development of these skills forecast later differences in reading skills and are useful for identifying young students’ risk for reading difficulties.

The science of reading has provided numerous clear answers about the type and focus of reading instruction for the subskills of reading, depending on where students are on the continuum of reading development and their linguistic backgrounds. Much of this knowledge has been summarized in the practice guides produced by the Institute of Education Sciences (Baker et al., 2014; Foorman, Beyler, et al., 2016; Gersten et al., 2007, 2008; Kamil et al., 2008; Shanahan et al., 2010) and in meta-analytic summaries of research (e.g., Berkeley, Scruggs, & Mastropieri, 2010; Ehri, Nunes, Stahl, & Willows, 2001; Ehri, Nunes, Willows, et al., 2001; National Institute for Literacy, 2008; Therrien, 2004; Wanzek et al., 2013, 2016). Whereas the practice guides list several best practices, here we emphasize those practices classified as supported by strong or moderate evidence based on WWC standards.

Since the publication of the National Reading Panel’s (2000) report, and supported by subsequent research (e.g., Foorman, Beyler, et al., 2016; Gersten, Jayanthi, & Dimino, 2017), it is clear that a large evidence base provides strong support for the explicit and systematic instruction of the component and foundational skills of decoding and decoding itself. That is, teaching students phonological awareness and letter knowledge, particularly when combined, results in improved word-reading skills. Teaching students to decode words using systematic and explicit phonics instruction results in improved word-reading skills. Such instruction is effective both for monolingual English-speaking students and students whose home language is other than English (i.e., dual-language learners; Baker et al., 2014; Gersten et al., 2007), as well as students who are having difficulties with learning to read or who have an identified reading disability (Ehri, Nunes, Stahl, & Willows, 2001; Gersten et al., 2008). Additionally, providing students with frequent opportunities to read connected text supports the development of word-reading accuracy and fluency, as well as comprehension skills (Foorman, Beyler, et al., 2016; Therrien, 2004).

Similarly, a number of instructional activities to promote the development of reading comprehension have strong or moderate supporting evidence based on WWC standards. For younger students, teaching how to use comprehension strategies and how to utilize the organizational structure of a text to understand, learn, and retain content supports better reading comprehension (Shanahan et al., 2010). For older students, teaching the use of comprehension strategies also enhances reading comprehension (Kamil et al., 2008), as does explicit instruction in key vocabulary, providing opportunities for extended discussion of texts and providing instruction on foundational reading skills when students lack these skills; such instructional approaches are also effective for students with significant reading difficulties (Berkeley et al., 2010; Kamil et al., 2008).

**Lack of Compelling Evidence in the Science of Reading**

In the previous section, we highlighted practices that have sufficient evidence to warrant their widespread use. In this section, we address reading practices for which there is a lack of compelling evidence. Some practices have simply not yet been scientifically evaluated. Other practices have been evaluated, but either the evidence does not support their use based on the generalizability of the results or the studies in which they were evaluated were not of sufficient quality to meet a minimal standard of evidence (e.g., WWC standards). Although we lack sufficient space to present a comprehensive list of practices that do not have compelling evidence, we provide examples of practices that are commonplace and vary in the degree to which they have been scientifically studied.

Evidence-based decision making regarding effective literacy programs and practices for classroom use can be difficult. Often, there is no evidence of effectiveness for a program, or the evidence is of poor quality. For instance, of the five most popular reading programs used nationwide (i.e., Units of Study for Teaching Reading, Journeys, Into Reading, Leveled Literacy Intervention, Reading Recovery; Schwartz, 2019), only Leveled Literacy Intervention and Reading Recovery, both interventions for struggling readers, have had studies on them that meet WWC standards. The evidence indicates that there were mixed effects across outcomes for Leveled Literacy Intervention and positive or potentially positive effects for Reading Recovery (e.g., Chapman & Tunmer, 2002).
Classroom reading programs are typically built around the notion of evidence-informed practices—teaching approaches that are grounded in quality research—but have not been subjected to direct scientific evaluation. As a consequence, it is currently impossible for schools to select basal reading programs that adhere to strict evidence-based standards (e.g., ESSA, 2015). As an alternative, schools must develop selection criteria for choosing classroom reading programs informed by the growing scientific evidence on instructional factors that support early reading development (e.g., Castles et al., 2018; Foorman, Smith, & Kosano, 2017; Rayner et al., 2001).

Common instructional approaches that lack generalizable empirical support include such practices as close reading (Welsch, Powell, & Robnolt, 2019), use of decodable text (Jenkins, Peyton, Sanders, & Vadasy, 2004), sustained silent reading (National Reading Panel, 2000), multisensory approaches (Birsh, 2011), and the three-cueing system to support word recognition development (Seidenberg, 2017). Some of these instructional approaches rest on sound theoretical and pedagogical grounds. For example, giving beginning readers the opportunity to read decodable texts provides practice in applying the grapheme–phoneme relations that these students have learned to successfully decode words (Foorman, Beyler, et al., 2016), thus building lexical memory to support word-reading accuracy and automaticity (Ehri, 2020). However, the only study to experimentally examine the impact of reading more versus less decodable texts as part of an early intervention phonics program for at-risk first graders found no differences between the two groups on any of the posttest measures (Jenkins et al., 2004). Such a result does not rule out the possibility of the usefulness of decodable texts but rather indicates the need to disentangle the active ingredients of effective interventions to specify what to use, when, how often, and for whom.

Similarly, multisensory approaches (e.g., Orton–Gillingham) that teach reading by using multiple senses (i.e., sight, hearing, touch, movement) to help students make systematic connections among language, letters, and words (Birsh, 2011) are commonplace and have considerable clinical support for facilitating reading development in students who struggle to learn to read. However, there is little scientific evidence indicating that a multisensory approach is more effective than similarly structured phonological-based approaches that do not include a strong multisensory component (e.g., Boyer & Ehri, 2011; Ritchey & Goeke, 2006; Torgesen et al., 2001). With further research, we may find that a multisensory component is a critical ingredient of intervention for struggling readers, but we lack this empirical evidence currently.

Instruction in reading comprehension is another area where, despite some studies showing moderate or strong support (see the previous section), other practices are employed despite limited support for them (e.g., Boulay, Goodson, Frye, Blocklin, & Price, 2015). The complexity of reading comprehension relies on numerous cognitive resources and background knowledge; as a result, intervention directed exclusively at one component or another is not likely to be that impactful. For example, consider intervention directed at improving vocabulary. Despite a solid theoretical foundation (Perfetti & Stafura, 2014) and numerous studies showing a sizable relation between vocabulary and reading comprehension (Protopapas, Mouzaki, A., Sideridis, Kotsolakou, & Simos, 2013; Sénéchal, Ouellette, & Rodney, 2006; Wagner, Muse, & Tannenbaum, 2007), evidence of the effectiveness of primarily working vocabulary to improve reading comprehension is weak. Numerous studies have sought to facilitate comprehension using instructional approaches that vary from teaching words in isolation to practices that involve instruction in the use of context to learn the meanings of unfamiliar words. Instruction has also included strategies to determine the meanings of words through word study and morphological analysis (e.g., Beck & McKeown, 2007; Lesaux, Kieffer, Kelley, & Harris, 2014).

Although these practices have been effective in increasing vocabulary knowledge of the words taught, there is limited evidence of transfer to untaught words (as measured by standardized measures) or to improvement in general reading comprehension (Elleman, Lindo, Morphy, & Compton, 2009; Lesaux, Kieffer, Faller, & Kelley, 2010). Such findings do not mean that vocabulary instruction is not a useful practice; rather, by itself, it is not sufficient to improve reading comprehension. To make meaningful gains, intervention for reading comprehension likely requires addressing multiple components of language and teaching content knowledge (see the next section) to make sizable gains.

Other instructional practices go directly against what is known from the science of reading. For example, isolating the three-cueing approach to support early word recognition (i.e., relying on a combination of semantic, syntactic, and graphophonemic cues simultaneously to formulate an intelligent hypothesis about a word’s identity) ignores 40 years of overwhelming evidence that orthographic mapping involves the formation of letter–sound connections to bond the spelling, pronunciation, and meaning of a specific word in memory (see Ehri, 2020). Moreover, relying on alternative cueing systems impedes the building of automatic word recognition skill that is the hallmark of skilled word reading (Stanovich, 1990, 1991). The English orthography, being both alphabetic-phonemic and morphophonemic, clearly privileges the use of various levels of grapheme–phoneme correspondences to read words (Frost, 2012), with rapid context-free word recognition being the process that most clearly distinguishes good from poor readers (Perfetti, 1992; Stanovich, 1980).

Guessing at a word amounts to a lost learning trial to help students learn the orthography of the word and thus
reduce the need to guess the word in the future (Castles et al., 2018; Share, 1995).

Similarly, alternative approaches to improving reading skills for struggling readers often fall well outside the scientific consensus regarding sources of reading difficulties. Some of these approaches are based on the tenet that temporal processing deficits in the auditory (e.g., Tallal, 1984) and visual (e.g., Stein, 2019) systems of the brain are causally related to poor word-reading development. Although there is some evidence that typically developing and struggling readers differ on measures tapping auditory (Casini, Pech-Georgel, & Ziegler, 2018; Protopapas, 2014) and visual (e.g., Eden et al., 1996; Olson & Datta, 2002) processing skills, there is little evidence to support the use of instructional programs designed to improve auditory or visual systems to ameliorate reading problems (Strong, Torgerson, Torgerson, & Hulme, 2011). Further, interventions designed to decrease visual confusion (e.g., Dyslexie font) or modify transient channel processing (e.g., Irlen lenses) to improve reading skill for students with reading disability have also failed to garner scientific support (Hyatt, Stephenson, & Carter, 2009; Iovino, Fletcher, Breitmeyer, & Foorman, 1998; Marinus et al., 2016). Similarly, although use of video games to improve reading via enhanced visual attention has been reported to be an effective intervention for students with reading disability (Peters, De Losa, Bavin, & Crewther, 2019), studies of this supplemental intervention approach have not compared it with standard supplemental approaches. Finally, studies of interventions designed to enhance other cognitive processes, such as working memory, have also lacked evidence effectiveness in terms of improved reading-related outcomes (e.g., Melby-Lervåg, Redick, & Hulme, 2016).

Promising but Not (Yet) Compelling Evidence in the Science of Reading

There are many promising areas of research that are poised to provide compelling evidence to inform the science of reading in the coming years. As we do not have space in this article to provide a comprehensive list, we highlight only a few promising areas in prevention research and elementary education research.

Promising Directions in Prevention Research

Research on the prevention of reading problems is critical for our ability to reduce the number of students who struggle in learning to read. One area of prevention research that has great promise but needs more evidence is how to more fully develop preschoolers’ language abilities that support later reading success. Both correlational and experimental findings indicate that providing young learners with opportunities to engage in high-quality conversations, coupled with exposure to advanced language models, matters for language development (Cabell, Justice, McGinty, DeCoster, & Forston, 2015; Dickinson & Porche, 2011; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011; Wasik & Hindman, 2020). Yet, most programs have had a more robust impact on proximal language learning (i.e., learning taught words) than on generalized language learning as measured with standardized assessments (Marulis & Neuman, 2010).

Promising studies that have demonstrated significant effects on young learners’ general language development elucidate potential points of leverage. First, improving the connection between the school and home contexts by including parents as partners can promote synergistic learning for students as language-learning activities in school and home settings are increasingly aligned (e.g., Lonigan & Whitehurst, 1998). A second leverage point is increasing attention to students’ active use of language in the classroom to promote a rich dialogue between children and adults (e.g., Lonigan et al., 2011; Wasik & Hindman, 2020). A third leverage point is integrating content area instruction into early literacy instruction to improve language learning, such as building students’ conceptual knowledge of the social and natural world and teaching vocabulary words within the context of related ideas (e.g., Gonzalez et al., 2010).

Promising Directions in Elementary Education Research

Here, we present two promising areas in reading research with elementary-age students: one focused on improving linguistic comprehension and one focused on improving decoding, consistent with the simple view of reading.

The knowledge a reader brings to a text is the chief determinant of whether the reader will understand that text (Anderson & Pearson, 1984). Thus, building knowledge is an essential, yet neglected, part of improving linguistic comprehension (Cabell & Hwang, 2020, this issue). Teaching reading is most often approached in early elementary classrooms as a subject that is independent from other subjects, such as science and social studies (Palincsar & Duke, 2004). As such, reading is taught using curricula that do not systematically build students’ knowledge of the social and natural world. Instruction in reading and the content areas does not have to be an either/or proposition. Rather, the teaching of reading and of content area learning can be simultaneously taught and integrated to powerfully impact students’ learning of both reading and content knowledge (e.g., Connor et al., 2017; Kim et al., 2020; Williams et al., 2014). This area of
research is promising but not yet compelling due to the small number of experimental and quasi-experimental studies that have examined either integrated content area and literacy instruction or content-rich English language arts instruction in K–5 settings (approximately 31 studies). Through meta-analysis, this corpus of studies has demonstrated that combining knowledge building and literacy approaches has a positive impact on both vocabulary and comprehension outcomes for elementary-age students (Hwang, Cabell, White, & Joiner, 2019). Further rigorous studies are needed that test widely used content-rich English language arts curricula (Cabell & Hwang, 2020); also required is new development of integrative and interdisciplinary approaches in this area.

There has also been promising research on helping students decode words more efficiently. It is widely accepted that students with reading difficulties often have underlying deficits in phonological processing (e.g., Brady & Shankweiler, 1991; Stanovich & Siegel, 1994; Torgesen, 2000; Vellutino et al., 1996), and these deficits are believed to disrupt the acquisition of spelling-to-sound translation routines that form the basis of early decoding skill development (e.g., Rack, Snowling, & Olson, 1992; van IJzendoorn & Bus, 1994). For a developing reader, decoding an unfamiliar letter string can result in either full or partial decoding. During partial decoding, the reader must match the assembled phonology from decoding with his or her lexical representation of the word (Venezky, 1999). For example, encountering the word island might render the incorrect but partial decoding attempt “izland.” Flexibility with the partially decoded word is referred to as the reader’s set for variability or the reader’s ability to go from the decoded form to the correct pronunciation of the word. This skill serves as a bridge between decoding and lexical pronunciations and may be an important second step in the decoding process (Elbro, de Jong, Houter, & Nielsen, 2012).

The matching of partial phonemic-decoding output is facilitated by the reader’s decoding skills, the quality of the reader’s lexical word representation, and the potential contextual support of the text (Nation & Castles, 2017). Correlational studies have indicated that students’ ability to go from a decoded form of a word to a correct pronunciation of it (their set for variability) predicts the reading of irregular words (Tunmer & Chapman, 2012), regular words (Elbro et al., 2012), and nonwords (Steacy, Compton, et al., 2019). Set for variability has also been found to be a stronger predictor of word reading than phonological awareness in students in grades 2–5 (e.g., Steacy, Wade-Woolley, et al., 2019). Studies in this area have suggested that students can benefit from being encouraged to engage with the irregularities of English (Dyson, Best, Solity, & Hulme, 2017) to promote the implicit knowledge structures needed to read and spell these complex words. Additional research has suggested that training in set for variability can be effective in promoting early word-reading skills (e.g., Savage, Georgiou, Parrila, & Maiorino, 2018; Zipke, 2016). The work done in this area to date has suggested that set for variability requires knowledge structures and strategies, which can be developed through instruction, that allow successful matching of partial phonemic-decoding output with the corresponding phonological, morphological, and semantic lexical representations.

Where Do We Go Next in the Science of Reading?

Basic Science Research

The “science of reading” community has reached some consensus on the typical development of reading skill and how individual differences may alter this trajectory (e.g., Boscardin, Muthén, Francis, & Baker, 2008; Hjetland et al., 2019; Peng et al., 2019). Less is known about factors and mechanisms related to reading among diverse learners, a critical barrier to the field’s ability to address and prevent reading difficulty when it arises. Investigations with large and diverse participant samples are needed to improve understanding of how child characteristics additively and synergistically affect reading acquisition (Hernandez, 2012; Lonigan, Farver, Nakamoto, & Eppe, 2013). Research on disentangling the influence of English learner status for students who also have identified disabilities (Solari, Petscher, & Folsom, 2014; Wagner, Francis, & Morris, 2005) has been insufficient. Greater attention to how language variation (e.g., dialect use) and differences in language experience affect reading development is crucial (Seidenberg & MacDonald, 2018; Terry, Connor, Thomas-Tate, & Love, 2010; Washington, Branum-Martin, Sun, & Lee-James, 2018). New realizations of the interaction between child characteristics and the depth of the orthography have also highlighted the importance of understanding the role of statistical learning in early reading development (Seidenberg, 2005).

Consider, for example, the different manners by which two alphabetic writing systems (Spanish and English) capture the phonological properties of the words they represent. In Spanish, there is a nearly one-to-one mapping between letters and phonemes, whereas in English, phonemes can be represented by either a single letter (e.g., the p in pan) or a letter cluster (e.g., the ph in graph), and many graphemes, particularly vowels, can be pronounced in more than one way (cf. pint vs. hint, bead vs. head). In English, much of the ambiguity is associated with the pronunciation of vowels; for instance, ea is pronounced as /i/ in beat, /ɛ/ in head, and /eɪ/ in steak. This ambiguity in the English orthographic-to-phonological mapping system poses significant challenges to beginning readers of English (e.g., Seymour, Aro,
Erskine, 2003). However, studies have suggested that students become sensitive, through statistical learning, to the probabilistic regularities representing context-dependent orthographic–phonological relations that exist in the English orthography and, further, that students who are better readers exhibit greater sensitivity to these constraints of the orthography when reading words (Steacy, Compton, et al., 2019; Steacy, Wade-Woolley, et al., 2019; Treiman, Kessler, Zevin, Bick, & Davis, 2006).

More work is certainly needed exploring how best to promote this type of learning in developing readers (see Seidenberg, Cooper Borkenhagen, & Kearns, 2020, this issue). Likewise, a better understanding of the role of executive function, socioemotional resilience factors, and biopsychosocial risk variables (e.g., poverty, trauma) on reading development is critical. Additional research like this, in English and across languages, is needed to develop effective instruction and assessments for all learners.

A clearer understanding of child and contextual influences on the development of reading also will support improvements in how early and accurately students at risk for reading difficulties and disabilities are identified. Currently, numerous challenges remain in identifying students early enough to maximize benefits of interventions (Colenbrander, Ricketts, & Kearns, 2018; Gersten, Newman-Gonchar, Haymond, & Dimino, 2017). Investigators often use behavioral precursors or correlates of reading to estimate students’ risk for reading failure. Whereas this work has shown some promise (Catts, Herrera, Nielsen, & Bridges, 2015; Compton et al., 2010; Compton, Fuchs, Fuchs, & Bryant, 2006; Lyytinen, Erskine, Hämäläinen, Torppa, & Ronimus, 2015; Thompson et al., 2015), identification of risk typically involves high error rates, especially for preschoolers and kindergartners who might benefit most from early identification and intervention. Similar challenges to accuracy have emerged when identifying older students with reading disabilities. Historically, this process has relied on discrepancy models (e.g., between reading skill and general cognitive aptitude), often yielding just a single comparison on which decisions are based (Brown Waesche, Schatschneider, Maner, Ahmed, & Wagner, 2011).

Challenges to identification for both younger and older students may be best met with frameworks recognizing that reading problems have multiple contributing causes (Pennington et al., 2012). Future research is needed that explores how models of early identification and risk can be informed, and improved, by consideration of the many individual, family, and contextual characteristics that may contribute in an additive or interactive manner to a student’s early reading trajectory (e.g., Erbeli, Hart, Wagner, & Taylor, 2018; Spencer et al., 2014; Wagner et al., 2019). In particular, evaluation is needed of models that include both risk factors and factors that may be protective, to see if these models increase the likelihood of correctly identifying those students most in need of additional instructional support (e.g., Catts & Petscher, 2020; Haft, Myers, & Hoeft, 2016). Even if beneficial, it is likely that for early identification to be maximally effective, early risk assessments will need to be combined with progress monitoring of response to instruction (Miciak & Fletcher, 2020). Of course, for such an approach to be successful, all students must receive high-quality reading instruction from the beginning, and interventions need to be in place to address students who show varying levels of risk (Foorman, Beyler, et al., 2016). Identifying students at risk and providing appropriate intervention early on has the potential to significantly improve reading outcomes and reduce the negative consequences of reading failure.

**Intervention Innovations**

Despite successes, too many students still struggle to read novel text with understanding, and intervention design efforts have not fully met this challenge (Compton, Miller, Elleman, & Steacy, 2014; Phillips, Connor, Lonigan, Willis, & Crowe, 2016; Vaughn et al., 2017). Greater creativity and integration of research from a broader array of complementary fields, including cognitive science and behavioral genetics, may be required to deal with long-standing problems. For example, genetic information may have causal explanatory power, so randomized trials are needed to evaluate the efficacy of using such information to select and individualize instruction and intervention (Hart, 2016). The field would benefit from increased attention to the problem of fading intervention effects over time. Although there can be detectable effects of interventions several years after they are completed (e.g., Blachman et al., 2014; Vadasy, Nelson, & Sanders, 2011; Vadasy & Sanders, 2013), invariably effect sizes reduce over time. A meta-analysis of long-term effects of interventions for phonemic awareness, fluency, and reading comprehension found a 40% reduction in effect sizes within one year post-intervention (Suggate, 2016). Perhaps reading interventions with larger initial effects or sequential reading interventions with smaller but cumulating effects would be more resistant to fade-out.

Solutions to the problem of diminishing effects may be inspired by examples from other fields. The field of memory includes examples of content that appears immune from forgetting. This phenomenon has been called permastore (Bahrick, 1984). For example, people only meaningfully exposed to a foreign language in school classes will still retain some knowledge of the language 50 years later. Additionally, expertise in the form of world-class performance appears to result from cumulative effects of long-term deliberate practice (Ericsson, 1996), and skilled reading can be viewed as an example of expert performance (Wagner & Stanovich, 1996). Informed by these concepts and by advances in early math instruction
(e.g., Kang, Duncan, Clements, Sarama, & Bailey, 2019; Sarama, Clements, Wolfe, & Spitler, 2012), reading intervention studies should prioritize follow-up evaluations, including direct comparisons of follow-through strategies aimed at sustaining benefits from earlier instruction. For example, studies should evaluate booster interventions, professional development that better aligns cross-grade instruction, and how reteaching and cumulative review may consolidate skill acquisition across time (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Smolen, Zhang, & Byrne, 2016).

**Translational and Implementation Science**

If the science of reading is to be applied in a manner resulting in achievement for all learners, the field must increase its focus on processes supporting implementation of evidence-based reading practices in schools. The field can leverage its considerable evidence base to systematically investigate, with replication, both the effectiveness of reading instructional practices with diverse learners and the processes that facilitate or prevent adoption, implementation, and sustainability of these practices (National Research Council, 2002; Schneider, 2018; Slavin, 2002). Research on these processes in educational contexts may be best facilitated by making use of methodological and conceptual tools developed within the traditions of translation and implementation science research (Eccles & Mittman, 2006; Gilliland et al., 2019). For example, these frameworks can support studies on whether and how educators and policymakers use information about evidence to inform decision making (e.g., Farley-Ripple, May, Karpyn, Tilley, & McDonough, 2018), and studies on how institutional routines may need to be adapted to best integrate new procedures and practices (e.g., scheduling changes in the school day; Foorman, Dombek, & Smith, 2016).

Reading research that uses translational and implementation science frameworks and methodologies (e.g., Solari et al., 2020, this issue) will make more explicit the processes of adoption, implementation, and sustainability and how these interact within diverse settings and with multiple populations (Brown et al., 2017; Fixsen, Blase, Metz, & Van Dyke, 2013; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). This work will be guided by new questions, not only asking what works but also what works for whom under what conditions and what factors promote sustainability of implementation. Innovative studies would adhere to rigorous scientific standards; prioritize hypothesis testing within a deductive, experimental framework; and leverage qualitative methodologies to systematically explore implementation processes and factors (Brown et al., 2017). Results could iteratively inform the breadth of scientific reading research, including basic mechanisms related to reading and the development of novel assessments and interventions to support achievement among diverse learners in diverse settings (Cook & Odom, 2013; Douglas, Campbell, & Hinckley, 2015; Forman et al., 2013).

**Conclusion**

There has recently been a resurgence of the debate on the science of reading, and in this article, we described the existing evidence base and possible future directions. Compelling evidence is available to guide understanding of how reading develops and identify proven instructional practices that impact both decoding and linguistic comprehension. Whereas there is some evidence that is either not compelling or has yet to be generated for instructional practices and programs that are widely used, the scientific literature on reading is ever expanding through contributions from the fields of education, psychology, linguistics, communication science, neuroscience, and computational sciences. As these additions to the literature mature and contribute to the evidence base, we anticipate that they will inform and shape the science of reading and the science of teaching reading.

**NOTES**

This work was supported by funding from the Chan Zuckerberg Initiative, the Institute of Education Sciences (grants R305A160241, R305A170430, R305A170635, R305F100005, R305F100027, R324A1 80020, R305H170054, and R324B19002), the Spencer Foundation (grant 201700094), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grants P50HD52120, P20HD091013, HD095193, and HD072286). The opinions expressed are those of the authors and do not represent views of the funding agencies or the university. The first author was determined by group consensus; the other authors contributed equally and are listed alphabetically.

**REFERENCES**


Hernandez, D.J. (2012). Double jeopardy: How third-grade reading skills
Haft, S.L., Myers, C.A., & Hoeft, F. (2016). Socio-emotional and cogni-
Psychology
sonalized education.
Haft, S.L., Myers, C.A., & Hoeft, F. (2016). Socio-emotional and cogni-
Precision education initiative: Moving toward personal-
ized education. Mind, Brain, and Education, 10(4), 209–211.
Hernandez, D.J. (2012). Double jeopardy: How third-grade reading skills
A systematic review of the research on the effect of knowledge building in
literacy instruction on comprehension and vocabulary in the elementary
years. Paper presented at the annual meeting of the Literacy Research
Association, Tampa, FL.
troversial educational practices: Perceptual motor programs, sensory
integration, and tinted lenses. Education and Treatment of Children,
32(2), 313–342. https://doi.org/10.1353/etc.0.0054
Colored overlays for visual perceptual deficits in children with reading
disability and attention deficit/hyperactivity disorder: Are they differentially
effective? Journal of Clinical and Experimental Neuropsychology, 20(6),
791–806. https://doi.org/10.1076/jcen.20.6.791.1113
Israel, S.E., & Duffy, G.G. (Eds.). (2014). Handbook of research on reading
reading decodable texts in supplemental first-grade tutoring. Scientific
Studies of Reading, 8(1), 53–85. https://doi.org/10.1207/
s1352799xssr0801_4
Kamil, M.L., Borman, G.D., Dole, J., Kral, C.C., Salinger, T., & Torgesen,
J. (2008). Improving adolescent literacy: Effective classroom and inter-
vention practices. A practice guide (NCEE 2008–4027). Washington,
DC: National Center for Education Evaluation and Regional Assistance,
Institute of Education Sciences, U.S. Department of Education.
(2019). The roles of transfer of learning and forgetting in the presence
and fadeout of early childhood mathematics interventions. Journal of Clinical
and Experimental Neuropsychology, 111(4), 590–603. https://doi.org/
10.1017/ecp.2017.133
the simple view of reading. Reading and Writing, 25, 433–464.
Kim, J.S., Burkhauer, M.A., Mesite, L.M., Asher, C.A., Relyea, J.E.,
Fitzgerald, J., & Elmore, J. (2020). Improving reading comprehension,
sCognitive domain knowledge, and reading engagement through a
first-grade content literacy intervention. Journal of Educational
Psychology. Advance online publication. https://doi.org/10.1037/
edu000465
Language and Reading Research Consortium & Chiu, Y.D. (2018). The
simple view of reading across development: Prediction of grade 3
reading comprehension from prekindergarten skills. Remedial and Special
32518762055
Lesaux, N.K., Kieffer, M.J., Faller, S.E., & Kelley, J.G. (2010). The effec-
tiveness and ease of implementation of an academic vocabulary
intervention for linguistically diverse students in urban middle schools.
Reading Research Quarterly, 45(2), 196–228. https://doi.org/
10.1598/RRQ.45.2.3
Lesaux, N.K., Kieffer, M.J., Kelley, J.G., & Harris, J.R. (2014). Effects of
academic vocabulary instruction for linguistically diverse adoles-
cents: Evidence from a randomized field trial. American Educational
31214532165
McQueen, J. (2009). The nature of preschool phonological processing
abilities and their relations to vocabulary, general cognitive abili-
ties, and print knowledge. Journal of Educational Psychology, 101(2),
345–358. https://doi.org/10.1037/a0013837
Development of phonological sensitivity in two- to five-year-old
10.1037/0022-0663.90.2.294
simple view of reading with elementary school children: Still simple
after all these years. Remedial and Special Education, 39(5), 260–273.
https://doi.org/10.1177/0741932518764833
Developmental trajectories of preschool early literacy skills: A com-
parison of language-minority and monolingual-English children.
1037/a0031408
Promoting the development of preschool children’s emergent liter-
acy skills: A randomized evaluation of a literacy-focused curriculum and
two professional development models. Reading and Writing, 24,
305–337. https://doi.org/10.1007/s11145-009-9214-6
and teacher involvement in a shared-reading intervention for pre-
school children from low-income backgrounds. Early Childhood
Research Quarterly, 13(2), 263–290. https://doi.org/10.1016/S0885-
2006(99)00388-6
Lyytinen, H., Erskine, J., Hämäläinen, J., Torppa, M., & Ronimus, M.
(2015). Dyslexia-early identification and prevention: Highlights of
the Jyvaskyla Longitudinal Study of Dyslexia. Current Developmental
Disorders Reports, 1, 330–338. https://doi.org/10.1007/s40474-015-
0067-1
comprehension for struggling readers: The case of Spanish-speaking
children.
Journal of Educational Psychology, 102(3), 701–711. https://doi.org/10.1037/a0019135
Marinus, E., Mostard, M., Segers, E., Schubert, T.M., Madelaine, A.,
& Wheldall, K. (2016). A special font for people with dyslexia: Does it
1002/dys.1527
Marulis, L.M., & Neuman, S.B. (2010). The effects of vocabulary inter-
vention on young children’s word learning: A meta-analysis. Review of
Educational Research, 80(3), 300–335. https://doi.org/10.3102/
0034653410377087
training does not improve performance on measures of intelligence or
other measures of “far transfer” evidence from a meta-analytic