The Language and Literacy Spectrum

Volume 34 | Issue 1 Article 1

May 2024

Examining the Evidence for Selecting Reading Programs in a Large Urban School System

Francine Falk-Ross
Pace University, ffalkross@pace.edu

Kathleen A. Gormley
Russell Sage College, gormlk@sage.edu

Peter McDermott
Pace University, pmcdermott@pace.edu

Examining the Evidence for Selecting Reading Programs in a Large Urban School System

Cover Page Footnote/Acknowledgements

I will serve as the corresponding author, but professors Fran Falk-Ross (Pace University) and Kathleen A Gormley (Russell Sage College) are co-authors.

Public policy makers and educators have long affirmed the importance of literacy in developing and sustaining a well informed and democratic society (Apple et al., 2022; Dewey, 1916; Freire, 1970; Goodlad et al., 2004). Literacy nourishes people's abilities and skills in making thoughtful decisions about their everyday lives, and high levels of literacy empowers people to critically evaluate issues affecting themselves, their loved ones, and communities. Yet, despite the general agreement about the importance of literacy in today's world, there exist passionate debates about the best way to teach children to read. Notably, in the last year the country's largest school system, the New York City's Department of Education (NYCDOE), departed from its decades old support of "balanced literacy" in favor of more skill-based approaches for teaching children to read. In this manuscript we examine the empirical evidence supporting the adaptation of the three reading programs now used in the country's largest school system.

Research Focusing on Teaching Children to Read

Historically, there have been extensive debates about the best way to teach children how to read. These "reading wars" (Pearson, 2004) focus on the place basic skills have in learning to read, and the conflict over how to teach reading is evident in classic mid-20th-century publications, such as Flesch's (1955) *Why Johnny Can't Read* and Chall's (1967), *Reading the Great Debate*, where both authors maintained that programs emphasizing decoding skills were superior to those that did not. Yet, during the same period, research by Bond and Dykstra (1967), which became known as the "First Grade Studies," found that there was no single best method for teaching reading.

The 1970s witnessed significant changes to U.S. reading instruction. At that time many school districts relied on basal programs (Shannon, 1983; Smith, 2002) containing grade-level texts with workbooks and skill sheets as resources for teaching reading. However, proponents of psycholinguistics theory and "whole language" (Smith, 1971; Goodman, 1967; Goodman & Burke, 1970) criticized the emphasis of skill instruction found in many of the basal programs, and, instead, argued for teaching with authentic literature that would engage and motivate to read, as well as serve as models for children's learning to write. The "whole language" movement came to represent the dominant model of teaching reading throughout the 1980s.

In the early 1980s the Reagan administration charged a panel of experts to examine the U.S. educational system so that schools would become more competitive and accountable in the global context. The report of this panel, "A Nation at Risk" (National Commission on Excellent in Education, 1983), which received widespread and long-lasting attention, argued that the U.S. education system was failing its children and weakening the country's competitive edge globally. This report stimulated subsequent debate and change in terms of

establishing national standards and accountability policies that would emerge in subsequent years and precipitate reform in reading education.

Several studies in the 1990s produced a rethinking of methods for teaching reading. Adams' (1990) review of existing research about teaching children to read argued that programs containing systematic phonics instruction led to higher levels of reading achievement than programs that did not have the same emphasis. Research reports from the National Reading Council (Snow, Burns & Griffin,1998) and the National Reading Panel (NRP) (2000) similarly supported the importance of teaching word identification skills in beginning reading, but these reports also argued that phonics was only one element of a comprehensive literacy program with vocabulary and comprehension instruction being equally essential in children's learning to read. As a result of the arguments made in these studies, many school districts shifted their methods to what was commonly known as "balanced reading" instruction (Pressley, 1998) where decoding skills were taught in conjunction with meaningful oral and written language activities. Subsequent studies pertaining to the increasing social, cultural and linguistic diversity of U.S. children indicated the importance of integrating higher level thinking skills that drew upon children's life experiences for improving their reading achievement (e.g., Taylor, B.M., Pearson, Clark, & Walpole, 2000; Teale, & Gambrell, 2007, Walker-Dalhouse & Risko, 2008).

A momentous shift in reading instruction occurred during the Bush administration with the passage of the No Child Left Behind legislation (NCLB, 2002). To receive federal funding NCLB required that school districts provide evidence documenting the effectiveness of the reading programs used in their schools. The current emphasis of having scientific evidence to support methods of teaching reading has evolved from that legislation, and it is now often referred to as the "Science of Reading" (SOR) (Duke & Cartwright, 2021; Cervetti et al., 2020). However, the meaning of SOR varies depending upon the users' understanding of the term "science." For some, science means "proof" that a particular method of teaching works, whereas for others it "informs" teaching practices. Yet, there is wide variability in what proof or inform actually mean; almost all educational research will inform one's thinking in one way or another about teaching reading, but only randomized studies with control and treatment groups can be used as evidence that one method works better than others (Duke & Martin, 2011; Shanahan, 2020a). Yet, despite these concerns, the SOR movement has gained national momentum and popularity, especially with the public, despite there being many criticisms of it among literacy researchers (e.g., Goldenberg, 2020; Hoffman, Hikida & Sailors, M., 2020; Shanahan 2020b; Terry, 2020; Thomas, 2022).

The SOR movement flows from previous calls for having evidence to support methods of teaching reading in schools. Historically, the argument for

having supporting evidence to document instructional effectiveness was derived from medicine (Baron, 2018; Slavin, 2002). At the turn of 20th century medicine moved to evidence-based models of practice, rather than anecdotal and intuitive treatments of patients that had been in effect for so many years. The shift to evidence-based treatment represented a major change in medical practice with its "gold standard" becoming experimental research to document the effectiveness of one particular medical intervention over others. The use of evidence-based practices in medicine resulted in enormous gains in research and public health. The promise was that a medical model could also be applied to education research and advance classroom teaching, just as the results of randomized experimental studies revolutionized medicine in the previous century. As Slavin (2002) wrote

> Evidence-based policies for education would be important at any time, but they are especially important today, given the rise of accountability. State and national governments are asserting stronger control over local education, primarily by establishing consequences for schools based on gains or losses on state assessments. (p.19)

Since then, policymakers have prioritized the use of evidence when endorsing classroom methods for teaching reading. The aforementioned National Reading Panel (2000), for example, examined experimental and quasi-experimental studies of reading, but it excluded case and qualitative inquiries when producing its report of best methods for teaching reading. The popularity of evidence-based instruction and the SOR movement, in particular, are displayed in online publications such as "Reading Rockets" (Farrell, Hunter, Davidson, & Osenga, 2019) and in widely read NYTimes essays by established columnists (e.g., Kristoff, 2023) and guest opinion writers (Hanford, 2022). Yet despite the wide-spread public enthusiasm for the SOR perspective on teaching methods, the research community has been hesitant to rush to generalize SOR findings into pedagogical applications for classroom teaching.

The rise of evidence-based practices has similarly influenced classroom teachers' decision-making (Harris et al., 2017; Lewis et al., 2007; Rasinski, Homan & Biggs, 2009; Roehling, Hebert, Nelson & Bohaty, 2017). Many teachers are now required to systematically document students' growth in their use of reading skills, and in our experience, some building leaders require teachers to enter students' performance data daily into online portals after which curriculum coordinators monitor and assess student learning and classroom teaching practices. Although there have been cautions to such evidenced-based teaching, particularly with the over reliance of the time needed for data collection rather than instruction (e.g., Neuman, 2016), many of today's teachers are expected to routinely collect and analyze children's performance data in reading, as well as use evidence-based pedagogies in their classroom practices (Fien

Chard, & Baker, 2021). Still, the process by which school districts adopt new reading programs is an under-researched area of investigation and may not include key stakeholders, such as building principals, teachers and families, in the decision-making process (Vaughn et al., 2021).

Teaching and Assessing Reading in New York City: A Shift in Focus

The largest school system in the country is the New York City Department of Education (NYCDOE) with approximately 1,200,000 students. Historically, the NYCDOE allowed its 32 school districts autonomy in selecting the reading curriculum felt to best fit their students' literacy needs. However, in 1987, after adopting a strong mayor model of school control, the city encouraged balanced literacy for use in all of its school districts (Stein & D'Amico, 2002), and until recently more than half of the city schools continued to use a balanced literacy model of reading instruction in the majority of its schools.

With the appointment of a new chancellor in 2022, the NYCDOE mandated a change in its elementary reading programs. The new chancellor directed all of the city's elementary schools to implement an approved phonics program to be used for 30 minutes a day, which would be in addition to their established reading curriculum, and in the 2023-24 school year more than half of the city schools were required to adopt one of three reading programs for its elementary schools. The three approved reading programs were the Great Minds publishing company's "Wit & Wisdom," Houghton-Mifflin-Harcourt's "Into Reading" and EL education's "Expeditionary Learning." The reason these reading programs were adopted was that they reportedly contained a stronger and more consistent emphasis on phonics instruction than found in the balanced literacy curriculum (Zimmerman, 2023). Moreover, the school system believed the new reading programs would improve children's performance on national and statewide examinations.

Trends in NYC Children's Reading Achievement

The NYC elementary schools participate in two widely known assessments of children's reading performance. The National Assessment of Educational Progress (NAEP) and the New York State Education Department's (NYSED) English Language Arts exam present achievement data regarding children's reading performance, and this data is publicly available and often reported in the media. We analyzed this data to identify patterns in NYC children's reading performance. In particular, we wanted to learn if there was clear evidence signaling the need for immediate change in the city's reading curriculum.

The National Assessment of Educational Progress (NAEP) examinations are administered every two years for students in grades 4, 8 and 12. NAEP offers

state and large city data for the country's largest school systems, including NYC. NAEP has used the same assessment framework since 2009, which is designed to elicit students' comprehension with informational and literary texts, as well as other comprehension skills, such as integrating, interpreting and evaluating textual information.

Analysis of NYC children's NAEP results reveal that approximately 25% of its fourth and eighth graders achieved "proficient or advanced" performance levels in the years 2013, 2015, 2017, 2019 and 2022. Additionally, NYC students' performance on these exams has been relatively stable since 2013 and similar to other large cities, with some increase in reading scores in the 2022 test results. Table 1 displays NYC children's testing results with comparisons to 25 other large U.S. school systems, which were located in urban areas of populations exceeding 250,00; these districts consented to participate and included such cities as Baltimore, Dallas, Los Angeles, Chicago, and Philadelphia.

Table 1:
NAEP Exams - Percent of New York City Students' Scoring Proficient & Advanced 2013 -2022 Calendar Years

Year	4th grade	8th Grade	Other Large Cities
2013	28%	25%	26%
2015	26%	26%	27%
2017	29%	28%	27%
2019	27%	27%	26%
2022	30%	32%	26%

The New York State Education Department's (NYSED) ELA exams are required each spring of all children in grades 3-8 throughout the state. In 2013 NYSED revised the content of its examinations to reflect the Common Core Learning Standards, and in 2018 it rescaled the exams to account for a change in test administration from 3 days to 2. The exam consists of multiple-choice items eliciting inferential comprehension with short and extended constructed response items requiring students to synthesize, evaluate, and provide evidence of their thinking when reading (NYSED, 2023). As with the NAEP scores, NYC children's performance on the NYS ELA examinations has been largely constant since the rescaling in 2018. Results revealed that 46.7%, 47.4% and 49% of NYC children were considered proficient or advanced in reading according in the 2018,

2019 and 2022 testing years. However, when children's performance was considered by borough, results revealed that Bronx children scored about half as well as children in the other boroughs. Specifically, only 24.2% of Bronx children scored proficient in 2018, 26.8**%** in 2019 and 33.2% in 2022. Table 2 displays NYC children's performance by borough on the NYS ELA examination.

Table 2:
NYS ELA Exams - Percent of New York City Student Scoring at or Above Proficient 2018-2022

Year	Borough	Percent at or Above Proficient
2018	Bronx	24.2
2018	Brooklyn	47.7
2018	Manhattan	55.2
2018	Staten Island	53.6
2018	Queens	52.1
2019	Bronx	26.8
2019	Brooklyn	49.8
2019	Manhattan	56.5
2019	Staten Island	53.3
2019	Queens	52.1
2022	Bronx	33.2
2022	Brooklyn	51.4
2022	Manhattan	57.4
2022	Staten Island	55.3
2022	Queens	53.8

An Inquiry into Reading Program Efficacy

We investigated the empirical evidence available to justify the change from balanced literacy to the Great Minds, the Houghton-Mifflin-Harcourt, and the EL reading programs. Our underlying goal was to determine whether empirical evidence existed for adopting the newly adopted reading programs, and whether there was substantive data indicating that these programs would improve NYC children's ability to read. Given the national movement toward evidence-based

practices, we wondered what data existed for selecting these three reading programs to replace balanced literacy. Our guiding questions were the following:

- Is there empirical evidence confirming the effectiveness of the new reading programs for an urban school system?
- What is the nature of the evidence that publishers use to confirm program efficacy of their reading programs?

Method of Analysis

Our analyses consisted of review of the publishers' website data pertaining to the effectiveness of their reading programs. We accessed information available on the three publishers' websites: EL Education's *Expeditionary Learning*, Great Minds (2023) *Wit and Wisdom* and Houghton Mifflin Harcourt's (2023) *Into Reading*. The publishers' websites were used because we believed this is where the companies would certainly present their best evidence regarding the efficacy of their program's reading methods. Although other information sources for selecting reading programs, such as consultations with publishers' representatives, observational visits to schools using their programs, or trial use of program materials could occur in other contexts, we assumed the publishers would champion their program effectiveness on their websites, if such evidence were available.

Previous discussions in the literature regarding the use of research evidence informed our analyses. Essays by Duke and Martin (2011) and Shanahan (2020a) examined the kinds of research that can be legitimately used when making comparisons of program effectiveness, and their ideas informed our thinking. These researchers argued there are many kinds of research that can inform one's understanding about the processes involved in learning to read. For example, surveys, case studies, correlational and observational studies, as well as others, can all be helpful in improving one's understanding of the characteristics and benefits of particular methods of teaching reading. These kinds of research studies could certainly inform and offer insight into one's understanding of a particular method of teaching reading. However, terms such as "research-tested" or "research-proven" require more rigorous use of methods of study.

Duke and Martin (2011) argue that for a method to be considered research-tested or research-proven, the program needs to have been tested and compared with other teaching methods. They explain that when making program comparisons the research must consist of experimental or quasi-experimental methods with students randomly assigned or matched to treatment and control groups. Without use of comparison groups, one cannot argue that one particular reading program is better than others at helping children learn to read.

Publishing companies sometimes assert that their methods of teaching are research-based. However, the phrase "research-based" is vague and insufficiently precise for claims arguing that one teaching method was better than others (Shanahan, 2020a). Consequently, we scrutinized the publishers' websites to determine if their reading program's effectiveness were assessed against other programs involving similar student populations. Absent this type of research design, it is not justifiable to claim superiority of one teaching method over others (Duke & Martin, 2011; Shanahan, 2020a).

The three publishers' websites served as our data source for evaluating the reported evidence regarding the effectiveness of each of the reading programs. To do this, we adapted Shanahan's (2020a) recommendations for assessing the quality and applicability of research results to one's teaching needs. Shanahan identified ten questions or criteria for determining the quality of a study, and we selected three that were most salient for our purposes. These three criteria would provide sufficient initial evidence about the efficacy of the publishers' program evidence.

The first Shanahan criterion is that of peer review and publication; peer review requires anonymous examination of a study's merits by outside reviewers before acceptance and publication into a journal. Publication in a peer-reviewed journal seemed a reasonable first criterion for discerning the quality of evidence regarding program effectiveness. Our assumption was that publishers would forefront such evidence on their webpages, if it were available.

Shanahan's second criterion pertained to having matched clinical trials for documenting effectiveness. That is, the reading program needs to have been compared with other programs having similar student populations. This second criterion is essential because certain kinds of research methods, such as case studies, surveys and correlations, although valuable in their own right, cannot be used as empirical evidence that one program was better than others. For a program to be described as better than others, the program needs to be compared with other methods of teaching where similar student populations and resources were available.

The third criterion from Shanahan that we applied in our analysis focused on replication of research results. This involved scrutinizing the publishers' evidence to see if the outcomes of their research studies were consistent across various environments. The significance of replication stems from the observation that students might excel with a new program simply because of the freshness of the learning approach, but not necessarily because of the method being used. Or teachers might exhibit increased effectiveness because of their enthusiasm for being selected to implement a new reading program, but not because of the actual efficacy of the teaching methods. If the findings of a study did not recur in other contexts and with other student populations, then it could not be justifiable to

claim that the reading program would be equally successful elsewhere. Concrete, empirical proof that the programs were examined in multiple school settings would satisfy this criterion of replication.

In sum, our three criteria for determining the quality of research evidence regarding the three reading programs were that of publication in professional journals, use of matched clinical trials and replication of results in multiple settings. This method, we believed, created an effective and sufficiently robust framework for assessing the quality of the evidence about the three reading programs. With evidence-based decision-making becoming a standard practice nationwide, we were eager to determine if there was empirical evidence supporting the effectiveness of these three reading programs that the city schools adopted.

Results

The Great Minds Program

The Great Minds' website indicates that its program's teaching practices are grounded in "research-based" practices. Its online resource page offers a "Knowledge Hub" where studies of program quality and effectiveness are available. The "Knowledge Hub" includes multiple case studies of teachers' program experiences, descriptions of teachers' program fidelity, and descriptions of children's test score progress from one academic year to the next. The website posted one quasi-experimental study comparing "Wit and Wisdom" students with a national data source of similar students and schools. This study was conducted by an outside educational group, which reported modest results indicating that in the first year of implementation Wit and Wisdom "…had a positive effect on students' early literacy outcomes and a positive effect on upper-elementary students' state reading test scores, on average. (Institute for Educational Policy, 2021, p.7). Our analysis of the Great Minds' website revealed that a program comparison with a national data bank occurred, the statistical results were positive but modest. The study was not replicated or published in a peer-reviewed journal.

The Houghton-Mifflin-Harcourt Program

Houghton-Mifflin-Harcourt (HMH)'s website presents a comprehensive perspective on the research about teaching elementary reading, including the five elements from the NRP report. In addition, HMH's program description provides detailed resources about SOR, including links to various resources explaining how its "Into Reading" integrates SOR research.

The HMH website indicates that a study of the effectiveness of "Into Reading" was undertaken, but it was discontinued because of the pandemic. Our analysis indicated that the HMH's study used a matched design of students

involving non-participating schools. The HMH website indicates that a new study is being planned.

EL Education

EL Education's website presents summaries of four clinical trials of its reading language arts program. These research studies involved schools in Detroit, New York City, Washington, D.C., Tennessee, and in Rochester, N.Y. All of the studies were conducted by outside evaluation group. The EL website summarizes a 2011 study occurring over a two-year period with participating middle schools in Rochester and New York City. Children's test performance in the participating schools were compared with control schools on the NYSED English Language Arts exam. The EL website reports that the treatment and control group were matched by income and percentage of English language learners. Results indicated that treatment children outperformed the control groups. The website states "…the effect of being in an EL Education school was positive, substantial, and significant ($p < .01$)".

A more recent study (Dolfin et al. 2019), which also appears on its website, took place in Washington, D.C. and New York City. The "Teacher Potential Project" involved control and treatment groups in five middle schools with students matched economically and by achievement; a related component of the study involved teachers' profession development with the EL curriculum. The EL website reports: "…EL Education students experience positive impacts that are roughly equal in magnitude to an extra five months of learning growth after two years and an extra seven months of learning after three years."

Summary of Our Analyses

Table 3 displays the summary of our analyses of the reading programs according to the three criteria used in our analysis. Only one of the reading programs (EL Education) used clinical trials with comparative groups, and its studies are available through its website.

Table 3:
Comparing the Empirical Evidence of the Three Reading Programs

Reading Program	Comparison Group	Peer-reviewed Replication

Wit and Wisdom	Yes	No	
No			
Into Reading	No	No	
No			
Expeditionary Learning	Yes	No	
Yes			

Discussion

At a time when evidence-based decision-making has accelerated in importance throughout education, our examination of the three publishers' websites revealed only one of the companies produced empirical data of program's effectiveness. EL Education's website showcased four comparative studies revealing students in its reading programs outperformed their counterparts in other schools where different reading programs were used. The Great Minds' website presented multiple case studies attesting to the quality of its program, and it featured a link to one year-long clinical trial with a comparison group; however, the study results were slight and not replicated. Lastly, Houghton-Mifflin-Harcourt's website reported a large-scale study that was halted due to the pandemic, although there were indications that the study would resume.

We suspect that decision by the NYC schools to replace balanced literacy with the three reading programs was likely driven by a variety of forces: (1) Although children's performance on the NAEP and statewide reading tests have been relatively constant for the last decade, the school system's leaders were dissatisfied with the lack of progress in children's reading achievement. (2) We inferred that the NYCDOE now held theoretical models of learning to read that differed from balanced literacy, and this difference pushed program change in the city schools. Such a change would be consistent with national criticism of balanced literacy programs, especially as voiced through "Science of Reading" advocates (e.g., Hanford, 2022, Seidenberg, 2017), and these criticisms likely accelerated change, including in New York State (Ashford, 2024). (3) The lack of progress in children's reading achievement reflected badly on the city school

system. Moreover, the disparity in test results between wealthy and low-income boroughs and neighborhoods suggested inequitable opportunities for children's learning to read. Such dissatisfaction was likely fueled by test scores in the Bronx where only about half of its children performed as well as students living in the other boroughs. Given these issues of stagnate test scores, change in theoretical models about teaching children to read, and the inequalities in children's reading achievement, program change in the city schools became inevitable.

The current debate about effective methods of teaching beginning reading remains a critical issue in research. The basic argument is that, similar to the medical field, teaching practices should be validated through clinical trials to be deemed effective. This stance challenges often vague claims that teaching methods are "research-informed" or "research-based." Such terms, although widely employed in professional literature, lack precision because we know that any theory or practice can be loosely viewed as beneficial for informing teaching practices. The results of our inquiry highlight a significant gap in the selection of the reading programs for the NYC school system. That is, aside from EL Education, the other two reading programs lacked rigorous evidence about the efficacy of their curricula. Without such empirical evidence, claims that the newly adopted reading programs will produce greater reading achievement than balanced literacy remain unsubstantiated.

There is significant research indicating that the most important influence in learning to read, excluding the home, is the quality of teaching children receive (e.g., Cardichon, et al., 2020; Darling-Hammond, 2000 & 2009). Or stated more succinctly, reading programs do not teach - teachers do. Although program design might advance learning, teachers have the greatest impact on children's success in reading. Importantly, there is no experimental research indicating that the three reading programs selected for the city's school system would be more effective than the balanced literacy program that has been previously used.

We recognize some of the limitations of our analyses. We did not interview school officials about the rationale for changing reading programs in the city schools. Although the decision seemed abrupt to us, we have not yet had the opportunity to learn about the processes the school system used to change reading programs. We acknowledge that school leaders have the right, as well as responsibility, to lead the system in the direction that it believes will best help the children it serves. Although a program change might have been merited, little empirical evidence was provided to the public supporting the adoption of the new reading programs.

In today's world, learning to read is significantly shaped by social and cultural forces that, we believe, are more intense than those faced by previous generations of children. These forces include the shift from paper to digital texts, the growing number of families living in poverty that makes access to books at

home more difficult than a decade ago, and the widespread use of multimedia texts as sources of information, as opposed to traditional print. Additionally, there is a cultural trend towards seeking information in short, abbreviated texts on social media platforms like TikTok, Facebook, and Instagram, rather than reading longer texts found in newspapers and books. The ongoing effects of the pandemic have certainly impacted children's social, emotional, and cognitive development, which are crucial in learning to read. All of these forces have likely interfered with children's learning to read and affected their test performance.

For children to succeed in today's world, they must receive high quality literacy instruction that offers varied and rich opportunities to learn to read and write. However well-intentioned, overly simplified methods of teaching reading lacking empirical evidence of effectiveness are unlikely to accomplish this important goal.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.

Apple, M. W., Biesta, G., Bright, D., Giroux, H. A., Heffernan, A., McLaren, P., Riddle, S., & Yeatman, A. (2022). Reflections on contemporary challenges and possibilities for democracy and education. *Journal of Educational Administration and History, 54*(3), 245–262. https://doi.org/10.1080/00220620.2022.2052029

Ashford, G. (2024**,** January 3). As literacy lags, Hochul proposes changing how schools teach reading. The New York Times. https://www.nytimes.com/2024/01/03/nyregion/new-york-schools-reading-hochul.html?smid=url-share

Baron, J. (2018). A brief history of evidence-based policy. *The Annals of the American Academy*, July, 40-50. https://doi.org/10.1177/0002716218763128

Bond, G.L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, *2*(4), 5–142. https://doi.org/10.2307/746948

Cardichon, J., Darling-Hammond, L., Yang, M., Scott, C., Shields, P. M., & Burns, D. (2020). *Inequitable opportunity to learn: Student access to certified and experienced teachers*.: Learning Policy Institute.

Cervetti, G., Pearson, P.D., Palincsar, A., Afflerbach, P., Kendeou, P., Biancarosa, G., Higgs, J., Fitzgerald, M. & Berman, A. (2020). How the reading for understanding initiative's research complicates the simple view of reading invoked in the science of reading. *Reading Research Quarterly*, *55*(1), 161-172. https://doi.org/10.1002/rrq.343

Chall, J. (1967). *Learning to read: The great debate*. McGraw-Hill.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. Macmillan.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, *8*(1), 1-44. https://doi.org/10.14507/epaa.v8n1.2000

Darling-Hammond, L. (2009). Recognizing and enhancing teacher effectiveness. *The International Journal of Educational and Psychological Assessment*, 3(December), 1-25. https://doi.org/10.14507/epaa.v8n1.2000

Dolfin, S., Richman, S., Choi, J. Streke, A., DeSaw, C., Demers, A. & Poznyak, D. (1999). *Evolution of the Teacher Potential Project*. Washington: Mathematica.

Duke, N. & Cartwright, K. (2021). The science of reading progresses:

Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, *56*(1), 25-44. https://doi.org/10.1002/rrq.411

Duke, N. & Martin, N. (2011). 10 things every literacy educator should know About research. *The Reading Teacher*, *66*(1), 9-22.

Farrell, L., Hunter, M. Davidson, M. & Osenga, T. (2019). The simple view of reading. *Reading Rockets*. https://www.readingrockets.org/topics/about-reading/articles/simple-view-reading

Fien, H., Chard, D. J., & Baker, S. K. (2021). Can the evidence revolution and multi-tiered systems of support improve education equity and reading achievement? *Reading Research Quarterly*, *56*(1), S105–S118. https://doi.org/10.1002/rrq.391

Flesch, R. (1955). *Why Johnny can't read*. Harper & Row.

Freire, P. (1970). *Cultural action for freedom*. Harvard University Press.

Goldenberg, C. (2020). Reading wars, reading science, and English learners. *Reading Research Quarterly*, *55*(S1), 131–144. https://doi.org/10.1002/rrq.340

Goodlad, G., Mantle-Bromley, C & Goodlad, S. (2004). *Education for everyone: Agenda for education in a democracy.* Jossey-Bass.

Goodman, K. S. (1967). A linguistic study of cues and miscues in reading. *Elementary English*, *42*, 639-643.

Goodman, Y. M., & Burke, C. L. (1970). *Reading miscue inventory manual procedure for diagnosis and evaluation*. Macmillan.

Great Minds (2023). *Wit and Wisdom*. https://greatminds.org/the-facts-about-wit wisdom-and-its-impact

Hanford, E. (2022). School is for Learning to Read. *The New York Times*. Retrieved from https://www.nytimes.com/2022/09/01/opinion/us-school-reading.html?mwgrp=c-dbar&smid=url-share

Harris, K., Graham, S., Aiken, A., Barkel, A., Huston, J. & Ray, A. (2017). Teaching spelling, writing and reading for writing: Powerful evidenced-based practices. *Teaching Exceptional Children*, 49(4), 262-272. https://doi.org/10.1177/00400599176972

Hoffman, J., Hikida, M. & Sailors, M. (2020). Amplifying equity, agency and design research in literacy teacher preparation. *Reading Research Quarterly, 55*(1), 255–S266. https://doi.org/10.1002/rrq.353

Houghton-Mifflin-Harcourt (2023). *Into Reading*. https://www.hmhco.com/programs/into-reading

Institute for Educational Policy (2021). *Implementing Wit and Wisdom: An Evaluation Research Study*. John Hopkins University.

Kristoff, N. (2023). Two-thirds of kids struggle to learn to read, and we know

how to fix it. New York Times, Feb 11. https://www.nytimes.com/2023/02/11/opinion/reading-kids-phonics.html

Lewis, K., McColskey, W., Anderson, K., Bowling, T., Dufford-Melendez, K., and Wynn, L. (2007). *Evidence-based decision-making: assessing reading across the curriculum interventions* (Issues & Answers Report, REL 2007–No. 003). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. U.S. Department of Education. https://www2.ed.gov/rschstat/research/pubs/accountable/accountable.pdf

National Reading Panel. (2000). *Teaching children to read: An evidence- based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/documents/report.pdf

Neuman, S. (2016). Code red: The danger of data-driven instruction. *Educational Leadership*, Nov., 24-29.

New York State Education Department (2023). *Educator Guide to the 2023 Grades 3–8 English Language Arts Tests*. Albany, N.Y. https://www.nysed.gov/sites/default/files/programs/state-assessment/38-ela-educator-guide-2023.pdf

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002). http://www. ed.gov/policy/elsec/leg/esea02/index.html

Pearson, P.D. (2004). The reading wars. *Educational Policy*, *18*(1), 216– 252. http:// 10.1177/0895904803260041.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. The Guilford Press.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health and Human Development.

Rasinski, T., Homan, S., & Biggs, M. (2009). Teaching reading fluency to struggling readers: Method, materials, and evidence. *Reading & Writing Quarterly*, *25*(2-3), 192-204. http://doi:10.1080/10573560802683622

Roehling, J., Hebert, M., Nelson, J. R., & Bohaty, J. (2017). Text structure

strategies for improving expository reading comprehension. *The Reading Teacher, 71*, 71-82. http://doi:10.1002/trtr.1590
Seidenberg, M. (2017). *Language at the speed of light*. Basic Books.
Shanahan, T. (2020a). The science of reading: Making sense of research. *The Reading Teacher*, *74*(2), 119-125. https://doi.org/10.1002/trtr.1940
Shanahan, T. (2020b). What constitutes a science of reading instruction? *Reading Research Quarterly*, *55*(S1), S235–S247. https://doi.org/10.1002/rrq.349
Shannon, P. (1983). The use of commercial reading programs in American elementary schools. *Reading Research Quarterly*, *19*(1), 68-85.
Slavin, R. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15-21. https://doi.org/10.3102/0013189X031007015
Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read*. Holt, Rinehart & Winston.
Smith, N.B. (2002). *American reading instruction (Special Edition).* International Reading Association.
Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in Young children*. National Academy Press.
Stein, M.K., & D'Amico, L. (2002). Inquiry at the crossroads of policy and learning: A study of district-wide literacy initiative. *Teachers College Record*, *104*(7). 1313-1344. https://doi.org/10.1111/1467-9620.
Taylor, B.M., Pearson, P.D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary grade reading instruction in low-income schools. *The Elementary School Journal*, *101*, 121–165.
Teale, W., & Gambrell, L. (2007). Raising urban students' literacy. *Reading Teacher*, *60*, 728–739. https://doi.org/10.1598/RT.60.8.3
Terry, N. (2020). Delivering on the promise of the science of reading for all children. *The Reading Teacher*, *75*(1), 83-90. https://doi.org/10.1002/trtr.2031
Thomas, P.L. (2022). The Science of Reading movement: The never-ending debate and the need for a different approach to reading instruction. Boulder, CO: National Education Policy Center. Retrieved [December 22, 2023]. http://nepc.colorado.edu/publication/science-of-reading
U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, *National Assessment of Educational Progress* (NAEP), 2022 Reading Assessment**.** https://nces.ed.gov/nationsreportcard/about/
Vaughn, M., Scales, R., Stevens, E., Kline, S., Barret-Tatum, J., Van Wig, A., Yoder, K. & Wellman, D. (2021). Understanding literacy adoption policies across contexts: A multi-state examination of literacy curriculum decision-

making. *Journal of Curriculum Studies*, 53(3), 333-352. https://doi.org/10.1080/00220272.2019.1683233

Walker-Dalhouse & Risko, V. (2008). Learning from literacy successes in high-achieving urban schools. *The Reading Teacher, 61*(5), 422–424.

Zimmerman, A. (2023). Hundreds of NYC Elementary schools used a Teachers College Reading curriculum Banks said 'Has not worked.' *Chalkbeat*, February 12. https://www.thecity.nyc/2023/2/14/23598696/nyc-teachers-college-lucy-calkins-balanced-literacy-david-banks