Building a Stronger Case for Independent Reading at School

Adriana G. Bus (iD)

University of Stavanger
ELTE Eötvös Loránd University

Yi Shang
Kathleen Roskos

John Carroll University

The effectiveness of incorporating independent reading practice in schools has long been a subject of uncertainty. To shed light on this ongoing debate, this meta-analysis seeks to investigate the impact of in-school independent reading on three crucial measures—attitudes toward reading, word recognition, and comprehension—focusing on K–10 students. The analysis encompasses (quasi-)experimental studies conducted between 1970 and 2020, examining a total of 7,493 students across 47 studies. Because most studies contain more than one outcome measure or effect size, we used a meta-analytic model with a three-level structure. The findings reveal a statistically significant overall effect size (Hedges' g = .08). Specifically, the effect sizes are more pronounced when considering word recognition (i.e., word attack, word identification, decoding, and fluency; Hedges' g = .21) and students' reading attitude (Hedges' g = .18) as outcome measures. However, the effect size for comprehension—the most commonly assessed outcome measure—was approximately zero (Hedges' g = −.014).

Keywords: *independent reading, three-level meta-analytic model, word recognition, reading attitudes, text comprehension, primary and secondary education*

Independent reading as a form of reading practice in schools gained recognition in the early 1970s and has been known by various terms such as *free voluntary reading, independent silent reading, supported independent reading, sustained silent reading, uninterrupted sustained silent reading*, and *drop everything and read* (DEAR), among others. Ongoing discussions have centered on the importance of integrating independent reading into the school curriculum during instructional hours. Advocates for IR strongly advocate for its inclusion, underlining the necessity for IR by proposing that a portion of each school day, typically around 15 minutes, should be devoted to what some have termed *free voluntary reading* (e.g., Krashen, 2004). By "free," they refer to independent reading that entails minimal accountability, without the requirement for book reports or strict limitations on reading materials. The objective is to cultivate an environment that motivates students to engage in reading activities while at school, reflecting how they typically interact with reading outside the classroom.

Debates Surrounding IR

One topic in debates surrounding IR (see, e.g., Pennington, 2011) is the control over students' reading activities. Proponents of IR typically advocate for providing students with access to a classroom library and encouraging them to read self-selected materials, eschewing tests or other mechanisms such as leveled books (e.g., Fielding et al., 1986; Krashen, 2004; Morrow, 1983). Their argument is that this approach fosters reading engagement and complements regular reading instruction. However, many educators argue against unrestricted IR, believing that guided or restricted practices, such as offering tutorial support or limiting book choices, may lead to better outcomes (e.g., Gambrell, 2007;

Kelley & Clausen-Grace, 2006; Reutzel et al., 2008, 2010; Shanahan, 2018; Topping et al., 2007; Weber, 2018).

Another aspect of the discussion surrounding IR is the concern that it may not constitute proper teaching and could place excessive responsibility on students for their reading development. However, proponents of IR argue that this concern stems from a misunderstanding of its purpose (see, e.g., Pennington, 2011). IR is not intended to replace a comprehensive reading program, whether for beginning or more advanced readers. Typically, IR involves only a short daily period of about 15 minutes, which may not allow sufficient time to practice all the necessary reading strategies. Nonetheless, proponents believe that by fostering essential abilities and attitudes for increasing reading volume, IR may contribute significantly to reading development. These abilities and attitudes include the motivation to choose books voluntarily, the discovery of inherent interest and excitement in reading material, and the cultivation of sustained concentration on text (Bryan et al., 2003; Merga, 2015; Van der Sande et al., 2022). Hence, IR might be particularly significant for students lacking external incentives to nurture these skills and attitudes beyond the school environment (Gottfried et al., 2003; Mullis et al., 2012; Willingham, 2015).

Rationales for IR

A primary objective of IR is to enhance reading volume, either directly through daily reading or indirectly by fostering interest in reading. Reading volume is considered a fundamental principle of the science of reading, encompassing a wide array of research on effective literacy learning strategies. Research consistently indicates that the quantity of reading material consumed plays a pivotal role in reading development (Seidenberg, 2017). Even novice readers who engage extensively in reading demonstrate notable advancements in literacy and language skills compared with their less engaged counterparts (Hiebert, 2024). Earlier studies by Stanovich and colleagues have illustrated that increased reading volume leads to an upward spiral in reading proficiency: as reading volume expands, avid readers exhibit enhanced knowledge in literature and history (Stanovich & Cunningham, 1992) and demonstrate elevated levels of "cultural literacy" (West et al., 1993), both of which are essential for fostering a deeper enjoyment of new texts and, consequently, further increasing reading volume.

A recent report on adolescents' reading performance by the Organisation for Economic Co-operation and Development (OECD) underscores the importance of in-school activities that ignite students' intrinsic interest and enthusiasm for reading. The Programme for International Student Assessment (PISA) 2022 results across 81 countries unveiled a noteworthy decline in the average reading scores, marking the first recorded instance of a decrease of up to 10 points compared with PISA 2018. This decline, surpassing any previous drops, suggests a long-standing trend predating the COVID-19 pandemic. One potential explanation is the noticeable decrease in the amount of in-depth reading of longer texts compared with superficial skimming of brief messages among newer generations (e.g., Wolf, 2018). Despite findings from numerous earlier surveys indicating that many students did not frequently choose to read (e.g., Anderson et al., 1988), we now seem faced with an even more serious decline in reading volume.

Effects of IR

Does IR in schools truly contribute not only to essential language and literacy development but also to fostering inherent interest and passion for reading, consequently influencing future reading abilities? Several studies have compared the progress of students engaged in language arts programs with and without IR supplements, primarily using standardized tests to evaluate word recognition, reading comprehension, and vocabulary (e.g., Pressley et al., 2002). According to Krashen (2004), the predominant observation, with a few exceptions, is that there are significant disparities in reading progress between the two approaches, favoring IR supplements. He notes that studies reporting limited effects of IR often have short durations. Brief periods as short as 2 months may not provide ample time for students to obtain interesting reading materials and fully engage in the reading process.

Indirect quantitative evidence comes from an analysis of predictors for achievement on the Progress in International Reading Literacy Study (PIRLS) reading test, administered to 10-year-olds in more than 40 countries, with a focus on reading in their native language (Krashen et al., 2012). The findings reveal that the presence of a school library with a minimum of 500 books emerged as a robust positive predictor for time spent on reading. Similarly, a Dutch experiment found that schools equipped with libraries containing a substantial number of books, specifically at least five books per student, create an optimal setting for fostering reading proficiency (e.g., Nielen & Bus, 2015).

Additionally, there is evidence supporting the hypothesis that a school promoting free voluntary reading fosters a propensity for engaging in more extensive reading in the future. For instance, a qualitative study by Ivey and Johnston (2013) that investigated students' responses to self-selected, self-paced reading of compelling young-adult literature in classrooms reported increased purposeful and extended absorption in books. The findings highlighted a strong sense of agency regarding their reading, because students pushed themselves to their limits and deliberately used available scaffolds, particularly from peers, when encountering difficulty.

Reviews on IR

However, the outcomes of reviews conducted on IR offer a varied viewpoint. In an early research review by Wiesendanger and Birlem (1984), the effect of IR on word recognition and reading comprehension was inconclusive in the eight studies testing such effects. However, it did demonstrate a positive impact on attitudes toward reading, aligning with the anticipated outcomes of engaging in daily 15-minute sessions of IR. Of the 11 studies, nine reported a positive effect on attitudes toward reading. This suggests that the additional value of IR lies in students' appreciation and enjoyment of reading, indicating that IR provides an opportunity for students to explore interesting books and discover how reading can align with their interests. Therefore, a significant impact of IR may be its influence on students' attitude toward reading and, consequently, their engagement in voluntary reading outside of school that will contribute to the further development of their literacy and language skills.

Subsequent research syntheses focused only on reading skills instead of also including reading attitude and behavior. The widely cited report of the National Reading Panel (NRP, 2000) specifically examined the impact of IR on reading fluency. The researchers reached the conclusion that the 14 studies comparing students who participated in IR with those who did not were inconclusive due to significant flaws. They encountered challenges in identifying enough IR studies that met their rigorous criteria for experimental research, which ultimately prevented them from determining an overall effect size. They determined that there is a lack of adequate data from well-designed studies capable of investigating causation, thus preventing the substantiation of causal claims. Garan and DeVoogd (2009) suggest that they might have found more studies had they focused on a broader set of reading achievements instead of fluency as the primary outcome measure. The strict criteria regarding experimental design received criticism. Garan and DeVoogd suggested that if medical science had applied a similarly rigorous methodology, it could have hindered the determination that smoking poses a threat to our well-being.

In a quantitative meta-analysis conducted by Yoon (2003) that examined 10 studies specifically investigating the influence of IR, an average effect size greater than zero was observed, specifically .11 (standard error = .04). This study focused primarily on reading comprehension rather than fluency. According to the classic benchmarks established by Cohen (1962) based on tightly controlled laboratory experiments, this effect size is considered small. However, Kraft (2020), noting that experiments conducted in school settings using general tests rather than intervention-specific measures tend to yield much smaller effect sizes than researcher-controlled laboratory experiments, argues that effect sizes between .05 and .20, classified as small by Cohen's criteria, actually may be large and meaningful in the context of educational interventions. Considering Yoon's reported effect size of .11 in this perspective, it can be seen as quite substantial.

Despite Yoon's findings, the use of IR decreased for some time following release of the NRP report in 2000, which was critical of the impact of IR. Nevertheless, in recent years, there has been a resurgence of IR in schools, with popular initiatives such as Accelerated Reader and equivalent programs being implemented (Hiebert et al., 2014). Consequently, new reviews are conducted to assess the continued validity of the NRP's findings. Erbeli and Rice (2021) conducted a review using similar selection criteria regarding the studies' design as the NRP but included a range of reading outcome measures, aiming to assess the impact IR can have on children's reading achievement. The authors concluded that many studies did not yield statistically significant results, likely due to the overall small sample sizes. However, the forest plots in their publication do suggest a positive trend, although this was not analyzed with meta-analytic tools. Meta-analysis enables a quantitative synthesis of accumulated studies, facilitating the combination of effects and the identification of trends.

This is precisely why we undertook another quantitative meta-analysis. Especially when dealing with studies characterized by limited sample sizes, as seen in many studies that test the effects of IR, a quantitative synthesis would significantly enhance the comprehensiveness of the analysis (Bus et al., 2021). Because print exposure correlates with various technical reading skills, language proficiency, passage comprehension, writing, and attitudes toward reading (Mol & Bus, 2011), we considered all as potential outcome measures. However, unlike previous reviews, we rigorously selected (quasi-)experiments that specifically emphasize IR as a brief daily activity lasting approximately 15 minutes. Our focus was on cultivating unrestricted free reading experiences within the school context as opposed to at home or during the summer (e.g., Kim & White, 2008). Additionally, we excluded studies that use free reading to promote learning English as a second language, even though we acknowledge its potential efficacy (e.g., Matsui & Noro, 2010).

This Meta-analysis

This meta-analysis aims to investigate the contribution of IR to students' reading development and its mechanisms. It is recognized that IR may offer valuable practice for reading skills and that the time spent on it may be as effective as explicit reading instruction during the same timeframe. Therefore, we anticipate examining the effectiveness of IR as a pedagogic strategy for enhancing reading proficiency, encompassing aspects such as word recognition and comprehension. However, of particular interest is its potential impact on cultivating an appreciation for reading, fostering

enjoyment in reading books, and promoting voluntary reading outside of school. The time spent engaging in IR at school is expected to facilitate the development of a positive attitude that nurtures students' interest in reading.

Our analysis focuses on experimental and quasi-experimental studies conducted within the last 50 years, aiming to contribute to the growing body of scientific knowledge on this subject. We are particularly interested in exploring the effectiveness of unconstrained IR, as advocated by Krashen and other proponents of IR. This approach involves *in-school free reading*—self-selected and devoid of external requirements. Our hypothesis is that this type of "pure" IR, characterized by dedicated time for self-directed silent reading practice within the school environment, can yield benefits. Through our meta-analysis, we seek to elucidate the advantages and effectiveness of IR, comparing it with more constrained reading approaches, while acknowledging the indispensable role of systematic teaching.

The objective of this meta-analysis is to examine and evaluate the following hypotheses pertaining to the effects of IR:

1. The first hypothesis is that providing students with authentic reading opportunities during school hours, even for brief periods (~15 minutes per day), without imposing constraints such as accountability, book reports, or strict limitations on reading materials can have a positive effect on their reading development.

2. The second hypothesis delves into the specific aspects of reading development influenced by IR. It is reasonable to assume that technical reading skills, such as word recognition and fluency, may improve with regular reading practice facilitated by IR. Additionally, while comprehension skills typically require a substantial amount of print exposure that may not be fully achievable during IR time alone, IR may indirectly contribute to comprehension through enhanced word recognition. However, the most significant expected outcome lies in the impact of IR on reading attitudes. The opportunity to experience the pleasure of reading and understand its social and emotional significance in everyday life is anticipated to have a substantial effect on students' attitudes toward reading.

3. The third hypothesis proposes that the effects of IR, specifically on reading proficiency, may vary depending on the control condition. The impact of IR on reading proficiency, especially regarding word-recognition skills (such as decoding, sight words, and fluency), may be influenced by the reading experiences provided in the control condition. If students in the control condition receive teacher-led supplemental reading skills practice during the same period, we might expect no effect or even a reduced effect of IR on reading skills.

However, if IR is genuinely additional and students in the control condition are not engaged in reading activities during the designated IR time, IR will outperform the control condition.

Method

Search Strategy

We searched several bibliographic databases, including Academic Search Complete, Education Research Complete, Education Full Text, Education Resources Information Center, Professional Development Collection, Web of Science, PsychINFO, and PubMed. We used the following search terms: "independent reading*," "sustained silent reading," "uninterrupted sustained silent reading," "drop everything and read," "free reading," "pleasure reading," "reading practice," "recreational reading," "R5 (read, relax, reflect, respond, rap)," "million minutes," and "choice reading." Figure 1 presents a flow diagram of study selection. The initial round of search yielded a list of 9,475 citations, including 802 duplicates. A two-tier screening of titles and abstracts reduced the list to 446 and then 83 articles, of which we were able to locate the full texts of 66 articles. Finally, by examining the bibliographies of the selected articles as well as several review articles, including Erbeli and Rice (2021), Krashen (2001), NRP (2000), and Yoon (2003), we were able to add another 80 publications to the pool. With three authors independently reviewing the full texts of the 146 publications, including journal articles, dissertations, conference contributions, and unpublished reports, we selected 29 articles for the meta-analysis based on the following inclusion criteria:

- Focus on K–12 students;
- Involvement in IR at school during class, center, or study hall time;
- IR approach as a brief activity, typically lasting 10–15 minutes per day, conducted in addition to regular reading instruction;
- During IR, teachers having limited influence over students' reading activities, deliberately avoiding tests or other mechanisms such as leveled books;
- A group engaging in IR compared with a control group exposed to various forms of reading instruction or alternative activities not involving reading (the specific nature of these activities may not be clearly defined);
- Including experimental, prospective causal-comparative, and ex post facto studies;
- Outcome measures including proficiency in reading (e.g., word attack, language skills, passage comprehension, and others), writing skills, and reading attitude; and
- The provision of effect sizes or sufficient information to enable their calculation or estimation.
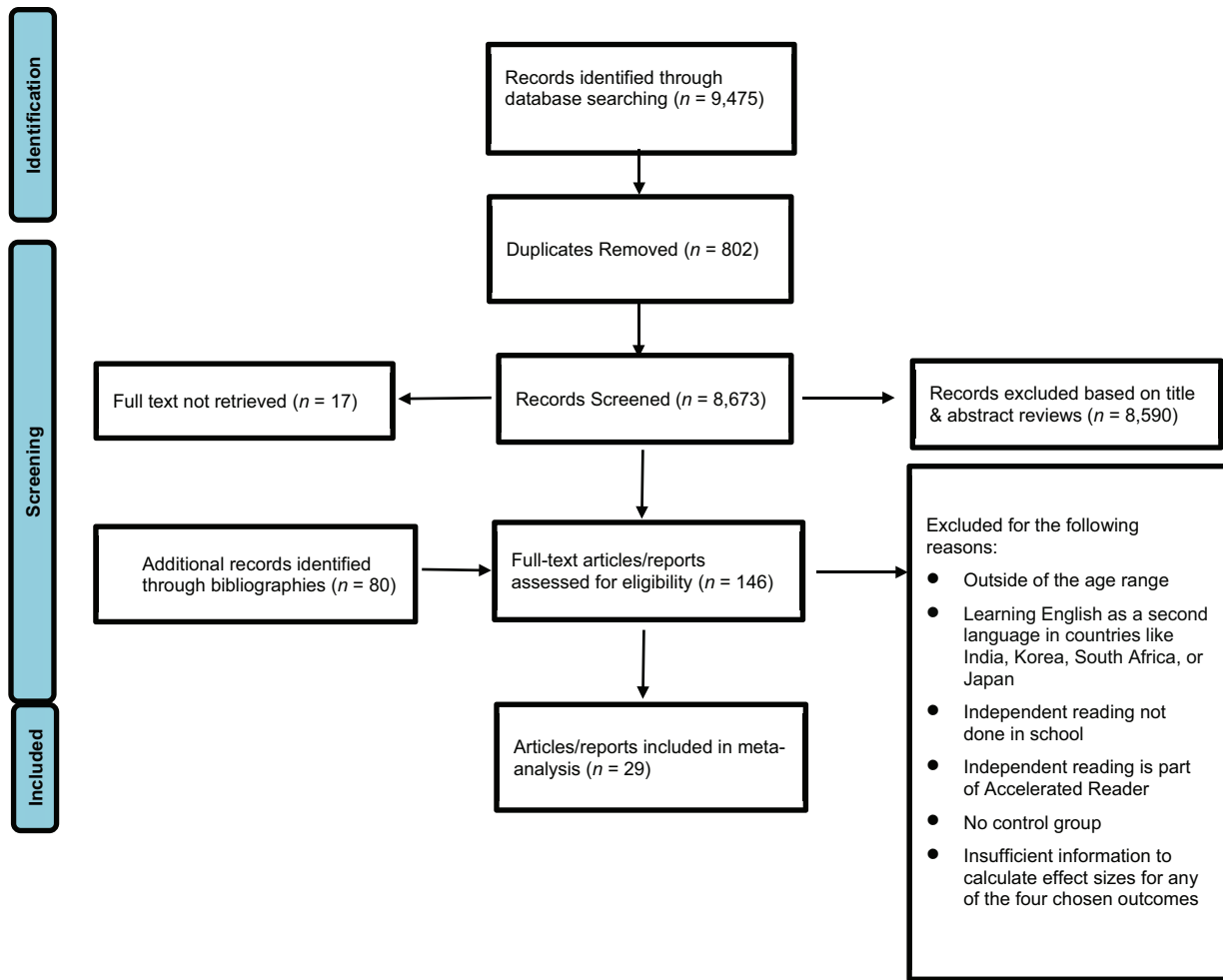
FIGURE 1. *Flow diagram of study selection.*

Thus, we excluded studies that

- focus on free voluntary reading at home or during the summer;
- deviate from the essence of IR programs by imposing limitations on students' book choices, restricting them to those available within the program and tailored to their independent reading levels (as a result, we excluded accelerated reader programs that involve goal setting facilitated by the program or teacher rather than the autonomy of the student and routine assessments such as practice quizzes per book to measure progress); and
- are correlational studies based on surveys or large-scale international assessments such as the Progress in International Reading Literacy Study (PIRLS).

The final pool of 29 primary research articles (47 studies) was published between 1975 and 2019, of which 13 were published after 2000 and eight after 2010. The pool consists of 14 unpublished dissertations, 12 peer-reviewed journal articles, one unpublished report, and two conference contributions (Table 1).

Coding the Studies

The studies encompassed a variety of measures, addressing fundamental reading skills at the word level, writing skills, word and passage comprehension, and evaluations of students' attitudes toward reading. Studies evaluating word-level skills, involving word identification, word attack strategies (e.g., decoding using phonics), and fluency (i.e., measured in words read per minute) were collectively coded as *word recognition*. A restricted number of studies addressed *writing skills*, covering both text composition (Reedy, 1994) and spelling proficiency (Mostow et al., 2013). Although these scores were considered in our examination of overall effects, the pool of studies was insufficient to quantify the impact of IR on this specific outcome.

TABLE 1

Selected Characteristics of Publications Included in the Meta-analysis

Publication	Publication status	Grade	Program	Reading in control group	Duration	Design	Independent reading used as control	*N*	Outcome measures	Hedges' *g*
Allen (2017)	Dissertation	5	IR	Yes	30 min daily, 36 wks	Quasi	No	116	Comprehension	−.26 (.19)
Borjes (2009)	Dissertation	3	ISR	Yes	30 min daily, 8 wks	Quasi	Yes	30	Word recognition	.19 (.37)
									Comprehension	.00 (.37)
Cirucci (2017)	Dissertation	3	ISR	Yes	30 min, 1 session/ week, 8 wks	Quasi	Yes	44	Language	.01 (.30)
									Language	.61 (.31)
Cline & Kretke (1980)	Journal	7–9	SSR	No	Session details unknown; program lasts 3 years	Ex post facto	No	249	Word recognition	.00 (.13)
									Attitude	.25 (.13)
Collins (1980)	Journal	2–6	SSR	No	20 min daily, 15 wks	Quasi	No	220	Comprehension	.07 (.14)
									Word recognition	.17 (.14)
									Language	.28 (.14)
									Attitude	−.18 (.14)
Cuevas (2012)	Journal	10	ISR	Yes	60 min weekly, 14 wks	Experiment	No	117	Comprehension	.61 (.20)
									Attitude	.62 (.19)
Davis 1 (1988)	Journal	8	ISR	Yes	40 min 3 times/ week, 1 year	Experiment	No	22	Comprehension	.99 (.45)
Davis 2 (1988)	Journal	8	ISR	Yes	40 min 3 times/ week, 1 year	Experiment	No	25	Comprehension	.10 (.40)
Evans & Towner (1975)	Journal	4	SSR	Yes	20 min daily, 10 wks	Experiment	No	48	Comprehension	.24 (.29)
Faggella-Luby 1 (2011)	Journal	5	SSR	Yes	30 min daily, 18 wks	Experiment	Yes	17	Comprehension	.15 (.53)
									Comprehension	.31 (.53)
Faggella-Luby 2 (2011)	Journal	5	SSR	Yes	30 min daily, 18 wks	Experiment	Yes	22	Comprehension	−.17 (.55)
									Comprehension	.02 (.55)
Faggella-Luby 3 (2011)	Journal	6	SSR	Yes	30 min daily, 18 wks	Experiment	Yes	19	Comprehension	−.86 (.54)
Faggella-Luby 4 (2011)	Journal	6	SSR	Yes	30 min daily, 18 wks	Experiment	Yes	23	Comprehension	−.74 (.56)
									Comprehension	−1.09 (.57)
Gray (2012)	Dissertation	4	SSR	Yes	20 min daily, 12 wks	Quasi	No	42	Word recognition	.20 (.33)
									Attitude	−.60 (.33)
Harris-Mobley 1 (2015)	Dissertation	6	DEAR	No	20 min daily, 1 year	Ex post facto	No	305	Comprehension	−.35 (.12)
Harris-Mobley 2 (2015)	Dissertation	6	DEAR	No	20 min daily, 1 year	Ex post facto	No	326	Comprehension	−.13 (.11)
Harris-Mobley 3 (2015)	Dissertation	6	DEAR	No	20 min daily, 1 year	Ex post facto	No	310	Comprehension	−.09 (.11)
Harris-Mobley 4 (2015)	Dissertation	7	DEAR	No	20 min daily, 1 year	Ex post facto	No	346	Comprehension	.03 (.11)
Harris-Mobley 5 (2015)	Dissertation	7	DEAR	No	20 min daily, 1 year	Ex post facto	No	312	Comprehension	−.23 (.11)
Harris-Mobley 6 (2015)	Dissertation	7	DEAR	No	20 min daily, 1 year	Ex post facto	No	326	Comprehension	.04 (.11)
Harris-Mobley 7 (2015)	Dissertation	8	DEAR	No	20 min daily, 1 year	Ex post facto	No	325	Comprehension	−.04 (.11)
Harris-Mobley 8 (2015)	Dissertation	8	DEAR	No	20 min daily, 1 year	Ex post facto	No	359	Comprehension	.02 (.11)
Harris-Mobley 9 (2015)	Dissertation	8	DEAR	No	20 min daily, 1 year	Ex post facto	No	310	Comprehension	−.11 (.11)

(continued)

TABLE 1 (CONTINUED)

Publication	Publication status	Grade	Program	Reading in control group	Duration	Design	Independent reading used as control	N	Outcome measures	Hedges' g
Higgens (1981)	Dissertation	5	SSR	No	20 min daily, 6 months	Quasi	No	366/341	Word recognition	.89 (.11)
									Comprehension	.02 (.11)
									Word recognition	−.03 (.11)
									Language	−.06 (.11)
Holt 1 (1988)	Conference	7	SSR	Yes	20 min 3 days/ week, 10 wks	Experiment	No	97/87	Comprehension	.91 (.21)
									Attitude	.49 (.21)
Holt 2 (1988)	Conference	8	SSR	Yes	20 min 3 days/ week, 10 wks	Experiment	No	104	Comprehension	.22 (.20)
									Attitude	.10 (.20)
Ibarra (2016)	Dissertation	4	Structured SSR	Yes	30 min daily, 24 wks	Ex post facto	No	190	Comprehension	.06 (.15)
Kariuki (2002)	Conference	7	SSR	No	20 min daily, 12 wks	Quasi	No	40	Comprehension	.34 (.32)
Langford (1983)	Journal	5–6	USSR	No	30 min daily, 6 mos	Quasi	No	250	Attitude	.18 (.13)
									Attitude	.03 (.13)
									Attitude	.90 (.13)
									Word recognition	1.00 (.13)
Melton (1993)	Unpublished report	3–4	SSR	Yes	10 min daily, 6 mos	Quasi	No	12	Word recognition	.52 (.59)
									Comprehension	.33 (.58)
Morgan (2013)	Dissertation	8	SSR	No	15–30 min, 3–5 days/week, 1 year	Ex post facto	No	64	Comprehension	.70 (.26)
Mostow (2013)	Journal	1–4	SSR	Yes	20–25 min daily, 20 wks	Experiment	Yes	178	Word recognition	−.36 (.15)
									Word recognition	−.19 (.16)
									Word recognition	.16 (.15)
									Comprehension	−.73 (.15)
									Word recognition	−.14 (.15)
									Writing	−.17 (.15)
									Attitude	.17 (.15)
Osborn (2007)	Dissertation	2	SSR	Yes	15 min daily, 12 wks	Quasi	No	82	Word recognition	.04 (.22)
									Word recognition	.15 (.22)
Reedy 1 (1994)	Dissertation	3	SSR	No	15 min daily, 12 wks	Quasi	No	74	Attitude	.60 (.29)
									Comprehension	−.06 (.28)
									Writing	.35 (.28)
Reedy 2 (1994)	Dissertation	3	SSR	No	15 min daily, 12 wks	Quasi	No	80	Attitude	.25 (.28)
									Comprehension	−.18 (.28)
									Writing	.30 (.27)
Reis (2007)	Journal	3–6	SSR	Yes	30 min daily, 12 wks	Experiment	No	226	Attitude	.26 (.13)
									Comprehension	.03 (.13)
									Word recognition	.26 (.13)
Reis 1 (2010)	Journal	3–5	SEM-R	Yes	30 min daily, 14 wks	Experiment	No	424	Comprehension	−.02 (.10)
									Attitude	−.09 (.10)
									Word recognition	−.00 (.10)
Reis 2 (2010)	Journal	3–5	SEM-R	Yes	30 min daily, 14 wks	Experiment	No	118	Comprehension	.37 (.18)
									Attitude	−.01 (.18)
									Word recognition	.29 (.19)
Reutzel 1 (1990)	Journal	4 & 6	SSR	Yes	25–30 min daily, 10 days	Experiment	Yes	29	Comprehension	−.10 (.46)
									Comprehension	−.16 (.46)
									Comprehension	−.02 (.46)
									Comprehension	−.26 (.46)
Reutzel 2 (1990)	Journal	4 & 6	SSR	Yes	25–30 min daily, 10 days	Experiment	Yes	31	Comprehension	−.05 (.45)
									Comprehension	−.30 (.46)
									Comprehension	−.15 (.46)
									Comprehension	−.17 (.46)

(continued)

7

TABLE 1 (CONTINUED)

Publication	Publication status	Grade	Program	Reading in control group	Duration	Design	Independent reading used as control	N	Outcome measures	Hedges' g
Reutzel 3 (1990)	Journal	4 & 6	SSR	Yes	25–30 min daily, 10 days	Experiment	Yes	26	Comprehension	.00 (.47)
									Comprehension	.04 (.47)
									Comprehension	−.24 (.47)
									Comprehension	−.35 (.47)
Reutzel 4 (1990)	Journal	4 & 6	SSR	Yes	25–30 min daily, 10 days	Experiment	Yes	28	Comprehension	−.03 (.46)
									Comprehension	−.11 (.46)
									Comprehension	−.29 (.46)
									Comprehension	−.41 (.47)
Siracuse (1991)	Unpublished report	2	SSR	Yes	15 min daily, 10 wks	Quasi	No	51	Attitude	.37 (.28)
									Attitude	.71 (.29)
Spichtig et al. (2019)	Journal	4–5	Web-based SSR	Yes	25 min daily, 24 wks	Experiment	No	426	Comprehension	.14 (.10)
									Word recognition	.15 (.11)
Walters (2006)	Dissertation	9	SSR	Yes	100 min biweekly, 16 wks	Quasi	No	70	Attitude	−.19 (.25)
									Attitude	.00 (.25)
									Attitude	−.17 (.25)
									Comprehension	−.20 (.25)
Williams (2010)	Dissertation	4	SSR	No	15 min daily, 16 wks	Ex post facto	No	41	Attitude	−.04 (.31)
Wilmot (1975)	Dissertation	4	SSR	Yes	<30 min daily, 6 mos	Quasi	No	576	Attitude	.11 (.09)
									Word recognition	.01 (.10)
									Comprehension	−.15 (.10)

IR = independent reading; ISR = independent silent reading; SSR = sustained silent reading; DEAR = drop everything and read; USSR = uninterrupted sustained silent reading; SEM-R = schoolwide enrichment model reading

Many studies have used standardized tests such as the Gates–MacGinitie Reading Achievement Test (e.g., Collins, 1980; Cuevas et al., 2012; Fagella-Luby & Wardwell, 2011), subtests of the Metropolitan Achievement Test (e.g., Evans & Towner, 1975), the Iowa Tests of Basic Skills (e.g., Reis et al., 2007), or responses to comprehension passages in a qualitative reading inventory (e.g., Melton, 1993) to assess reading levels, with word and passage comprehension being the primary components. These results were categorized as *comprehension*. Sometimes studies presented a separate language test. Because the number of studies reporting *language* was too small (four studies) for a separate test, we combined these scores with comprehension.

As an indicator for reading *attitudes*, studies used tools such as the Estes Reading Attitude Scale (e.g., Langford & Allen, 1983), the Adolescent Motivation to Read Profile, or the Elementary Reading Attitude Survey (e.g., Mostow et al., 2013; Reis et al., 2010; Siracuse, 1991). For example, Elementary Reading Attitude Survey consists of 20 statements about reading, such as "How do you feel about reading on a rainy Saturday?" accompanied by four pictures of Garfield depicting different emotional states ranging from very positive to very negative.

Concerning the publication, we coded the *author, publication year*, and *publication status* (i.e., dissertation, journal article, unpublished report, or conference contribution).

When assessing the characteristics of the studies, we incorporated coding for *sample size* and *study design* (i.e., randomized, controlled trial or not) alongside an evaluation based on the five Cochran *dimensions used to assess the risk of bias*. These dimensions encompassed baseline equivalence, control of confounding factors, assessment of implementation fidelity, missing data, and the use of validated and unbiased measurement tools. In the included studies, IR was not always the intended intervention. We also found several studies using *IR as the control condition* for other reading interventions. These interventions often focused on specific comprehension strategies, such as finding the main idea. In these instances, researchers anticipated that some form of comprehension instruction would be more effective than IR (e.g., Reutzel & Hollingsworth, 1990).

We also coded intervention characteristics such as the *number of IR sessions per week, session duration*, and the *duration of the evaluated intervention in weeks* and student characteristics such as the *grade*, enabling us to distinguish between primary and secondary education. Additionally, we endeavored to code the *participants' socioeconomic status* and whether they encountered *specific learning-related challenges* at school. Unfortunately, we were unsuccessful in coding the socioeconomic status of the groups because not all studies provided sufficient information for a founded
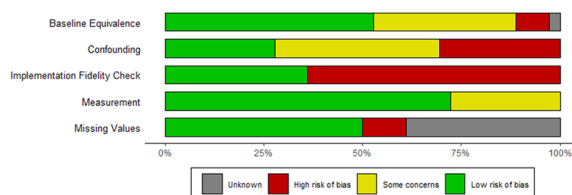
FIGURE 2.   *Results of risk-of-bias analysis.*

estimate. Some studies mentioned the inclusion of students who encountered challenges with reading (e.g., Fagella-Luby & Wardwell, 2011; Mostow et al., 2013). However, in most cases, this inclusion applies to only a small subset of participants rather than the entire group. Therefore, we chose to exclude both socioeconomic status and learning-related challenges from our analysis.

To test the third hypothesis, we looked for a description of the *activities in the control group* (e.g., undefined, literary arts, spelling, oral reading, title of a reading program, and remedial reading). Based on this information, we coded whether the control condition was involved in any form of reading instruction (e.g., Gray, 2012; Walters-Parker, 2006), an activity other than reading (e.g., health activities [Langford & Allen, 1983]), or something undefined (e.g., Higgens, 1981; Kariuki & Replogle, 2002) during the time the experimental condition spent on IR.

Finally, we computed the standardized differences between the mean of the IR group and that of the control group at posttests for each study. Due to the relatively small samples in quite a few studies—26 studies had sample sizes <50—we preferred Hedges' *g* to other measures. A positive effect size indicates a favorable outcome of IR.

We coded findings from the same publication as two or more independent studies when we could not calculate overall effect sizes but only for subsamples, such as different reading proficiency levels (Davis, 1988), different grades (Fagella-Luby & Wardwell, 2011; Harris-Mobley, 2015), various years (Harris-Mobley, 2015), different schools (Reedy, 1994), or various comparison conditions (Fagella-Luby & Wardwell, 2011). Table 1 summarizes the main coding items across the samples.

Interrater Reliability

Two authors each independently coded all 29 publications. The mean raw agreement across the total category matrix was 97.2%, with single-variable agreement rate ranging from 77.8% to 100%. Cohen's kappa's ranged from .67 to .72.

Risk-of-Bias Analysis

Figure 2 presents the risk of bias in a summary bar plot with all 29 primary publications, distinguishing five dimensions. Bias was most severe for implementation fidelity.

Only 12 primary publications reported conducting a check on the implementation of IR, whereas the remaining studies may have conducted a check but did not report it. Most studies controlled for potential confounders such as curriculum, school, teacher, and time devoted to literacy-related activities. A "high risk" rating indicates that the study controlled for at most one of the four potential confounders, whereas a "low risk" rating indicates that all four confounders were controlled. Many studies received a rating of "some concerns" because they only controlled for two or three confounders. Regarding baseline equivalence, nearly all studies reported demographic equivalence, whereas ~30% failed to report pretest equivalence. The amount of missing data was notably high (>33%) in only a few studies, but it often remained unspecified. Concerning measurement, >70% of the studies used validated and unbiased tests.

Meta-analytic Procedures

Because studies with larger sample sizes provide more reliable estimates of the population mean due to a smaller standard error, effect sizes are pooled by weighting each outcome by the inverse of its variance. Most studies in the current set contain more than one outcome measure or effect size, so we must deal with the interdependency of effect sizes. We therefore applied a three-level structure to the meta-analytic model accounting for sampling variance of the extracted effect sizes at level 1, variance between effect sizes extracted from the same study at level 2, and variance between studies at level 3 (Cheung, 2014). The three-level approach allows for examining within-study heterogeneity as well as between-study heterogeneity.

Conducting the analysis, we assumed that individual effect sizes (level 2, defined by an effect size ID) were nested within studies (level 3, defined by a study ID). We performed two one-sided log-likelihood-ratio tests to determine whether the within-study and between-study variances were significant. We compared the fit of the original three-level model to the fit of a two-level model in which within-study or between-study variance was no longer modeled.

If there appeared to be greater variability in effect sizes (both within and between studies) than what could be attributed to sampling variance alone, we proceeded with moderator analyses to examine variables that might account for this variance. To that end, we extended the three-level random-effects model with study and effect size characteristics, thus turning the model into a three-level mixed-effects model. To fit multilevel meta-analytic models, we used the rma.mv function of the metafor package that can be extended by including moderators (Harrer et al., 2021).

Interrelated moderators can result in significant multicollinearity in the analyses. Therefore, it is advisable to explore the potential moderating effects of multiple variables, including both sample and design characteristics. In the final

phase of the moderator analyses, we extended the meta-analytic model by incorporating all significant moderating variables using the metafor package in R (Harrer et al., 2021).

Results

Overall Effects of IR

The studies incorporated outcome measures such as word identification, word attack strategies skills, decoding, fluency, writing, comprehension (including vocabulary and passage meaning), and students' attitude toward reading. The overall association between IR and outcome measures in 47 independent samples, involving 7,493 students, was .081 (expressed in Hedges' g) with a standard error of .04. Because the 95% confidence interval, .0004–.16, did not include zero, this overall effect differed significantly from zero ($t(103) = 1.99$, $p = .049$). The sampling error variance on level 1 made up ~28% of the total variation in the data. The value of level 2, heterogeneity variance within studies, was much higher, totaling ~47%. Finally, the share falling to level 3, the between-study heterogeneity, made up ~25%. This result indicates substantial within-study heterogeneity on the second level, followed by differences between studies, indicating a need for a moderator analysis.

Comparing the three-level model with the fit of two-level models, we concluded that we need to include both within-study and between-study variables. If we set the smallest amount, the level 3 variance, representing the between-study heterogeneity, to zero, the full, three-level model showed a better fit than the two-level model. The Akaike information criterion and the Bayesian information criterion were lower for the three-level model, which indicates favorable performance. The likelihood-ratio test comparing both models was significant ($\chi^2 = 5.70$, $p = .017$), indicating that the three-level model provided a better fit. Hardly surprising considering the heterogeneity variance within studies, results were similar when we set level 2 variance to zero. The likelihood-ratio test was significant, comparing the reduced model with the full model ($\chi^2 = 61.74$, $p = .0001$). The added complexity of a three-level model seems to be justified.

Effects per Outcome Measure

Next, we tested whether the three combined sets of outcome measures—word recognition, comprehension, and reading attitude—exhibited similar effects of IR. Among 104 effect sizes (47 studies), most—59 effect sizes (40 studies)—were obtained from comprehension assessments. Surprisingly, their number was much lower for word recognition, with 20 effect sizes (14 studies), and students' attitudes toward reading, with 22 (17 studies; see Table 2). Several studies reported more than one effect size within each of the three sets. Therefore, we also used the meta-analytic model with a three-level structure for the separate

outcomes even though, in some cases, the three-level models were not performing significantly better than a reduced model including only two levels.

For comprehension, the pooled Hedges' g estimated with the three-level model was not significant (Hedges' $g = -.014$; 95% CI [−.10, .08]; $p = .76$). By contrast, for word recognition (Hedges' $g = .21$; 95% CI [.03, .39]; $p = .026$) and reading attitude (Hedges' $g = .18$; 95% CI [.04, .32]; $p = .016$), the Hedges' g values were much higher and different from zero. Hence, IR at school appears to be significantly more beneficial for word recognition (Hedges' $g = .21$) and students' attitudes toward reading (Hedges' $g = .18$) than for comprehension (Hedges' $g = −.01$). This is a surprising result considering that most researchers focused on comprehension (40 of 47 studies), indicating that they expected this measure to be the primary outcome of IR.

Program Implementation as Moderator

We first tested whether the characteristics of the intervention or control condition influenced the effect sizes. Most studies took place in primary education ($k = 30$), whereas a smaller part ($k = 17$) concerned secondary education (cf. Erbeli & Rice, 2021). Although primary education students scored lower than secondary education students when we included all outcome measures, the difference was not statistically significant ($F(1, 102) = 1.68$, $p = .20$). We could not test the effect of primary versus secondary education on word recognition because the studies with word recognition as an outcome measure included only two studies in secondary education. However, we could test the impact of education level on attitude and comprehension. The studies targeting attitude did not reveal a significant effect ($F(1, 20) = .01$, $p = .91$). However, the impact of IR on comprehension scores was significantly lower in primary education than in secondary education ($F(1, 57) = 4.42$, $p = .04$; see Table 2).

The timing of the interventions was quite similar across studies. In most studies, IR took place daily ($M = 4.63$ days per week, SD = 0.98), most lasting less than half an hour per session (median = 25 minutes). The median for the intervention period was <4 months (median = 15 weeks) with few very brief or lengthy exceptions. Program duration revealed the most variation; therefore, we carried out a meta-regression model with the number of weeks as the predictor. We did not find an effect of intervention duration whether we focused on all outcomes combined or on one of the clusters—word recognition, comprehension, or reading attitude.

Next, we tested the effect of activities in the control condition. In 15 of 47 studies, the control group took part in some form of reading described in various ways (e.g., basal instruction, reading tutor, regular instruction, or language arts). In the rest of the studies, the control group spent time on activities other than reading or undefined activities. We hypothesized that the effects of IR would be stronger if the control

TABLE 2

Effect Sizes Overall, per Outcome Measure, and Moderator/Measure insofar as Significant

	$N_{studies}$	N_{effect_sizes}	Estimate	SE	95% Confidence Interval LB	95% Confidence Interval UB	t Value	p Value
Aggregated effect sizes	47	104	.08	.04	.0004	.16	1.99	.0489
Comprehension/language	40	59	−.01	.05	−.10	.08	−0.30	.76
Word recognition	14	20	.21	.09	.03	.39	2.41	.0265
Attitude	17	22	.18	.07	.04	.32	2.63	.016
Moderator effects								
*Outcome measure: Aggregated effect sizes								
IR as intervention	36	69	.14	.04	.06	.22	3.37	.01
Focus on another intervention	11	35	−.15	.08	.03	.02	−1.74	.09
Publication year	—	—	−.006	.003	−.01	−.001	−2.24	.03
*Outcome measure: Comprehension								
IR as intervention	29	31	.03	.05	−.06	.13	0.74	.46
Focus on another intervention	11	28	−.24	.10	−.44	−.04	−2.44	.018
Primary education	27	46	−.09	.05	−.20	.02	−1.58	.12
Secondary education	13	13	.12	.08	−.04	.27	1.52	.13
*Outcome measure: Word recognition								
Reading in control group	10	14	.03	.08	−.13	.20	0.43	.67
Other or undefined activity in control group	5	6	.44	.12	.18	.69	3.60	.002

SE = Standard Error; LB = Lower Bound; UB = Upper Bound

group did not engage in any reading activities. To that end, we differentiated between studies in which the control group engaged in reading activities while the experimental group was involved in IR and studies in which the control group participated in other than reading or undefined activities. We used a meta-regression model with a dummy-coded predictor—reading activities in the control group or not. When we included all outcome measures, we did not find a difference between control groups with some form of reading instruction versus other activities ($F(1, 102) = .23$, $p = .63$).

Because the reading activities in the control group often targeted word recognition, we hypothesized that particularly the set of studies with word recognition as an outcome measure might be most sensitive to what happens in the control condition. So we also tested the effect of control condition per outcome measure—comprehension, reading attitude, or word recognition. The meta-regression model did not reveal significant effects for comprehension ($F(1, 57) = .69$, $p = .41$) or reading attitude ($F(1, 20) = .77$, $p = .39$). However, the activities in the control group made a difference for word recognition ($F(1, 18) = 6.6407$, $p = .0190$; see Table 2). With any form of reading activity in the control condition, the effect-size estimate equaled Hedges' $g = .034$ (95% CI [−.13, .20]), indicating that the control condition was as effective as IR in promoting word recognition. By contrast, the estimate was substantially higher for the studies that reported that nonreading activities occurred or studies

that did not specify what happened in the control condition. In those cases, the effect size equaled slightly less than half a standard deviation (Hedges' $g = .44$; 95% CI [0.18, 0.69])—according to Cohen's criteria, a moderately high effect size, and according to Kraft (2020), a high effect size.

Study Characteristics

Including all effect sizes, we found a higher score when the researcher had designed the study to test IR (35 studies) compared with studies in which the focus was on another reading intervention (12 studies; $F(1, 102) = 9.11$, $p = .0032$). The study's focus also was significant when we only included studies focused on comprehension ($F(1, 57) = 6.39$, $p = .0142$). Due to insufficient variation, we could not test this moderator's effects when we limited the studies' selection to those with word recognition or attitudes as outcome measures. These two sets included only two studies and one study where the focus of the intervention group was an intervention other than IR.

Regarding other study characteristics (e.g., study quality, publication status, study completion year, and sample size), we observed a negative effect only for publication year in the sample that included all outcome measures ($F(1, 102) = 5.34$, $p = .023$). This moderating effect indicated that more recent studies showed less strong effects than older studies. We did not find this moderating effect in the sets that

targeted one of the three outcome measures—word recognition, comprehension, or attitude.

When more moderators were significant, we built multiple moderator models to exclude overlap. Targeting the sample with all effect sizes, we found that the two significant moderators—study focus (IR versus another reading treatment as target intervention) and publication year—remained significant, with *p* values equal to .0085 and .0439. Targeting comprehension, we found that study focus and education level (secondary versus primary education) were significant moderators. However, in a multivariate model, the education level was no longer significant (*p* = .1842), and the study focus only approached significance (*p* = .0696). Word recognition and attitude did not have multiple significant moderators.

Publication Bias

We used several approaches to investigate publication bias. We did not find evidence for the hypothesis that published studies had higher effect sizes than unpublished studies ($F(1, 102) = .0078$, $p = .93$). Neither did we find evidence for a negative correlation of sample size with the magnitude of the effect size ($F(1, 102) = .39$, $p = .53$), which would indicate a bias against publishing findings that are not statistically significant (Levine et al., 2009).

Furthermore, we visually investigated the relationship between the studies' observed effect sizes on the *x*-axis against a measure of their standard error on the *y*-axis. To carry out such a procedure, we created a new data set with one effect size per study calculated by averaging multiple effect sizes. The effect size was nonsignificant (Hedges' *g* = .023; 95% CI, [−.05, .10]) and smaller than the one reported earlier, resulting from a multilevel approach. The funnel plot had a wide-spreading top with positive and negative effect sizes. Among these studies at the top with small standard errors, some used IR as the control condition. In particular, these studies often exhibited negative effect sizes. Conversely, when studies aimed to assess the impact of IR, they showed positive effect sizes. Thus, the data points at the top of the funnel plot were widespread. However, they also were asymmetrical. The overrepresentation of negative effect sizes was attributed to a considerable number of studies focusing on the evaluation of a reading intervention other than IR. Egger's (1997) regression test, which tested for asymmetry in the funnel plot, confirmed the asymmetry. The intercept, equal to −1.17 (SE = .55), was significant ($t = −2.15$, $df = 45$, $p = .037$). The trim and fill procedures' outcomes included the addition of 15 studies with high positive effect sizes, leading to a significant increase in the overall effect size from Hedges' *g* = .02 to Hedges' *g* = .23, with a 95% CI of [.12, .33].

Finally, we wanted to ensure that the effect we estimated is not spurious, an artifact caused by selective reporting. A *p*-curve precisely addresses this concern. It allows us to check
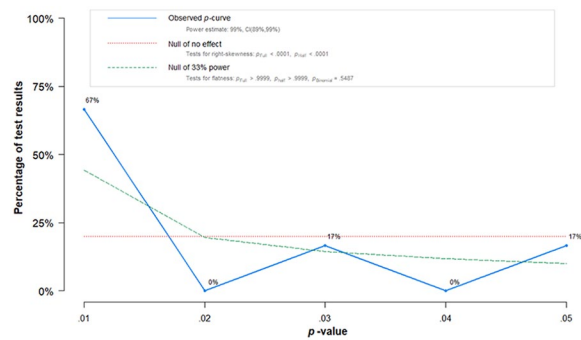

FIGURE 3. *The p-curve plot.*
Note. The observed *p*-curve includes six statistically significant (*p* < .05) results, of which four are *p* < .25. Forty-one additional results were entered but excluded from the *p*-curve because they were *p* > .05. The solid line represents the *p*-curve based on our data and is visibly right skewed, just as we may expect when the studies measure a true effect. The dashed line shows the expected *p* value distribution assuming 33% power, and the dotted line shows the uniform distribution we would expect when there is no effect.

whether *p* values slightly below .05 were overrepresented and highly significant results underrepresented (see Figure 3, the *p*-curve plot). The meta-analysis included six *p* < .05 and four *p* < .025 samples. Two right-skewness tests were significant: the full *p*-curve test (*p* < .001) and the test based on the half *p*-curve (*p* < .001), which indicated that our data contained evidential value: small and high *p* values were equally likely. In contrast, two of the three flatness tests were insignificant, with *p* values as low as 1, meaning that evidential value is neither absent nor inadequate.

Discussion

Impact of IR

Our first hypothesis proposed that IR is an effective strategy in reading pedagogy, resulting in improvements in reading proficiency, which is an important objective of primary and secondary education. The findings support the notion that allocating dedicated time, preferably on a daily basis, for IR makes a valuable contribution to the reading curriculum. Based on the conventional Cohen criteria, the overall relationship between IR and all outcome measures, as indicated by Hedges' *g* with a value of *g* = .08 and a 95% CI of [.0004, .16], falls within the low range. However, considering Kraft's (2020) benchmarks, which are more appropriate in the context of the current set of studies due to the studies being conducted in schools instead of the lab and the measures being standardized rather than intervention based, an effect size ranging between .05 and .20 is considered a medium effect.

No evidence was found to suggest an overestimation of the impact of IR due to the file-drawer problem. The file-drawer problem occurs when researchers face difficulties in publishing nonsignificant findings, leading to a situation where many

such studies remain unpublished and tucked away in researchers' file drawers. To address this issue, we included unpublished reports in our analysis, in particular dissertations, to mitigate any potential bias leading to overestimation of effect sizes. In the last three decades, IR has emerged as a prominent research area for doctoral students. Our search for relevant studies yielded 14 dissertations, collectively including 22 studies on this topic. Dissertations are considered the best source for unpublished work. The effects reported in these dissertations were generally smaller compared to those in published studies, which is consistent with previous meta-analyses, such as the one conducted by Rosenthal (1969). However, it is important to note that despite the smaller effects, the findings from these dissertations did not significantly differ from those reported in other published studies.

The overall effect size of IR in the current set of studies actually may underestimate its impact due to a group of studies where IR was not the target intervention. For example, in the study conducted by Reutzel and Hollingsworth (1990), interventions were implemented to train students in various skills, such as locating details, drawing conclusions, finding the sequence, and determining the main idea. The students receiving these trainings were compared with students who received IR. After the interventions, tests specifically designed to assess targeted intervention skills, such as identifying the main idea, were administered. It is therefore not surprising that students who underwent these interventions performed better than the control group, which received IR that prepared students less effectively for the test (e.g., Reutzel & Hollingsworth, 1990). Due to the inclusion of these studies, the effect size of IR was somewhat reduced.

How IR Influences Reading Proficiency

There was a significant amount of heterogeneity variance within the studies, totaling ~47%, which aligns with the expectation that the outcomes vary across the three distinct measures (i.e., word recognition, comprehension, and attitude toward reading). The effect size is highest when examining word recognition. According to Kraft's benchmarks, an effect size as substantial as Hedges' $g = .21$ is considered large. This finding provides support for the hypothesis that IR offers additional print exposure that children need to develop their word-recognition skills. Furthermore, the observation that the effects of IR on word recognition diminish when considering studies where the control group receives some form of systematic reading training aligns with this conclusion. This finding suggests that IR may yield effects on word recognition comparable with those of reading activities that are organized and guided by teachers for the purpose of practice. However, we cannot conclude from this that students do not need systematic instruction or that instruction is equivalent to IR. Instead, it confirms that a portion of the teacher-guided training includes the type of practice that is also present during IR.

When examining reading attitudes, we observe a similar effect size as we saw for word recognition (Hedges' $g = .18$; 95% CI [0.04, 0.32]; $p = .016$). IR has a rather strong effect on students' attitude toward reading. This indicates that IR offers an additional benefit in terms of students' appreciation and enjoyment of reading. This finding corroborates the expectation that IR provides students with an opportunity to explore captivating books and discover how reading can align with their individual interests. It is important to note that while many students may have this opportunity in their home environment, it may not be readily available to all students, making it essential to provide such experiences in schools. By enhancing children's attitude toward reading, IR has the potential to serve as an incentive for them to engage in more voluntary reading outside of school. However, it should be acknowledged that we were unable to test the hypothesis regarding whether students who participate in IR at school also engage in more reading activities outside of school, which would be consistent with this finding.

Despite the fact that most studies gave priority to comprehension as the primary measure, the observed impact (Hedges' $g = -.014$) is not statistically significant and is much lower than the effects observed for word recognition and reading attitude. This finding is not particularly surprising given that, unlike word recognition and reading attitude, comprehension relies on skills that are challenging to acquire. These skills include cultural and content knowledge, reading-specific background knowledge, and theory of mind, as proposed by Duke and Cartwright's (2021) active view of reading. It seems plausible that the duration of time devoted to IR in the current set of studies may not have been sufficiently long to produce a noticeable effect on such skills and knowledge. In other words, the average duration of IR across the analyzed studies in this meta-analysis, which was 25 minutes per day with a median program duration of 3 months, may not have allowed enough time for extensive reading, which is crucial for improving the complex knowledge unique to reading comprehension. Moreover, because comprehension involves complex cognitive processes, selecting appropriate tests can be challenging, which also may have contributed to the observed lack of impact. Furthermore, in a significant number of studies ($n = 12$) focusing on comprehension, IR was not the target intervention. In these cases, the estimated effect size showed a notably negative value ($-.24$, SE = .10). This pattern, particularly prevalent in studies examining comprehension, may have contributed to the low effects of IR on comprehension.

Limitations

One notable limitation of this meta-analysis is its inclusion of (quasi-)experimental studies, which exhibit several flaws. These studies may particularly lack verification of the implementation of IR and control over potential confounding

factors. Additionally, they fail to address documented sources of inequality in schooling experiences and reading instruction, such as socioeconomic class, in a manner that allows for reliable coding and inclusion in the analyses. The scarcity of (quasi-)experimental research in this domain may imply that IR has not been accorded high priority probably due to the pressing emphasis on ensuring that children initially acquire technical reading skills (Reinking et al., 2023).

The overall effect size potentially could have been significantly higher if more studies had focused on fundamental word-recognition skills such as decoding, sight words, and fluency, as well as reading attitude, instead of primarily assessing comprehension using measures related to literal understanding and inferential abilities.

Furthermore, it is important to consider that the evaluation of reading attitude relies primarily on questionnaires, which introduces the potential for social desirability biases. Due to the awareness of the significance placed on reading, children may tend to report more positive attitudes than they genuinely hold. To obtain a more precise evaluation, it would be beneficial to incorporate assessment measures that are less susceptible to such biases. Additionally, conducting long-term assessments that use print exposure lists could offer valuable insights into whether IR promotes increased reading beyond school and influences long-term reading habits.

The relatively small sample sizes in the included studies are notable, particularly considering that Yoon (2003) reported an average effect size of .11 twenty years ago. This effect size suggests that studies should have relatively large sample sizes to detect significant results. Based on the current findings, where the expected effect size is around Hedges' $g = .20$ or lower, an experiment would require slightly fewer than 200 participants to yield statistically significant results. Unfortunately, only a few studies in the analysis possess sample sizes large enough to address this issue, which explains why many studies yielded nonsignificant results, as also concluded by Erbeli and Rice (2021).

There is some variability across the studies, including differences in the availability and types of books or the level of teacher supervision. In some cases, students may even listen to audio-recorded books (Boeglin-Quintana & Donovan, 2013). Unfortunately, we were unable to examine the effectiveness of these variations and whether they have an impact on the results. Descriptions of IR were limited in many studies, which made it challenging to gather comprehensive information for coding potentially relevant variations and conducting a fair comparison across studies.

Conclusions

While developing word-recognition skills through IR is noteworthy, it is not the primary focus of IR as envisioned by Krashen (2004) and other proponents. The central aim of IR, supported by the current findings, is that incorporating authentic, unrestricted reading practice into the school curriculum leads to students perceiving reading as more exciting and enjoyable compared with situations where IR is not included and students receive only traditional reading instruction. This finding aligns with Krashen's (2011) theory, which posits that IR can deeply engage students in reading and cultivate a positive attitude toward it. As emphasized by Krashen in his numerous presentations and publications, providing engaging reading materials and regular opportunities for students to immerse themselves in books can serve as a motivational catalyst, inspiring students to embrace reading as a primary means of acquiring new knowledge and nurturing their ongoing personal growth.

The impact of IR on attitudes alone provides a compelling reason to incorporate it into reading education. The current findings support the hypothesis that IR satisfies students' curiosity and interests, enriching not only their intellectual but also their emotional lives. Our meta-analysis reinforces the argument that IR should be an essential component of the reading curriculum and a regular practice in reading pedagogy, applicable to both primary and secondary education. However, it is important to remember that IR is not intended to replace teacher-led reading instruction. Instead, it serves as a complementary activity that enhances students' engagement with reading, making it indispensable not only for supporting reading skill development but also for fostering a love of reading.

ORCID iD

Adriana G. Bus https://orcid.org/0000-0002-2836-176X

References

*Allen, K. (2017). *An evaluation study of fifth-grade independent reading and reading achievement* [Unpublished doctoral dissertation]. Southeastern University. ProQuest Dissertations & Theses Global.

Anderson, R. C., Fielding, L. G., & Wilson, P. T. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*(3), 285–304. https://doi.org/10.1598/RRQ.23.3.2

Boeglin-Quintana, B., & Donovan, L. (2013). Storytime using iPods: Using technology to reach all learners. *TechTrends*, *57*(6), 49–56. https://doi.org/10.1007/s11528-013-0701-x

*Borjes, J. A. (2009). *Repeated oral reading approach versus independent silent reading approach for reading fluency and comprehension* [Unpublished doctoral dissertation]. University of the Incarnate Word. ProQuest Dissertations & Theses Global.

Bryan, G., Fawson, P. C., & Reutzel, D. R. (2003). Sustained silent reading: Exploring the value of literature discussion with three non-engaged readers. *Literacy Research and Instruction*, *43*(1), 47–73. https://doi.org/10.1080/19388070309558400

Bus, A. G., van IJzendoorn, M. H., & Mol, S. E. (2021). Meta-analysis. In M. H. Mallette, & N. K. Duke (Eds.), *Literacy research methodologies* (3rd ed., pp. 234–263). Guilford Press.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

*Cirucci, C. L. (2017). *The effects of shared reading and independent reading on incidental word learning of new vocabulary for third grade proficient and non-proficient readers* [Unpublished doctoral dissertation]. Widener University. ProQuest Dissertations & Theses Global.

*Cline, R. K., & Kretke, G. L. (1980). An evaluation of long-term SSR in the junior high school. *Journal of Reading*, *23*(6), 503–506. https://www.jstor.org/stable/40028834

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal Social Psychology*, *65*(3), 145–153. https://doi.org/10.1037/h0045186

*Collins, C. (1980). Sustained silent reading periods: Effect on teachers' behaviors and students' achievement. *The Elementary School Journal*, *81*(2), 108–114. https://doi.org/10.1086/461213

*Cuevas, J. A., Russell, L. R., & Irving, M. A. (2012). An examination of the effect of customized reading modules on diverse secondary students' reading comprehension and motivation. *Educational Technology Research & Development*, *60*(3), 445–467. https://doi.org/10.1007/s11423-012-9244-7

*Davis, Z. T. (1988). A comparison of the effectiveness of sustained silent reading and directed reading activity on students' reading achievement. *The High School Journal*, *72*(1), 46–48. https://www.jstor.org/stable/40364822

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, *56*(Suppl. 1), S25–S44. https://doi.org/10.1002/rrq.411

Egger, M. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Erbeli, F., & Rice, M. (2021). Examining the effects of silent independent reading on reading outcomes: A narrative synthesis review from 2000 to 2020. *Reading & Writing Quarterly*, *38*(3), 253–271. https://doi.org/10.1080/10573569.2021.1944830

*Evans, H. M., & Towner, J. C. (1975). Sustained silent reading: Does it increase skills? *The Reading Teacher*, *29*(2), 155–156. https://doi.org/10.2307/20193969

*Fagella-Luby, M., & Wardwell, M. (2011). RTI in a middle school: Findings and practical implications of a tier 2 reading comprehension study. *Learning Disability Quarterly*, *34*(1), 35–49. https://doi.org/10.2307/23053295

Fielding, L. G., Wilson, P. T., & Anderson, R. C. (1986). A new focus on free reading: The role of trade books in reading instruction. In T. E. Raphael (Ed.), *Contexts of school-based literacy* (pp. 149–160). Random House.

Gambrell, L. (2007). Reading: Does practice make perfect? *Reading Today*, *24*(6), 16. https://doi.org/10.1598/RT.24.6.6

Garan, E. M., & DeVoogd, G. (2009). The benefits of sustained silent reading: Scientific research and common sense converge. *The Reading Teacher*, *62*(4), 336–344. https://doi.org/10.1598/RT.62.4.6

Gottfried, A. W., Gottfried, A. E., Bathurst, K., Guerin, D. W., & Parramore, M. M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. In M. H. Bornstein, & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development* (pp. 189–207). Erlbaum.

*Gray, H. L. (2012). *The effects of sustained silent reading on reading achievement and reading attitudes of fourth grade students* [Unpublished doctoral dissertation]. University of North Carolina at Chapel Hill. ProQuest Dissertations & Theses Global.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman & Hall/CRC Press.

*Harris-Mobley, I. J. (2015). *Closing gaps of disadvantaged readers using drop everything and read* [Unpublished doctoral dissertation]. Capella University. ProQuest Dissertations & Theses Global.

Hiebert, E. H. (2024). Enhancing opportunities for decoding and knowledge building through beginning texts. *The Reading Teacher*, *77*(6), 965–974. https://doi.org/10.1002/trtr.2303

Hiebert, E. H., Wilson, K. M., & Trainin, G. (2014). Are students really reading in independent reading contexts? An examination of comprehension-based silent reading rate. In E. H. Hiebert, & D. Ray Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers* (pp. 151–167). International Reading Association. https://doi.org/10.1598/0833.17

*Higgens, K. J. (1981). *Sustained silent reading in the fifth grade: The effects of sustained silent reading on speed, comprehension, word study skills, and vocabulary* [Unpublished doctoral dissertation]. Brigham Young University. ProQuest Dissertations & Theses Global.

*Holt, S. B., & O'Tuei, F. S. (1988, April). *The effects of sustained silent reading and writing on achievement and attitudes of seventh and eighth grade students reading two years below grade level* [Paper presentation]. Annual meeting of the American Educational Research Association, New Orleans, LA. ERIC Number: ED308484.

*Ibarra, S. (2016). *The effect of independent reading on STAAR reading scores of fourth grade students* [Unpublished doctoral dissertation]. Northcentral University. ProQuest Dissertations & Theses Global.

Ivey, G., & Johnston, P. H. (2013). Engagement with young adult literature: Outcomes and processes. *Reading Research Quarterly*, *48*(3), 255–275. https://doi.org/10.1002/rrq.46

*Kariuki, P., & Replogle, D. (2002, November). *The effects of independent reading on reading ability of seventh grade students* [Paper presentation]. Annual meeting of the Mid-South Educational Research Association, Chattanooga, TN. ERIC Number: ED474927.

Kelley, M., & Clausen-Grace, N. (2006). R5: The sustained silent reading makeover that transformed readers. *The Reading Teacher*, *60*(2), 148–156. https://doi.org/10.1598/RT.60.2.5

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading*, *12*(1), 1–23. https://doi.org/10.1080/10888430701746849

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Krashen, S. (2001). More smoke and mirrors: A critique of the National Reading Panel report on fluency. *Phi Delta Kappan*, *83*(2), 119–123. https://doi.org/10.1177/003172170108300208

Krashen, S., Lee, S. Y., & McQuillan, J. (2012). Is the library important? Multivariate studies at the national and international level. *Journal of Language and Literacy Education*, *8*(1), 26–38. http://jolle.coe.uga.edu/wp-content/uploads/2012/06/Is-the-Library-Important.pdf

Krashen, S. D. (2004). *The power of reading: Insights from the research* (2nd ed.). Libraries Unlimited.

Krashen, S. D. (2011). *Free voluntary reading*. Libraries Unlimited.

*Langford, J. C., & Allen, E. G. (1983). The effects of U.S.S.R. on students' attitudes and achievements. *Reading Horizons: A Journal of Literacy and Language Arts*, *23*(3), Article 10. https://scholarworks.wmich.edu/reading_horizons/vol23/iss3/10

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, *76*(3), 286–302. https://doi.org/10.1080/03637750903074685

Matsui, T., & Noro, T. (2010). The effects of 10-minute sustained silent reading on junior high school EFL learners' reading fluency and motivation. *Annual Review of English Language Education in Japan*, *21*, 71–80. https://doi.org/10.20581/arele.21.0_71

*Melton, E. J. (1993). *SSR: Is it an effective practice for the learning disabled?* [Unpublished research report]. ERIC Number: ED397569. Marywood College.

Merga, M. K. (2015). "She knows what I like": Student-generated best-practice statements for encouraging recreational book reading in adolescents. *Australian Journal of Education*, *59*(1), 35–50. https://doi.org/10.1177/0004944114565115

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/a0021890

*Morgan, M. S. (2013). *Sustained silent reading in middle school and its impact on students' attitudes and achievement* [Unpublished doctoral dissertation]. Wilmington University. ProQuest Dissertations & Theses Global.

Morrow, L. M. (1983). Home and school correlates of early interest in literature. *Journal of Educational Research*, *76*(4), 221–230. https://www.jstor.org/stable/27539975

*Mostow, J., Nelson-Taylor, J., & Beck, J. E. (2013). Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research*, *49*(2), 249–276. https://doi.org/10.2190/EC.49.2.g

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. TIMSS and PIRLS International Study Center. https://timssandpirls.bc.edu/pirls2011/downloads/P11_IR_FullBook.pdf

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

Nielen, T. M. J., & Bus, A. G. (2015). Enriched school libraries: A boost to academic achievement. *AERA Open*, *1*(4), 1–11. https://doi.org/10.1177/2332858415619417

Organisation for Economic Co-operation and Development (OECD). (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Organisation for Economic Co-operation and Development (OECD). (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://www.oecd-ilibrary.org/sites/53f23881-en/index.html?itemId=/content/publication/53f23881-en

*Osborn, D. F. (2007). *Developing oral reading fluency: Effects of daily use of word walls and daily independent silent reading on oral reading fluency development of second grade students* [Unpublished doctoral dissertation]. Liberty University. ProQuest Dissertations & Theses Global.

Pennington, M. (2011, August). Straight talk with Stephen Krashen on SSR. *Pennington Publishing Blog*. https://blog.pennington-publishing.com/straight-talk-with-stephen-krashen-on-ssr/

Pressley, M., Dolezal, S., Roehrig, A. D., & Hilden, K. (2002). Why the National Reading Panel's recommendations are not enough. In R. Allington (Ed.), *Big brother and the national reading curriculum: How ideology trumped evidence* (pp. 75–89). Heinemann.

*Reedy, J. D. (1994). *Effects of a sustained silent reading program with literature response journals on third graders' attitude, reading achievement, and writing* [Unpublished doctoral dissertation]. Baylor University. ProQuest Dissertations & Theses Global.

Reinking, D., Smagorinsky, P., & Yaden, D. B. (2023, May 24). New York City requires reading instruction to be phonics-based. *Washington Post*. https://www.washingtonpost.com/education/2023/05/23/phonics-reading-analysis/

*Reis, S. M., Eckert, R. D., McCoach, D. B., Jacobs, J. K., & Coyne, M. (2010). Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. *The Journal of Educational Research*, *101*(5), 299–315. https://doi.org/10.3200/JOER.101.5.299-315

*Reis, S. M., Gubbins, E. J., McCoach, M. C., Schreiber, F. J., Betsy, D., & Eckert, R. D. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency,

comprehension, and attitude toward reading: An evidence-based study. *The Elementary School Journal*, *108*(1), 3–23. https://doi.org/10.1086/522383

*Reutzel, D. R., & Hollingsworth, P. M. (1990). Reading comprehension skills: Testing the distinctiveness hypothesis. *Literacy Research and Instruction*, *30*(1), 32–46. https://doi.org/10.1080/19388079109558040

Reutzel, D. R., Fawson, P. C., & Smith, J. A. (2008). Reconsidering silent sustained reading: An exploratory study of scaffolded silent reading. *The Journal of Educational Research*, *102*(1), 37–50. https://doi.org/10.3200/JOER.102.1.37-50

Reutzel, D. R., Jones, C. D., & Newman, T. H. (2010). Scaffolded silent reading: Improving the practice of silent reading practice in classrooms. In E. H. Hiebert, & D. R. Reutzel (Eds.), *Revisiting silent reading: New directions for teachers and researchers* (pp. 129–150). International Reading Association.

Rosenthal, R. (1969). Task variations in studies of experimenter expectancy effects. *Perceptual and Motor Skills*, *29*(1), 9–10. https://doi.org/10.2466/pms.1969.29.1.9

Seidenberg, M. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. Basic Books.

Shanahan, T. (2018). For the love of reading: Independent reading at school. *Blogs about Reading*. https://www.readingrockets.org/blogs/shanahan-on-literacy/love-reading-independent-reading-school

*Siracuse, L. A. (1991). *The effects of a sustained silent reading program on the reading attitudes and habits of second-grade students* [Unpublished master's thesis]. College at Brockport. http://digitalcommons.brockport.edu/ehd_theses/109

*Spichtig, A. N., Gehsmann, K. M., Pascoe, J. P., & Ferrara, J. D. (2019). The impact of adaptive, web-based, scaffolded silent reading instruction on the reading achievement of students in grades 4 and 5. *The Elementary School Journal*, *119*(4), 443–467. https://doi.org/10.1086/701705

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*(1), 51–68. https://doi.org/10.3758/BF03208254

Topping, K. J., Samuels, J., & Paul, T. (2007). Does practice make perfect? Independent reading quantity, quality and student achievement. *Learning and Instruction*, *17*(3), 253–264. http://doi.org/10.1016/j.learninstruc.2007.02.002

Van der Sande, L., Wildeman, I., Bus, A. G., & Van Steensel, R. (2022). Personalized expert guidance of students' book choices in primary and secondary education. *Reading Psychology*, *43*(5), 380–404. https://doi.org/10.1080/02702711.2022.2113944

*Walters-Parker, K. (2006). *The effects of two reading interventions on the reading motivation and reading achievement of low-performing high school readers* [Unpublished PhD dissertation]. University of Kentucky. ProQuest Dissertations & Theses Global.

Weber, S. (2018). How teachers can guide library book selection to maximize the value of independent reading time. *The Language and Literacy Spectrum*, *28*(1), Article 4. http://digitalcommons.buffalostate.edu/lls/vol28/iss1/4

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, *28*(1), 34–50. https://www.jstor.org/stable/747815

Wiesendanger, K. D., & Birlem, E. D. (1984). The effectiveness of SSR: An overview of the research. *Reading Horizons: A Journal of Literacy and Language Arts*, *24*(3), Article 9. https://scholarworks.wmich.edu/reading_horizons/vol24/iss3/9

*Williams, N. (2010). *The effects of sustained silent reading on motivation to read* [Unpublished doctoral dissertation]. Walden University. ProQuest Dissertations & Theses Global.

Willingham, D. T. (2015). For the love of reading: Engaging students in a lifelong pursuit. *American Educator*, *39*(1), 4–13. https://www.aft.org/ae/spring2015/willingham-Feb-2021

*Wilmot, M. P. (1975). *An investigation of the effect upon the reading performance and attitude toward reading of elementary grade students, of including in the reading program a period of sustained silent reading* [Unpublished PhD dissertation]. University of Colorado. ProQuest Dissertations & Theses Global.

Wolf, M. (2018). *Reader come home: The reading brain in a digital world*. Harper.

Yoon, J. C. (2003). What a meta-analytic review of three decades of SSR says about reading comprehension. *The Journal of Curriculum & Evaluation*, *6*(2), 171–186. https://doi.org/10.29221/jce.2003.6.2.17140

Authors

ADRIANA G. BUS is a professor emerita at Leiden University, Netherlands; a professor at the University of Stavanger, Norway; and an honorary professor at ELTE Eötvös Loránd University, Budapest, Hungary. She studies ways to encourage reading habits to enhance both reading proficiency and enjoyment. Email: adriana.g.bus@uis.no.

YI SHANG is an associate professor in the Department of Education at John Carroll University. She is interested in researching quantitative methodologies and measurement issues in the field of education. Email: yshang@jcu.edu.

KATHLEEN ROSKOS is a professor emerita in the Department of Education and School Psychology at John Carroll University, Cleveland, Ohio. She studies the design and use of digital books as teaching and learning resources to support early literacy development and promote early literacy skills. Email: roskos@jcu.edu.