Journal of Cognitive Enhancement (2025) 9:455–479
https://doi.org/10.1007/s41465-025-00336-2

REVIEW

Digital Tools for Reading Success: Meta-analyses of Digital Interventions

Angela Pasqualotto[1,3] · Emma G. Cunningham[4] · Claire Holman[1,5] · Benoit Bediou[1,2] · Daphne Bavelier[1,2]

Received: 21 April 2025 / Accepted: 13 October 2025 / Published online: 12 December 2025

Abstract
This meta-analytic work, conducted in accordance with PRISMA guidelines, evaluates the efficacy of digital interventions in improving reading outcomes in children under 15 years of age. We included 41 studies ($k=194$) focused on poor readers and 15 studies ($k=69$) on general school readers, all employing randomized controlled or quasi-experimental designs published between 2000 and 2022. Multilevel meta-analyses with robust variance estimation, conducted using the open-access tool ShinyApp, revealed a medium effect for poor readers ($g=0.433$, CI = [0.3, 0.56], $p<.001$), corresponding to substantial literacy gains, and a smaller effect for general readers ($g=0.256$, CI = [0.08, 0.43], $p=.009$) reflecting more modest progress. Notably, interventions targeting reading-specific skills, such as decoding and comprehension, were effective across both groups, whereas those focused on domain-general cognitive skills, such as attention and working memory, benefited only poor readers. Results varied slightly by orthographic transparency, with numerically stronger effects in languages with simpler syllabic structures. No significant differences emerged between decoding and comprehension outcomes. These findings highlight the potential of tailored digital interventions to enhance reading—especially for children facing literacy challenges— while underscoring the need to adapt them to diverse learner and contextual characteristics to ensure scalability and cross-linguistic applicability.

Keywords Reading · Digital-based tools · Intervention effects · Poor readers

Introduction

Reading proficiency and its constituent abilities —such as reading with fluency and fully comprehending what has been read— are fundamental to academic achievement across domains. As Duke and colleagues stated, these abilities bridge "the gap between learning to read and reading to learn" (Duke et al., 2003).

Despite the importance of reading in every aspect of 21st-century life, a significant proportion of the school-age population fails to fully master this essential skill (McFarland et al., 2019). For instance, approximately 25% of 15-year-olds in OECD countries perform below the minimum reading standard needed for daily life (OECD, 2019). Poor reading proficiency limits life opportunities and contributes to broader societal inequalities, with increased risks of adverse outcomes such as physical and mental health issues (Carneiro & Gordon, 2013; Pape et al., 2011; Parsons & Bynner, 2007). Crucially, individuals with reading difficulties, often referred to as "poor readers", encompass a heterogeneous population that includes, but is not limited to children with developmental dyslexia, which affects 5–7% of the global school-age population (Snowling et al., 2020).

Addressing these challenges requires innovative, scalable solutions. Digital interventions, leveraging advances in Information and Communication Technologies (ICT), may be able to provide such crucial solutions. The current work

✉ Angela Pasqualotto
angela.pasqualotto@unige.ch

✉ Daphne Bavelier
daphne.bavelier@unige.ch

1 Faculty of Psychology and Education Sciences, University of Geneva, Geneva, Switzerland

2 Campus Biotech, Geneva, Switzerland

3 Present address: University of Applied Sciences and Arts of Southern Switzerland, Locarno, Switzerland

4 Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

5 Present address: Faculty of Medicine, University of Geneva, Geneva, Switzerland

presents an evaluation of the state of digital interventions for reading following the PRISMA framework. In the case of reading, different digital interventions have been specifically designed to address either the needs of participants with poor reading skills or the needs of participants within the typical reading range, hereafter termed "general readers". In doing so, these interventions have aimed to follow the best design principle of tailoring intervention content to the user's skill level (Mayer, 2020). As poor readers and general readers bring rather different reading and cognitive skills to the intervention, we present two separate meta-analyses, one for each group. Given our interest in reading acquisition during the school years, we limit our age range to children under the age of 15.

Determinants of Learning To Read

Designing digital interventions to support reading requires an in-depth understanding of the key processing components that support reading acquisition. A rich literature points to the importance of training reading and reading-related skills, which we term domain-specific skills. In addition, training more central cognitive skills such as executive functions and attentional control, which we term domain-general skills, has also been identified as a promising avenue for enhancing reading outcomes.

Domain-specific interventions focus on linguistic skills, such as phonological awareness, decoding, and reading comprehension, which are foundational to reading. Indeed, for alphabetic languages, there is much evidence that systematic instruction in letter-sound correspondences is essential for improving literacy skills in both poor readers and general readers (for a comprehensive meta-analysis, see Torgerson et al., 2019). Once children have developed proficiency in letter-sound mappings, instruction can then incorporate meaning-based strategies to further improve reading outcomes (Castles et al., 2018; Haft et al., 2019). Thus, these strategies offer a robust foundation for achieving literacy in general school populations (Landerl et al., 2021; Snowling, 2014). For digital interventions focusing on domain-specific training, the challenge then is to scaffold the training through different difficulty levels via carefully selected linguistic tasks and appropriate stimuli (Verhoeven et al., 2021). GraphoGame exemplifies this approach by systematically introducing grapheme-phoneme correspondences—the relationship between written letters or letter combinations (graphemes) and their corresponding sounds (phonemes)—through gamified tasks (Richardson & Lyytinen, 2014). GraphoGame adapts to different languages, such as French, by training mappings at various grain sizes, including syllables and whole words (Lassault et al., 2022).

In contrast, domain-general interventions target cognitive skills —such as attentional control and executive functions (EFs)— known to facilitate learning to read, as well as learning in other domains (Grainger et al., 2016; Pennington et al., 2012; Vidyasagar & Pammer, 2010). For instance, fluent reading requires attentional control to focus selectively on relevant parts of the text while inhibiting distractions and working memory to retain and process phonological, morphological, and semantic representations. Among these skills, attentional control is particularly critical for grapheme-to-phoneme conversion and filtering out irrelevant information, both of which underpin reading fluency (Gori & Facoetti, 2015; Grainger et al., 2016; Valdois et al., 2019; Vidyasagar & Pammer, 2010). While the precise mechanisms by which domain-general abilities may affect the acquisition of literacy remains an issue of debate in the field, it is well accepted that, if dysfunctional, they can act as roadblocks to reading acquisition (Cirino et al., 2019; Landerl et al., 2019; Moll et al., 2014). To address these barriers, digital interventions must carefully scaffold task difficulty to enhance cognitive fitness effectively in the target cognitive skill areas. Tools designed to train cognitive skills, such as working memory or attentional control through varied and adaptive tasks, exemplify domain-general interventions (for a review, see Titz & Karbach, 2014).

Because digital interventions for reading have been designed to address both of these possible intervention paths; accordingly, trained domain –specific or general– is used as a moderator in this meta-analytic work. Another moderator of interest, also motivated by theoretical work in reading, concerns the type of reading outcome considered. Decoding, the ability to fluently and accurately read words, contrasts with comprehension, which involves understanding text meaning. Some readers excel in comprehension despite poor decoding, while others demonstrate the reverse pattern, underscoring the distinction between these reading components (Oakhill et al., 2003; Perfetti & Stafura, 2014). This outcome type is therefore used as a second moderator in our meta-analytic work. Finally, the last moderator concerns a key linguistic factor in reading known as syllabic transparency. It refers to the consistency of letter-sound relationships in a language. Highly transparent languages (e.g., Finnish, Italian) have predictable mappings, while less transparent ones (e.g., English, French) are more irregular, influencing decoding and comprehension (Seymour et al., 2003; Ziegler & Goswami, 2005). Together, these moderators allow for a more nuanced analysis of intervention efficacy across diverse reading contexts.

Despite significant advancements and the prevalence of their usage, gaps remain in understanding how digital interventions cater to diverse reading needs and linguistic contexts. Few studies systematically compare domain-specific and domain-general approaches or address linguistic moderators like syllabic complexity. This study addresses these gaps by conducting separate meta-analyses for poor and general readers, offering a comprehensive evaluation of digital interventions for personalized and scalable literacy solutions.

To contextualize the role of digital interventions, we first review traditional paper-and-pencil approaches for poor and general readers, followed by a review of how digital tools seek to transform reading outcomes through individualized and engaging strategies.

Lessons from Paper-and-Pencil Reading Interventions

Before diving into the specifics of our two meta-analyses, it is important to highlight the traditional paper-and-pencil interventions for reading in each of the two groups we consider. Traditional interventions for poor readers, particularly in orthographic systems such as English, emphasize systematic phonics approaches (McArthur et al., 2018). These programs focus on teaching explicit relationships between graphemes and phonemes, which are foundational for decoding and spelling skills. Meta-analytic evidence supports their efficacy; with consistent findings that phonics-based instruction significantly improves literacy outcomes (Ehri et al., 2001; Compton et al., 2014; Galuschka et al., 2014; McArthur et al., 2012; Suggate, 2010). In addition, interventions that teach the interrelations between phonology, morphology and etymology (i.e., structured word inquiry) also show promising results for children with reading difficulties (Colenbrander et al., 2022). In terms of interventions aimed at improving comprehension, these have been found to be effective particularly for older students with reading disorders (Talbott, Wills, & Tankersley, 1994; Wanzek, Vaughn, Kim, & Cavanaugh, 2006). Thus, both phonics training and comprehension-focused training have been identified as providing benefits to poor readers (de Graaff et al., 2009; Melby-Lervåg & Lervåg, 2014, respectively). In sum, meta-analyses show that tailored interventions addressing decoding can achieve medium-to-large effect sizes, benefiting both struggling readers (Hedges $g = 0.49$; Scammacca et al., 2015) and those with severe difficulties (Hedges' $g = 0.54$; Wanzek et al., 2016). Explicit comprehension strategies focusing on inferencing and text summarization yield positive results ($g \approx 0.36$), with maintenance at follow-up ($g \approx 0.22$; Suggate, 2016).

For interventions focused on general readers, foundational literacy skills, including phonological awareness, vocabulary, fluency, and comprehension, are pivotal. Systematic phonics instruction, targeting grapheme-phoneme correspondences, is a cornerstone of effective literacy development (National Reading Panel, 2000). Hybrid approaches that integrate phonics and comprehension instruction have demonstrated substantial efficacy in fostering fluency and comprehension (Buckingham, 2020; Castles et al., 2018). Instructional strategies vary across languages with different levels of orthographic transparency and syllabic complexity (Landerl et al., 2019; Ziegler et al., 2010). Transparent orthographies like Italian allow for straightforward phonics instruction due to consistent letter-to-sound mappings. In contrast, opaque orthographies like English necessitate greater emphasis on irregular patterns and sequencing strategies (Verhoeven & Perfetti, 2021).

In sum, traditional paper-and-pencil interventions for general readers demonstrate modest to moderate impacts across various reading domains. Systematic phonics instruction, which targets foundational decoding skills, has shown a moderate overall effect size ($d = 0.41$) for improving early literacy, with even greater impacts for younger learners ($d = 0.55$; Ehri et al., 2001). For reading comprehension, effect sizes range from small ($d = 0.186$–$d = 0.431$) to medium ($d = 0.786$) depending on the assessment task used (Okkinga et al., 2018). However, traditional paper-and-pencil interventions also present notable limitations, such as restricted adaptability to individual learning trajectories, limited engagement, and logistical constraints in implementation. To overcome these challenges, digital approaches have emerged as promising tools offering greater flexibility, personalization, and scalability in reading interventions.

Digital Approaches in Reading Interventions: Advancements and Challenges

Digital interventions offer a versatile approach to enhancing reading skills with applications both during school instruction and in out-of-school settings, such as at home or in after-school programs (Andreev et al., 2009). Moreover, they can more naturally be built to include best practices from learning, such as keeping the learner in the zone of proximal development, thanks to individualization, boosting engagement through finely tuned interactivity, ease of accessibility, low costs, attractive graphics, and acceptance by users as enjoyable and engaging activities (Barr & Kirkorian, 2023; Jamshidifarsani et al., 2019; Lumsden et al., 2016; Pasqualotto et al., 2023; Plass et al., 2019; Vermeir et al., 2020). In the Project Tomorrow (2019), 75% of school-age students expressed a preference for learning using tablet devices. Beyond such clear gains in acceptability, digital interventions that offer individualized interaction modes and adapt the intervention content based on children's real-time performance may be most effective in enhancing reading skills as reviewed by Jamshidifarsani et al. (2019). Indeed, these authors reported overall low to moderate effects on decoding and reading comprehension, with a more pronounced impact observed among lower-achieving students. Individualized digital interventions are attractive as they offer learners the opportunity to progress at their own pace (Corbett, 2001; Kim, 2006) while remaining engaged with the learning activity (Bryant et al., 2015; Malouf, 1988) thanks to formative feedback and scaffolding (Pasqualotto et al., 2022). In addition to being relatively easy to deploy at scale, digital interventions can incorporate interactive and gamified elements, deliver real-time adaptive adjustments, and provide immediate individualized feedback and pacing, making them a particularly promising approach to reading training, especially for young learners and those with attentional or learning difficulties.

Digital interventions for reading have largely mirrored more traditional, paper-and-pencil approaches when it comes to content. Digital interventions targeting reading and reading-related skills such as phonics-based training appear most promising as reviewed by Ostiz-Blanco et al., 2021 (see also Alqahtani & Rajkhan, 2020; Cidrim & Madeiro, 2017; MacArthur et al., 2001; Ronimus et al., 2020). Similarly, a meta-analysis of 67 studies (Verhoeven et al., 2021) indicated that computer-assisted word reading interventions generally result in moderate positive effects on decoding ($k=694$; $g=0.36$, CI = [0.28, 0.43]), regardless of reading language. Evidence also indicated a transfer effect to reading comprehension, with 69 effect sizes yielding a mean value of 0.21 (CI = [0.13, 0.29]). While Verhoeven et al. (2021) reported moderate positive effects of computer-assisted word reading interventions across languages, a review of digital reading interventions by McTigue et al. (2020) suggested that these effects may be moderated by contextual factors such as participant engagement and goal orientation, underscoring the need for more nuanced evaluations that explicitly account for such moderators and the conditions under which digital interventions are most effective.

Beyond domain-specific interventions, other digital approaches have been developed to target cognitive skills, with a focus on the downstream effect of enhancing reading outcomes. Because some reading deficits have been linked to impaired attentional control (Bosse et al., 2007; Facoetti et al., 2006; Grainger et al., 2016; Valdois et al., 2019), implementing adaptive trainings for attention and related areas has been hypothesized to positively impact reading abilities (Debeer et al., 2021; Peters et al., 2019, but see Cartwright & Palian, 2024). For example, some research indicates that children with dyslexia who received training using a commercial action video game experienced improvements in both attentional control and reading speed, without any detriment to reading accuracy (Franceschini et al., 2013, 2017; Franceschini & Bertoni, 2019, but see, Łuniewska et al., 2018). Other work suggested that training attentional control and executive functions through a kid-friendly action-based video game can have promising effects on reading skills in general readers, at least in Italian speaking children (Pasqualotto et al., 2022). In addition, prior research has illustrated that training executive functions, such as working memory, may also improve reading abilities in both poor and general readers (Dahlin, 2011; Karbach et al., 2015; Loosli et al., 2012; St Clair-Thompson et al., 2010; Studer-Luethi et al., 2016). As reviewed by Titz and Karbach (2014), training working memory and other EFs improve reading skills in students of all reading levels, yet with variations depending on students age and reading skill level as well as the orthographic characteristics of the reading language. In particular, the effect of domain-general training may be strongest in more transparent orthographic systems such as Italian, where many of these studies have been carried.

Overall, evidence for domain-general digital training is relatively scarce and remains a topic of debate, with most studies focusing on poor readers and transparent languages (Peters et al., 2019; Puccio et al., 2023). While training executive functions can improve executive functioning, the evidence for transfer to reading or other academic tasks remains mixed and inconsistent (e.g., Diamond & Ling, 2016; Melby-Lervåg et al., 2016).

Moreover, despite promising findings, many digital reading interventions still lack a robust theoretical foundation (Savage et al., 2013), and the literature reflects considerable variability in design quality and implementation. The relative novelty and diversity of these tools has resulted in a proliferation of studies—many of which rely on small sample sizes, non-randomized designs, or limited follow-up (Ostiz-Blanco et al., 2021). These challenges, however, underscore the need for systematic syntheses to clarify what types of digital interventions are most effective, for whom, and under what conditions. In this context, meta-analytic approaches can play a crucial role in organizing the existing evidence, identifying promising directions, and highlighting gaps that should guide future research. Indeed, to determine whether digital tools can leverage their unique affordances for reading intervention—as originally envisioned (Santoro & Bishop, 2010)—it is essential to evaluate not only their average effects but also the moderating factors that shape their efficacy.

Study Purpose and Research Questions

This study presents two meta-analyses evaluating the effects of digital interventions on reading skills (decoding and/or reading comprehension) in children from early reading age to mid-adolescence, considering separately children with poor reading skills (including those with dyslexia) and the general school population. The dataset and code used to reproduce the analyses are publicly available via our meta-analysis app built under the Shinyapps.io (see https://sites.google.com/view/4reviewers/home).

Primary Research Questions

- ***R1***: What are the overall effects of digital interventions on reading skills, across quasi-experimental and experimental studies? This includes studies with at least one control group (either active or passive). Digital-based interventions are defined as those involving individual interaction with software, applications, or digital platforms accessed via personal computers, tablets, or gaming consoles.
- ***R2***: What is the impact of intervention type (domain-specific vs. domain-general) on the overall effects of intervention?
- ***R3***: What is the impact of language characteristics (i.e., syllabic complexity: simple vs. complex) of the spoken language, on overall intervention effects?
- ***R4***: Do intervention effects differ based on reading outcome type (decoding vs. Reading comprehension)?

This work also investigates a few other exploratory questions that have yet to be adequately addressed in the literature despite being of known importance, including training duration, the number of tasks and skills trained, the adaptivity of intervention content to user performance, the type of control intervention used, and participant characteristics such as age and pre-training reading skills.

These objectives align with emerging hypotheses about factors influencing reading intervention efficacy (Verhoeven et al., 2019) and leverage the growing potential for personalized digital tools to support reading skill development both within and beyond classroom settings. Importantly, the present analyses offer an initial meta-analytic assessment of the effectiveness of various digital intervention methods for enhancing reading skills in both poor readers and the general school population, thereby providing critical insights into the most effective approaches for supporting these distinct learner groups.

Method

Literature Search

A review of the literature was carried out in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses—PRISMA guidelines (Page et al., 2021) to ensure transparency and reproducibility. Specifically, PRISMA presents a flow diagram that consists of four phases: Identification, Screening, Eligibility, and Inclusion. Each of these steps was carried out in accordance with the PRISMA guidelines. Figure 1 provides a detailed PRISMA flow diagram summarizing the process, from initial database searches to final study inclusion.

Search Terms and Databases

The literature search covered the period between January 2000 and December 2022 by means of the following databases: Pubmed, PsycINFO, PsycARTICLES; PubPsych (including PsyINDEX, ERIC, MEDLINE), Web of Science, ScienceDirect, and Google Scholar. The search strategy combined terms related to digital tools, intervention, reading outcomes and learner populations (e.g., “cognitive intervention AND reading comprehension AND children”). Boolean strings and modifications for each database are detailed in Supplementary Information (SI) Sect. 1 - Search Strategy. The search was limited to papers written in one of the following languages: English, Spanish, Italian, French, and German due to the authors’ proficiency in these languages, but studies conducted in any alphabetic language were included. To gather gray literature, or studies not published in peer-reviewed journals, OpenGrey, ProQuest Dissertations and Theses Global were queried using the same search terms and strategy described in the SI. Authors known to have worked on the topic were also contacted to identify unpublished data. In addition to database queries, the authors’ lab groups were asked for recommendations of papers that met the inclusion criteria and may not have been retrieved, and authors of the included papers were asked if they had any papers that met the inclusion criteria. Finally, the reference lists of all the documents that passed the inclusion criteria were also consulted in accordance with the snowball principle.

This initial search, combining the results of all databases (including gray literature repositories) and languages as well as all keyword combinations described above, yielded 8,320 results, which, after removal of duplicates, yielded a total of 6,667 papers for initial screening (see Fig. 1). An additional 10 papers were retrieved via non-database methods (e.g., author recommendations and reference checks).

Eligibility Criteria

Eligible studies were group-designed randomized controlled trials or quasi-experimental trials (e.g., assignment based on the school/class level) which compared the effects of digital interventions targeting either domain-specific or domain-general skills to at least one control group (active

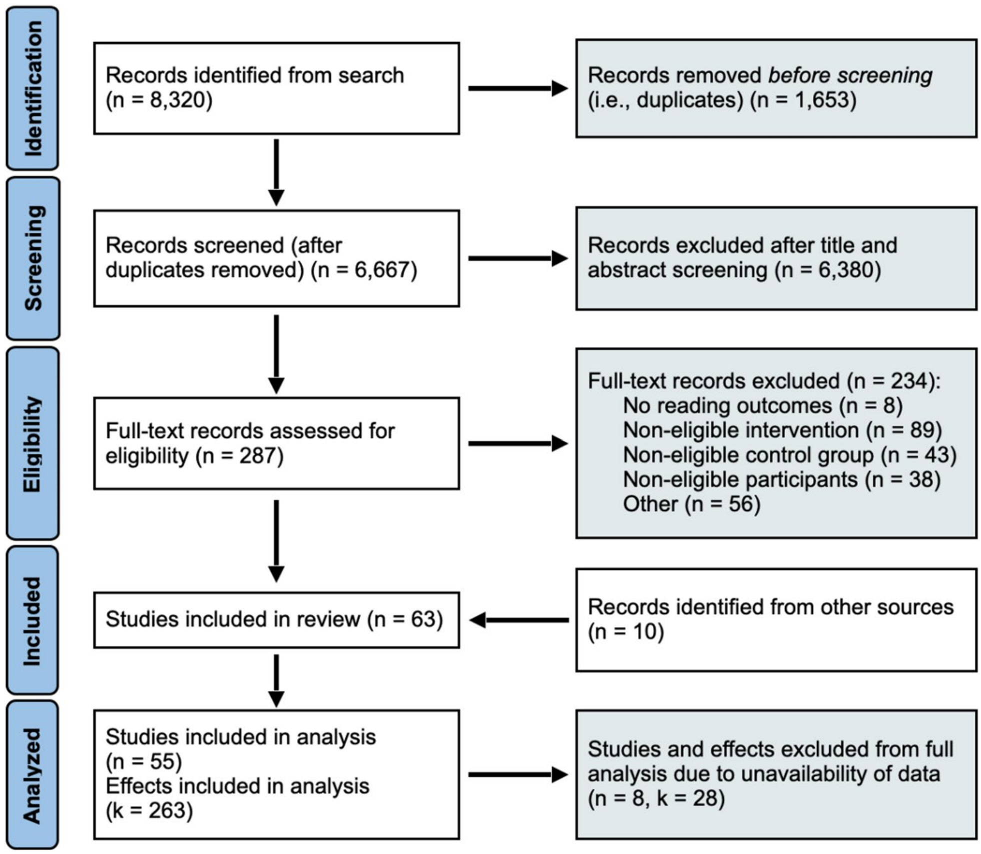

Fig. 1 PRISMA flow diagram of systematic search and study selection. This diagram illustrates the systematic review process, showing the number of studies identified through database searches, screened for eligibility, excluded at each stage with reasons specified, and ultimately included in the final analysis. Detailed reasons for exclusions are provided in Table S2

or passive; see Table S1 in the Supplementary Information for a description of the types of controls included) on reading outcomes. Papers regarding usability testing alone were excluded, as well as literature reviews, opinion pieces, single case interventions, observational reports, or newsletters.

Only studies that reported participants' performance in either decoding and/or comprehension as their dependent variable were included. Furthermore, the studies needed to include statistical information (sample size, mean, standard deviation before and after intervention) to allow for effect size calculations. When only part of this data was reported, contact was made with the corresponding authors to obtain the missing data. Inclusion and exclusion criteria are summarized in Table 1, with detailed descriptions available in Supplementary Information (SI) Table S1. Excluded studies with documentation as to why they were excluded, are outlined in SI Table S2.

Screening

Articles were screened for eligibility by five independent reviewers (AP, EC, CH, MM and BZ). The two master students (MM and BZ) were involved primarily in the first step of the screening process (i.e., title and abstract screening). All reviewers were trained by AP to identify study aspects that would warrant exclusion. All studies were coded independently by two reviewers, with discrepancies resolved through consensus. Agreement between reviewers was high (99.5%) for full-text inclusion decisions. In the cases of disagreement ($n = 31$), papers were discussed in group meetings with at least three authors until consensus was reached. Beyond percent agreement, no additional inter-rater reliability statistics (e.g., Cohen's kappa) were calculated, as consensus meetings ensured thorough resolution of discrepancies and alignment on final ratings. After titles and abstracts were screened and were agreed to meet the inclusion criteria, papers were read thoroughly to extract variables and moderators of interest. If, after full-text screening, one or more inclusion or exclusion criteria were not met, the paper was then excluded from further review. The reader is referred to Table S2 for examples of excluded studies and their reasons. When studies met inclusion criteria, but insufficient information was provided to extract effect sizes, corresponding authors were identified and emailed with a request to provide the missing information. When the authors did not provide the requested data (8 studies), studies were excluded from effect size analysis but retained in the counts of studies included in the review (see Table S3 for a summary of these studies).

Table 1 Summary of inclusion and exclusion criteria (see table S1 for full details)

Criterion	Inclusion	Exclusion
Population	Children aged 6–14 years, general and poor readers (including developmental dyslexia).	Pre-readers, non-alphabetic language users, children with unrelated disorders (e.g., ASD, ADHD).
Intervention	Digital interventions targeting reading domain-specific or domain-general skills (tablet, computer).	Eye/gaze interventions, hybrid (digital and non-digital), interventions on non-native languages.
Control Group	Active (non-reading digital tasks, math activities) or passive (classroom instruction, waitlist).	No control group, both groups receiving eligible interventions.
Outcomes	Quantitative measures of decoding (word, non-word, text) or written comprehension.	Phonological awareness, RAN, or decoding of isolated letters.
Study Design	RCTs and quasi-experimental studies with pre- and post-tests, > 10 participants.	Case studies, post-tests > 1 month after intervention.
Limits	Published 2000–2022 in English, Spanish, Italian, French, or German.	Non-alphabetic languages, published outside the defined time range.

RCT randomized controlled trial, *ASD* autism spectrum disorder, *ADHD* attention deficit hyperactivity disorder, *RAN* rapid automatized naming

Coding Procedure

Included studies were coded for study descriptors and moderators. Once all articles were first coded, the list of moderators and study descriptors was further refined, producing a final systematic coding scheme by which all included papers were cross-checked by a minimum of two independent authors. The studies underwent coding with *reading performance measured before and after intervention* being tabulated with sample sizes, means and standard deviations being extracted for effect size calculations (***primary outcome***). The following characteristics were also considered:

(i) ***Participant characteristics.*** age, gender distribution, reading skills at baseline, qualification of participants as poor or general readers (see SI Table S1 for description of the criteria used to determine reader group status).
(ii) ***Moderators***: (i) Intervention Type – Domain-specific vs. Domain-general (ii) Language characteristics (syllabic complexity) – Simple vs. Complex; (iii) Reading outcome – Decoding vs. Comprehension.
(iii) ***Primary exploratory variables*** include the type of control used, categorized as active (such as non-digital interventions or digital activities unrelated to reading or executive functions) or passive (no contact or waitlist controls). The intervention duration was measured in total hours of training and the length of each session.
(iv) ***Secondary exploratory variables*** examined the type of device used, whether a tablet, computer, or game console. Intervention scaffolding was assessed by determining whether difficulty was adaptive (automatically or manually adjusted based on individual participant performance), non-adaptive (progressed in a scaffolded manner, such as through preset levels), or fixed (unchanged throughout the intervention). The number of tasks within the intervention, as well as the number of skills trained, were noted, with skills categorized as domain-specific (such as reading) or domain-general (such as working memory).

Study characteristics and moderators were systematically coded using a pre-defined scheme as per PRISMA guidelines (see SI Table S4). There was no missing information for the primary moderators of interest, as sufficient details were consistently available to categorize studies along these dimensions: (a) participant type, coded as poor readers versus general readers (with only studies explicitly reporting a dyslexia-based inclusion criterion coded separately as poor readers–dyslexia for descriptive purposes only); (b) intervention type, coded as domain-general versus domain-specific; (c) language characteristics, coded according to Seymour et al.'s (2003) classification; and (d) outcome type, coded as decoding or comprehension. Each moderator was coded based on the descriptions provided in the original publications. For exploratory moderators, when information was insufficient, study authors were contacted for clarification. If no clarification was obtained, the study was retained in the overall analyses but coded as 'NA' for the relevant variable, and therefore excluded only from analyses involving that moderator. Missing data were relatively infrequent and did not affect the primary moderators, though they may have reduced statistical power for some exploratory analyses.

Quality Assessment

The methodological quality of individual studies was evaluated using the Physiotherapy Evidence Database (PEDro) scale. Although originally developed for physiotherapy and medical trials (Maher et al., 2003; De Morton, 2009), the PEDro scale has also been adopted in broader intervention research because it provides a transparent, structured checklist of methodological rigor that can be easily applied across randomized and quasi-experimental designs (Verhoeven et al., 2019).

We selected the PEDro scale because its quantitative scoring system facilitates comparability across multiple studies, which is advantageous in meta-analysis. Alternative tools such as the Cochrane Risk of Bias (RoB 2; Higgins et al., 2019) or the Quality Assessment Tool for Quantitative Studies (QATQS; Thomas et al., 2004) offer more detailed domain-based assessments, but are less suited for deriving summary scores that can be systematically integrated across trials. PEDro was therefore judged to provide an appropriate balance between methodological rigor and analytic feasibility in the present context, although it does not directly address aspects such as treatment fidelity or implementation quality that are often emphasized in educational research.

Using a 0–11 scale, quality is determined by assigning one or zero points to each of the following factors: randomized allocation of participants, concealed allocation of participants, comparable baselines, blinding of participants, therapists and assessors, completion of key outcome measurements by > 85% of participants, intention-to-treat analysis, between-group statistical comparisons, and measures of variability for at least one key outcome. Afterwards, studies were divided into three groups based on the total PEDro score (following Foley et al., 2003): high risk of bias (≤3 points), medium risk (4–7), or low risk (≥8).

The quality assessment was carried out by two independent reviewers (CH and AP). Any disagreements in scoring were resolved by consensus between the two independent reviewers. Detailed scoring for each study is provided in SI Tables S7 and S8.

Meta-analysis

Effect Sizes

For each included study, we quantified the difference in performance changes from pre- to post-training between an experimental group and a control group using a measure of effect size and variance.

We first computed Cohen's *d* using the formula for independent groups with pre–post design (Borenstein et al., 2009; Eq. 12.19–12.22, p. 227). Effect sizes were then converted to Hedges' *g* by applying the J correction factor, which adjusts for positive bias in small samples and thus provides more accurate estimates (Hedges, 1981). This yields a standardized measure of how strongly the intervention group outperformed the control group, where values around 0.2 are typically considered small, 0.5 medium, and 0.8 large (Cohen, 1988). The formula used is given below with the first half corresponding to Cohen's *d* and the 2nd half being the J factor to convert into Hedge's *g*.

$$g = \frac{(M_{post,\,T} - M_{pre,\,T})}{SD_{pooled,pre}}\; x \;\left(1 - \frac{3}{4\,x\,(N_T + N_C) - 9}\right)$$

$M_{pre,\,T}$, $M_{post,\,T}$ and $M_{pre,\,C}$, $M_{post,\,C}$ are respectively the pre, post mean scores of the treatment group and the pre, post mean scores of the control group. $SD_{pooled,\,pre}$ is the pooled pre-test standard deviation, computed considering only the pre-test standard deviation of the two groups, and N_T, N_C are the sample sizes of the treatment and control groups. The formula for the pooled standard deviation of the pretest and for the effect size variance (Vg) are given below.

$$SD_{\text{pooled,pre}} = \sqrt{\frac{(N_T - 1)\times\ SD_{pre,T} + (N_C - 1)\times\ SD_{pre,C}}{(N_T +\ N_C - 2)}}$$

$$V_g = \left(2\times(1 - r_{pooled})\times\frac{N_T+N_C}{N_T\times\ N_C} + \frac{d^2}{2\times\ (N_T\times\ N_C)}\right)\times\left(1 - \frac{3}{4\times\ (N_T+N_C)-9}\right)^2$$

In some studies, the standard deviation of change scores is reported instead of the standard deviation of the pretest and postest. In such cases, we used the following formula to convert the pooled standard deviation of the change scores into pooled standard deviation of pretest.

$$SD_{\text{pooled,pre}} = \frac{SD_{diff,\,T}}{\sqrt{2\times\ \ (1 - r_{pooled})}}$$

Estimating effect sizes and their variance with this method requires knowledge of the pre-post correlation which is rarely reported. To overcome this, we followed recommendations from the What Works Clearinghouse (What Works Clearinghouse Procedures Handbook, Version 4.1, 2020; see also Morris & DeShon, 2002) and used a correlation of 0.5 between pretest and posttest. This is equivalent to the d_{ppc2} from Morris et al. (2008) which was shown to perform better than all alternatives in terms of bias, precision, and robustness to heterogeneity of variance, and is thus likely to provide the most conservative estimates. In other words, this method minimizes the risk of overstating intervention effects while still retaining statistical accuracy.

When multiple tasks or measures were extracted, we computed an effect size for each and modeled the dependence structure between effect sizes from the same study using correlated and hierarchical meta-analytic models described below. This approach avoids 'double-counting' participants by explicitly modeling the fact that multiple measures from the same children are more similar to each other than to measures from different studies.

Statistical Analysis

All analyses were conducted in R using the metafor (Viechtbauer & Viechtbauer, 2015) and clubSandwich (Pustejovsky & Tipton, 2017) packages, as well as robumeta (Fisher et al., 2016). Our approach combined a multilevel meta-analytic model with robust variance estimation (RVE; Hedges et al., 2010; Tipton, 2015). A common feature of computerized intervention studies is to include several pretest and posttest measures in the same participants (e.g., decoding, fluency, comprehension). This results in clusters of related, non-independent effect sizes, which if treated as independent would artificially inflate the estimation of the averaged effect size. While it is possible to overcome this by either randomly selecting one effect or by aggregating non-independent effects from a given study, these approaches do not fully exploit the available data. Here, we use robust variance estimation to retain all available measures while properly accounting for their statistical dependence. Because effects are often both correlated (due to coming from the same participants) and hierarchically organized (e.g., two reading measures from the same child are more alike than measures from another study), we used the correlated and hierarchical effects model (CHE; Pustejovsky & Tipton, 2022) to model the dependence structure of the effect size data. Because the correlation between effect sizes from the same participants is typically unreported, we used the default value of rho = 0.8, reflecting the expectation that outcomes within the same participants are strongly related and following standard practice in meta-analysis when empirical values are unavailable.

We estimate standard errors and conduct hypothesis tests using robust variance estimation (RVE) which provides more accurate estimates than other alternative methods (e.g., chi-square, Q-statistics or meta-regression coefficients), even in the presence of model misspecifications. Small sample correction was applied to all tests.

For each meta-analysis, we first estimated the average effect size using an intercept-only multivariate meta-analytic model with the restricted maximum likelihood method. This approach provides estimates of total heterogeneity (i.e., between-study) and between clusters (i.e., within-study). Sources of heterogeneity were examined by adding moderators to the model and running publication bias analyses.

For moderator analyses, the intercept was removed, and the following moderators were added: intervention type (domain-general, domain-specific), syllabic complexity (simple, complex) and outcome type (decoding, reading comprehension). Contrasts between moderator levels were evaluated using the Wald F-test and the Approximate Hotelling's T-squared (AHT) approach (Pustejovsky & Tipton, 2018), followed by t-tests to assess the significance of each moderator level (i.e., against zero). This was achieved by changing the moderator of interest in the model – for example, modeling intervention type while controlling for the other two moderators (e.g., syllabic complexity and outcome type), and then using a loop across each moderator. In doing so, each moderator effect is assessed while correcting for the relative frequency of the reference levels of other moderators included in the model, thereby ensuring an accurate estimate for the relative distribution of all moderators in the present dataset.

Publication bias was assessed by testing funnel plot asymmetry with an alternative of Egger's Regression which combines meta-regression with robust variance estimates (Rodgers & Pustejovsky, 2021). Using a modified estimate of effect-size precision, this test outperforms other methods while maintaining type I error rate in the presence of dependent or sampled effect sizes, even when the dependence structure is unknown or mis-specified. This test relies on a modified covariate of effect size precision to Egger's test, estimating its slope using weighted least squared while accounting for dependent effect sizes with the CHE model (Pustejovksy & Tipton, 2022). This analysis was performed both with and without moderators. While the Egger's test performs well at detecting the presence of bias, the intercept does not provide an accurate estimate of the bias-adjusted effect size. For publication-bias correction, we thus relied on the 3-parameter selection mode (Vevea & Hedges, 1995), which provides great flexibility in estimating the effect sizes adjusted for publication bias. However, because the selection model methods (e.g., 3PSM) is unable to handle dependent effect sizes we applied the 3PSM method via bootstrapping. First, we randomly selected one effect size per study and repeat this cluster bootstrapped 1000 times, to calculate the mean effect and mean variance across the distribution of repetitions. For completeness we report the detection and correction for both Egger's CHE and 3-PSM methods. Additional analyses to detect publication bias were conducted for comparability with previous meta-analyses and to continue tracking publication bias as this literature grows (e.g., Trim & Fill; Duval & Tweedie, 2000; PET-PEESE; Stanley & Doucouliagos, 2014). These additional results are reported in the supplementary materials.

Results

Among the 63 distinct studies that fit our inclusion criteria, 55 studies were eligible for inclusion in the calculation of effect sizes. It should be noted that one study (Szili et al., 2022) is included in both meta-analyses due to its reporting of data for both Poor and General readers separately. This data set and corresponding code under shinyapps.io is available at MetaAna_app on ShinyApp. For a detailed presentation of the publication bias results and the qualitative analyses of risk of bias using the PEDro scale, we direct readers to SI Sect. 3.1 and 3.2, respectively. Additionally, for supplementary analyses on primary and secondary exploratory variables, please refer to SI Sect. 3.3. Below, we provide information on the number of effect sizes (k), clusters (m), and degrees of freedom (*df*) for each RVE analysis performed. Two meta-analyses were conducted, one focusing on poor readers and one on general readers.

Meta-Analysis with Poor Readers

The data set of poor readers included 41 studies and $k=194$ effect sizes, with a mean age of 9.30 years (SD = 1.68, range = 6–12.07) and a mean intervention duration of 12.7 h (SD = 11.5, range = 1–50). Across the 41 included studies the sample was 40% female, with 25 studies describing poor readers in general and 16 (38%) describing readers with an official dyslexia diagnosis.

The multilevel robust meta-analytic model showed a medium mean effect size ($g=0.43$, CI = [0.30, 0.56], $p<.001$, $k=194$), meaning that on average poor reading children in the digital intervention groups showed a greater improvement in reading outcomes than children in the control groups (Fig. 2). The prediction interval ($PI=g\pm1.96\sqrt{(\tau^2+\omega^2)}$), which gives the effect size predicted for a newly published study with poor readers and the same eligibility inclusion as this meta-analysis, ranges between −0.44 and 1.3. The effect sizes of the included studies are quite heterogeneous. Specifically, heterogeneity is high between studies ($w^2=0.185$, $Q(193)=1366.71$, $p<.001$) but not as high within-studies ($\tau^2=0.01$).

As Table 2 illustrates no publication bias was detected. The robust Egger's CHE test did not detect any significant asymmetry in the funnel plot (Fig. 3), either in the intercept-only model model (i.e., no moderators included, $\beta=-0.01$, SE = 0.60, $t(14.0)=-0.01$, $p=.10$) or in the full model ($\beta=-0.34$, SE = 0.77, $t(17.0)=-0.44$, $p=.67$); yet the bias-corrected effect obtained from 3-PSM produced a significant positive effect when applied to the intercept-only model ($g=0.51$) and a positive but nonsignificant effect with the full model ($g=0.60$). Additionally, visual inspection of the funnel plot (Fig. 3) showed a largely symmetrical distribution of effect sizes across studies by precision, providing no indication of systematic publication bias beyond the results of Egger's test. The SI-Sect. 3 presents a comprehensive publication bias sensitivity analysis highlighting the diverging results obtained with distinct detection and correction techniques, especially the PET/PEESE which is known to provide a biased estimate (Carter et al., 2019; Stanley, 2017).

Poor Readers - Moderators Analyses

Overall, the moderators included did not explain any of the variance in effect sizes (−6.48% with residual heterogeneity = 0.21 - the negative estimate of variance-accounted-for effect size is likely due to both the population effect and the sample size being small, while the between-study heterogeneity is high). We computed Approximate Hotelling's T-squared (AHT-F) tests to assess the effect of the primary moderators. We discuss below each moderator separately, reminding the reader that all moderators were included in the model (Table 3).

Intervention Type

No significant effect of intervention type on performance is observed (AHT-F = 0.00, $p=.10$), due to a medium effect size for both digital domain-specific ($g=0.44$, CI = [0.29, 0.59], $p<.001$, $k=167$, m = 34); and digital domain-general interventions ($g=0.44$, CI = [−0.17, 1.047], $p=.13$, $k=27$, m = 7). Note the large CI for domain-general interventions likely due to the limited number of studies in this cell (m = 7).

Language Characteristics – Syllabic Complexity

The effect of syllabic complexity is also non-significant (AHT-F = 1.97, $p=.17$, $k=194$, m = 41). A medium effect size of syllabic complexity is observed for both simple languages ($g=0.35$, CI = [0.17, 0.52], $p=.001$, $k=98$, m = 18) and complex languages ($g=0.54$, CI = [0.31, 0.76], $p<.001$, $k=96$, m = 23). These analyses included five alphabetic languages (English, German, Italian, Spanish, and French). Most studies in English and German focused on decoding and fluency outcomes, whereas those in Italian, Spanish, and French also assessed comprehension.

Reading Outcome Characteristics

Whether reading performance was measured via decoding or via comprehension measures led to comparable effects (AHT-F = 1.59, $p=.23$, $k=194$, m = 41), even if numerically decoding ($g=0.48$, CI = [0.31, 0.65], $p<.001$, $k=151$, m = 34) led to a medium effect and comprehension to a small one ($g=0.30$, CI = [0.06, 0.54], $p=.016$, $k=43$, m = 18).

Meta-Analysis with General Readers

The data set of general readers included 15 studies and $k=69$ effect sizes, with a mean age of 9.01 years (SD = 1.76, range = 6.4–12.5) and a mean intervention duration of 16.5

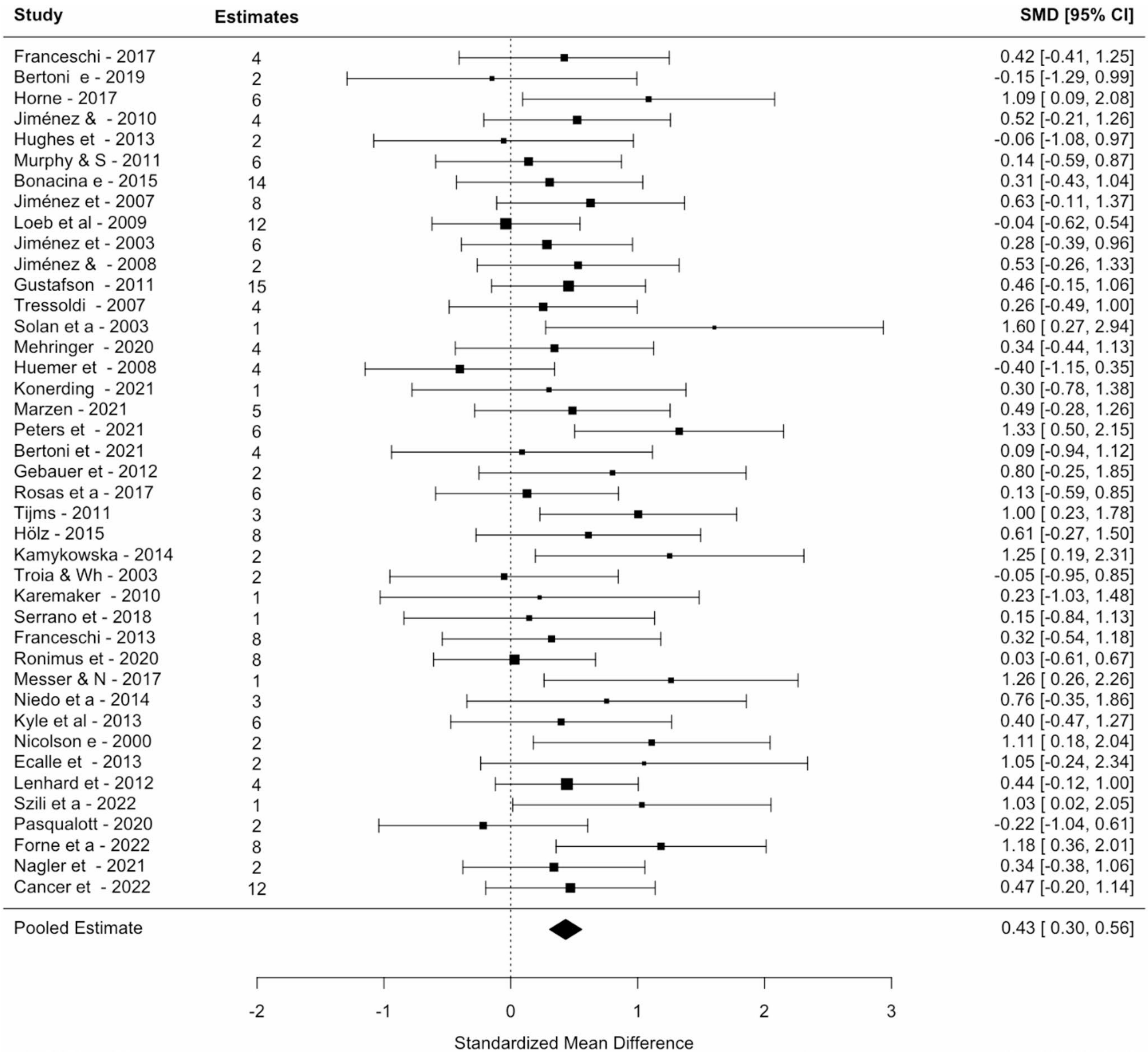

Fig. 2 Poor readers: Forest plot of aggregated effect sizes of the impact of computerized training on reading outcomes. Note. Each square shows the aggregated (averaged) effect-sizes from all dependent effects. The size of each square is proportional to the inverse variance (larger square = lower variance). The diamond shows the overall averaged estimate from the intercept-only model. Digits in the 2nd column next to the study ID indicate the number of aggregated effects

h (SD = 17, range = 2–72). Across all studies, 49% of the sample was female.

The overall multivariate robust meta-analytic model showed a small mean effect size (g = 0.26, CI = [0.08, 0.43], p = .01, k = 69, m = 15), meaning that on average children in the digital intervention groups showed an improvement in reading outcomes of about 1/4th of a standard deviation as compared to children in the control group (Fig. 4). The prediction interval (PI) indicated that for a newly published study (with general readers and eligible for inclusion in this meta-analysis), an effect size between −0.48 and 0.99 (i.e., $g \pm 1.96\sqrt{(\tau^2+\omega^2)}$) is expected. Overall, the effect sizes of the included studies are quite heterogeneous. Specifically, heterogeneity is high between studies (ω^2 = 0.01) but is not as high within-studies (τ^2 = 0.14; Q(68) = 586.83, p <. 001).

Again, no significant publication bias was found with the robust Egger's test using the CHE model (see Fig. 5; Table 4), whether in the intercept-only model (i.e., no moderators included, β = 1.07, SE = 0.94, t(6.52) = 1.15, p = .29) or in the full model with all primary moderators (β = 1.62, SE = 0.90, t(4.74) = 1.80, p = .14). Additionally, visual inspection of the funnel plot (Fig. 5) showed a largely symmetrical

Table 2 Poor readers: results of publication bias analysis

Method	Model	b	CI	t	dfs	*p*
Detection						
Egger CHE (slope)	NULL	−0.022	[−1.31, 1.265]	−0.037	14.0	0.971
	FULL	−0.350	[−1.982, 1.281]	−0.453	17.0	0.656
3-PSM	NULL			0.772		0.533
	FULL			0.525		0.599
Correction						
Egger CHE (intercept)	NULL	0.441	[−0.005, 0.88]	2.137	12.9	0.052
	FULL	0.526	[−0.496, 1.547]	1.119	12.2	0.285
3-PSM	**NULL**	**0.506**	**[0.308, 0.709]**			**0.004**
	FULL	0.592	[0.249, 0.951]			0.097

b is the beta coefficient (slope for detection methods or intercept for correction methods); CI is the 95% confidence interval; Stats corresponds to the t-test for intercepts and Likelihood Ratio chi-square test for 3-PSM; *df*s are the degrees of freedom; *p* is the *p* value. NULL refers to an intercept-only model; FULL refers to the model including all moderators

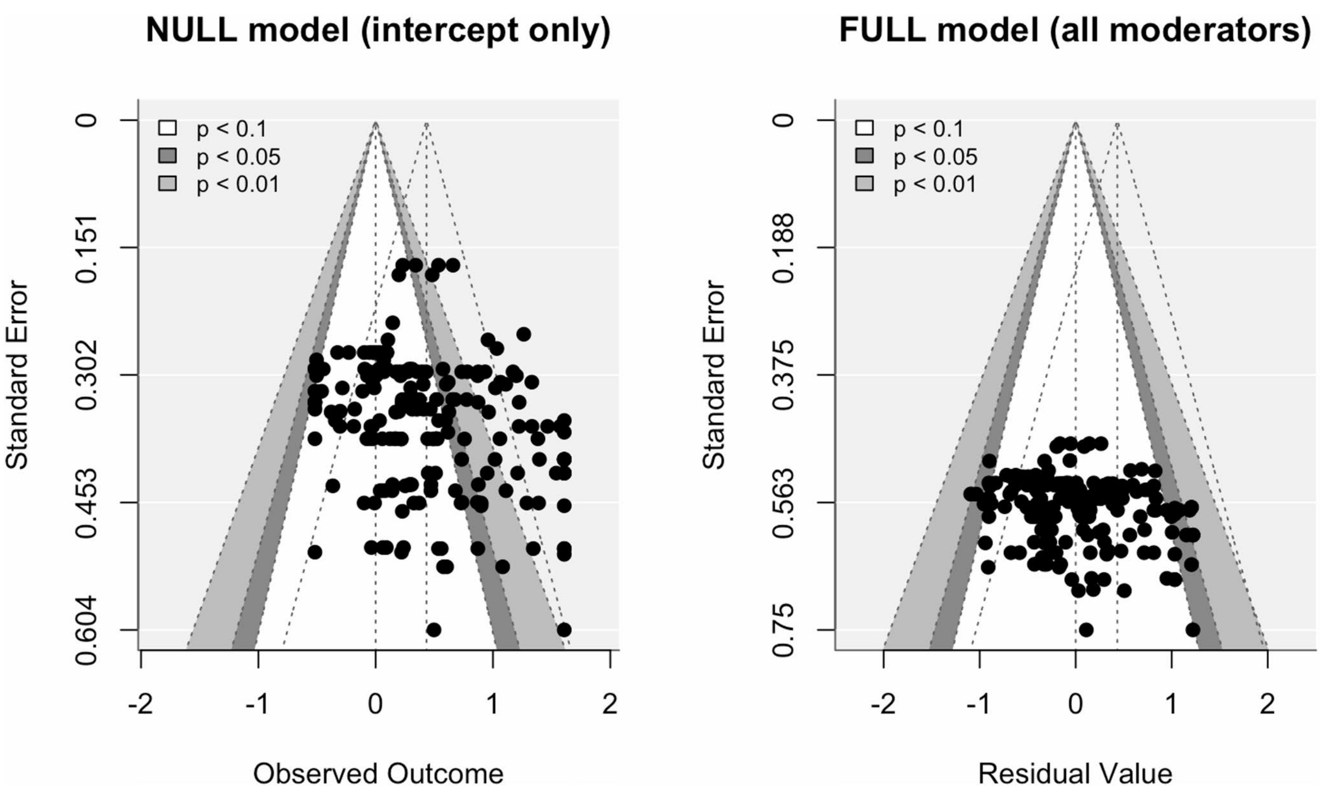

Fig. 3 Funnel plots: Each black dot represents one study included in the meta-analysis for Poor Readers

distribution of effect sizes across studies by precision, providing no indication of systematic publication bias beyond the results of Egger's test. Together these results suggest that the sample of studies does not suffer from publication bias issues. As shown in Table 3, the corrected average effects obtained from 3-PSM were non-significant with both intercept-only model ($g=0.11$) and the full model ($g=-0.24$). Sensitivity analysis with different detection and correction techniques are presented in SI-Sect. 3.

Moderators Analyses

For general readers (Table 5), the moderators included explain 11.21% of the variance in effect sizes (residual heterogeneity = 0.13). We computed Approximate Hotelling's T-squared (AHT-F) tests to assess the effect of the primary moderators. We discuss below each moderator separately, reminding the reader that all moderators were included in the model.

Table 3 Poor readers: summary table of moderator effects

Moderator	Level	Fstat	k	m	g	CI	df	p
No moderator	Intercept only		194	41	0.432	[0.3, 0.565]	33.955	0.000
Intervention type	AHT-F Test	0.00	194	41			7.352	0.997
	Domain-general		27	7	0.439	[−0.17, 1.047]	5.487	0.126
	Domain-specific		167	34	0.440	[0.29, 0.59]	26.870	0.000
Syllabic complexity	AHT-F Test	1.97	194	41			31.546	0.174
	Simple		98	18	0.345	[0.167, 0.524]	15.071	0.001
	Complex		96	23	0.536	[0.309, 0.764]	19.077	0.000
Reading outcome	AHT-F Test	1.59	194	41			12.457	0.231
	Decoding		151	34	0.480	[0.308, 0.652]	3.446	0.000
	Comprehension		43	18	0.299	[0.062, 0.536]	18.280	0.016

Each moderator summary includes the associated Approximate Hotelling's T² test (AHT F) statistics, and by-level Hedge's *g* values, in addition to k, m, 95% CI, *df* (per moderator level), and associated *p* values for each reported statistic

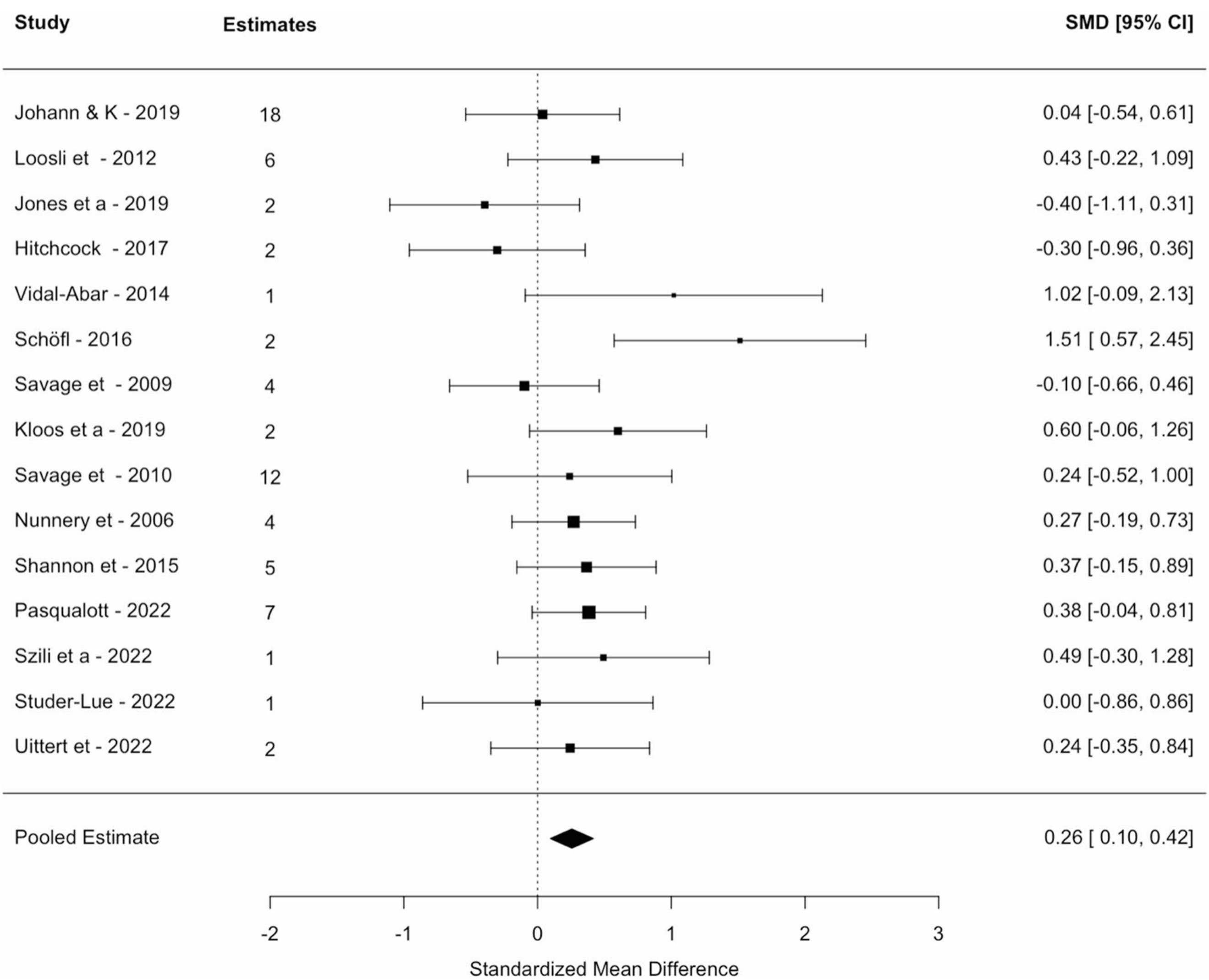

Fig. 4 General readers: Forest plot of aggregated effect sizes of the effect of digital intervention. Note. Each square shows the aggregated (averaged) effect-sizes from all dependent effects. The size of each square is proportional to the inverse variance (larger square = lower variance). The diamond shows the overall averaged estimate from the intercept-only model. Digits in the 2nd column next to the study ID indicate the number of aggregated effects

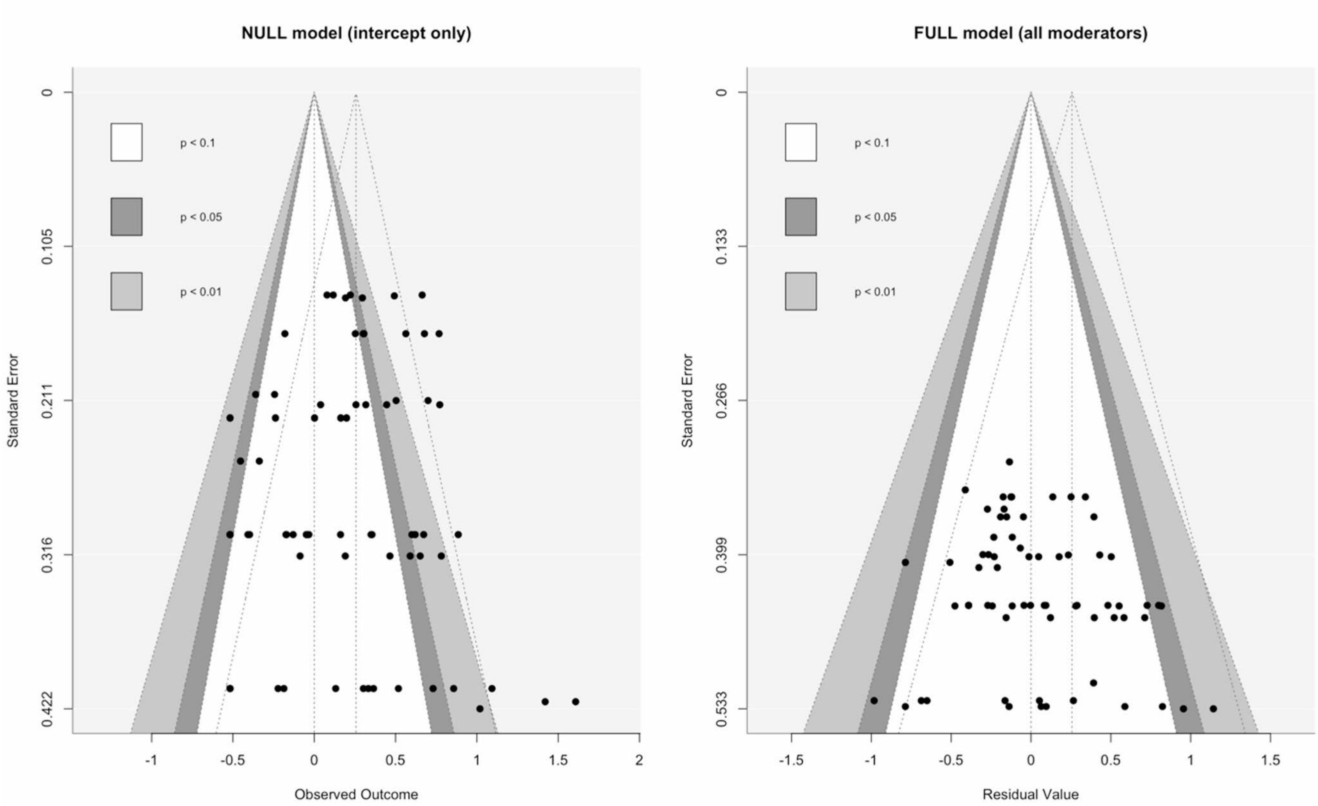

Fig. 5 Funnel plots: Each black dot represents one study included in the meta-analysis for General Readers

Table 4 General readers: Results of publication bias analysis

Method	Model	*b*	CI	*t*	*dfs*	*p*
Detection						
Egger CHE (slope)	NULL	1.071	[−1.176, 3.318]	1.144	6.52	.293
	FULL	1.623	[-.737, 3.982]	1.798	4.73	.135
3-PSM	NULL			2.265		.311
	FULL			5.248		.116
Correction						
Egger CHE (intercept)	NULL	.020	[-.439,.48]	.104	6.995	.920
	FULL	-.529	[−1.215,.157]	−1.898	5.844	.108
3-PSM	NULL	.107	[-.132,.333]			.481
	FULL	-.243	[-.401, -.084]			.117

b is the beta coefficient (slope for detection methods or intercept for correction methods); CI is the 95% confidence interval; Stats corresponds to the t-test for intercepts and Likelihood Ratio chi-square test for 3-PSM; *dfs*are the degrees of freedom; *p*is the *p*value. NULL refers to an intercept-only model; FULL refers to the model including all moderators

Intervention Type

A significant effect of intervention type on performance (AHT-F = 7.62, p=.02) is observed, due to a greater magnitude effect for digital domain-specific interventions: (g=0.40, CI = [0.14, 0.67], p=.01, k=33, m = 9) than for digital domain-general interventions: (g=0.01, CI = [−0.23, 0.25], p=.94, k=36, m = 6).

Language Characteristics – Syllabic Complexity

The effect of syllabic complexity is marginal (AHT-F = 1.01, p=.06). The low degrees of freedom (2.74) call for cautious interpretation and the use of a more conservative threshold of p<.01. The effect size for digital interventions was significant for syllabically simple languages such as Italian and Spanish (g=0.51, CI = [0.12, 0.89], p=.03, k=9, m = 3 – but

Table 5 General readers: summary table of moderator effects

Moderator	Level	Fstat	k	m	g	CI	df	p
No moderator	Intercept-only		69	15	0.256	[0.078, 0.435]	11.068	0.009
Intervention type	AHT F Test	7.62	69	15			8.925	0.022
	Domain-general		36	6	0.008	[−0.231, 0.247]	4.815	0.936
	Domain-specific		33	9	0.404	[0.135, 0.673]	6.505	0.010
Syllabic complexity	AHT F Test	9.96	69	15			2.741	0.058
	Simple		9	3	0.508	[0.124, 0.893]	1.570	0.033
	Complex		60	12	0.151	[−0.042, 0.343]	8.309	0.109
Reading outcome	AHT F Test	1.94	69	15			4.481	0.229
	Decoding		31	8	0.305	[0.03, 0.58]	7.393	0.034
	Comprehension		38	11	0.109	[−0.09, 0.308]	7.092	0.236

Each moderator summary includes the associated Approximate Hotelling's T^2 test (AHT F) statistics, and by-level Hedge's *g* values, in addition to k, m, 95% CI, *df* (per moderator level), and associated *p* values for each reported statistic

note the very low m), but not for more complex languages such as English, German, and French ($g=0.15$, CI = [−0.04, 0.34], $p=.11$, $k=60$, m = 12).

Reading Outcome Characteristics

Effect sizes were numerically larger for decoding outcomes ($g=0.30$, CI = [0.03, 0.58], $p=.03$, $k=31$, m = 8) compared to comprehension outcomes ($g=0.11$, CI = [−0.09, 0.31], $p=.24$, $k=38$, m = 11). However, this difference between outcome types was not significant (AHT-F = 1.93, $p=.23$).

Effects Size Distributions with Respect To Moderators

In both meta-analyses, the distribution of data across moderator categories was uneven (Fig. 6), with some levels represented by only a few studies, limiting statistical power and interpretability for those comparisons.

Discussion

The overarching objective of this meta-analytic work was to comprehensively assess the impact of digital interventions on reading performance, separately for poor readers and general readers. We incorporated potential moderating factors and publication bias as sources of heterogeneity.

This study distinguishes itself from prior reviews and meta-analyses on childhood reading interventions in several crucial respects. First, we show through our two separate meta-analyses a large impact of digital interventions in poor readers ($g=0.42$) and a small effect in general readers ($g=0.26$). Second, this study is among the few to include

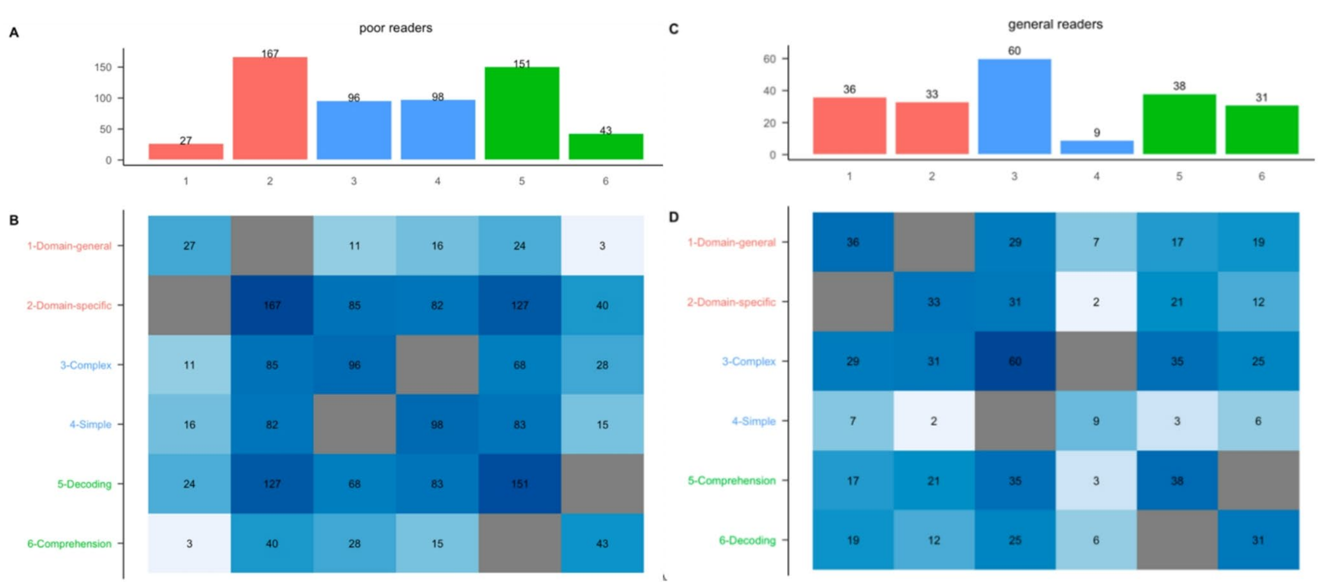

Fig. 6 Poor and general readers: Distribution of effect sizes across moderators. Note. *Poor Readers* - (**A**) Histogram of the number of effects in each level of each moderator. **B** Heat map showing the number of effects per moderator combinations. *General Readers* - (**C**) Histogram of the number of effects in each level of each moderator (**D**) Heat map showing the number of effects per moderator combinations.

and contrast domain-specific and domain-general digital interventions. Our findings highlight a robust impact of domain-specific interventions, with more variable impact from domain-general ones. Third, by considering decoding and reading comprehension as reading outcomes separately, our meta-analyses show a potential for impact on both outcomes, even if numerically the magnitude of the effect for decoding is about twice that for comprehension. Fourth, many syntheses akin to ours have overlooked an examination of intervention effects as a function of the language spoken, hindering the ability to draw conclusions about the generalizability of findings from English-speaking studies to other alphabetic languages with simpler syllabic structures (see, "Anglocentricity" issue noted in prior research; Share, 2008). Our results suggest a nuanced pattern by which numerically greater effects are noted in poor readers when using greater syllabic complexity languages such as English (g=0.54 vs. 0.34 for simple language) but, in general readers when using lesser syllabic complexity languages such as Italian (g=0.15 for complex languages vs. 0.51 for simple ones). Lastly, this study implements the stringent quality criteria for study inclusion that have been argued for in areas of psychology, cognitive science and educational research (Green et al., 2019; Johnson, 2021), In addition, it also relies on recent state-of-the-art multilevel and robust meta-analytic methods designed to handle the presence of dependent effects, both correlated and hierarchical dependencies, which are typical of interventions studies (Tipton et al., 2019; Tipton & Pustejovsky, 2015). In doing so, it represents one of the most statistically robust assessments of the impact of digital reading intervention one can generate using the latest advances in meta-analytic tools.

We discuss below in more detail these two distinct random effects meta-analyses of digital interventions and elaborate on the medium mean effect size for poor readers and the small mean effect size for general readers. While each moderator is considered separately, all moderators were entered simultaneously in the model.

Determinants of Digital Interventions for Poor Readers and for General Readers

Concerning domain-specific versus domain-general interventions, for poor readers, both intervention types led to benefits (Hedges' g of 0.43 and 0.44), while for general readers, only domain-specific interventions yielded significant benefits (Hedges' g of 0.4 versus 0.01). A note of caution is warranted regarding the latter given the very low number of contributing studies (m=6), which reduces statistical power and limits generalizability. Still, two mechanisms may account for this pattern. First, EF are less likely to be a limiting factor during reading acquisition in individuals who typically possess higher baseline cognitive and reading skills (i.e., general readers), leaving less room for improvement (a ceiling effect). Second, domain-general training may fail to translate into reading gains unless the targeted skills are closely aligned with reading processes—a limitation reflected in broader evidence on the lack of far transfer in working memory training (Melby-Lervåg et al., 2016). This pattern of findings likely echoes the significance of tailoring digital interventions to address specific reading needs. Future efforts could combine domain-specific and domain-general strategies adaptively, enabling personalized training that leverages the strengths of both approaches (Harrar-Eskinazi et al., 2022; Pasqualotto & Venuti, 2020; Verhoeven et al., 2019). Recent work by Cartwright and Palian (2024) supports this idea, in a meta-analysis of domain-general interventions for reading (including paper-and-pencil as well as digital ones), they found that domain-general interventions targeting EF that are specifically important for reading (g+=0.64) outperform domain-general ones (g+=0.59), likely due to better alignment between training and transfer tasks, as suggested by Ni et al. (2023). Thus, rather than contrasting domain-specific and domain-general interventions as is now typically done in the field, it would seem advisable in the future to combine them in an adaptative way to permit truly personalized training of each and every skill identified to mediate literacy acquisition, albeit to different extent depending on readers' profiles and the syllabic complexity of the target language. These advancements may provide critical support for reading attainment to all children, particularly when dedicated reading intervention time is scarce (Dean et al., 2021).

Concerning syllabic complexity, no statistically robust difference between levels of linguistic complexity was noted in poor readers. In general readers, the observed greater effects in syllabically simpler languages (e.g., Italian, Spanish) due to the very small number of contributing studies (m=3), which substantially limits statistical power and the strength of inference. Across the five alphabetic languages represented (English, German, Italian, Spanish, and French), most interventions in English and German emphasized decoding and fluency, whereas those in Italian, Spanish, and French also included comprehension outcomes. This imbalance limits the generalizability of our findings and underscores the need to broaden the range of languages studied, particularly by including non-English alphabets and non-European writing systems (Blasi et al., 2022). Nonetheless, the current findings align with prior work, which suggests promising outcomes of digital reading interventions in languages with simpler syllabic structures (Peters et al., 2019).

Interestingly, no significant differences were observed between decoding and reading comprehension outcomes, though decoding interventions showed numerically larger effect sizes. This result likely reflects the prevalence of interventions targeting decoding skills, which yield near-transfer benefits, compared to the broader demands of comprehension interventions (Klahr & Chen, 2011). Overall, the lack of differences by outcome is promising as the goal of all reading interventions remains reading comprehension, with decoding being only a steppingstone toward that goal. In future studies, the more systematic inclusion of both decoding and reading comprehension as outcome measures would be advantageous, ensuring that interventions are developed to cater to the multifaceted nature of reading skills and needs.

A key challenge when designing digital interventions concerns the duration of the training. Indeed, designing a one-hour experience is a very different design challenge than providing tens of hours of time on task. Here, the present work highlights the high heterogeneity of training duration. The studies examined in this analysis implemented training durations ranging from 1 to 70 h, with a mean training duration of 12.7 h for poor readers and 16.5 h for general readers. Most studies provided between 2 and 20 h of training (see SI Figure S7, panels A and B). This heterogeneity mostly reflects that the precise amount of training required to facilitate reading acquisition remains highly unclear in the literature (see SI Figure S7 panels C and D). Such heterogeneity highlights that the optimal training duration likely depends on several factors: (i) the population sample, with poor readers benefitting from shorter interventions due to greater room for improvement compared to general readers; (ii) the training content, whether domain-specific or domain-general; and (iii) the outcome measure, with decoding skills potentially requiring less time to improve than comprehension ones, which demands broader skill integration (Snowling & Hulme, 2011). Overall, the evidence suggests that digital interventions are most effective when conceptualized as delivering tens of hours rather than just a few, aligning with the observed trends across the reviewed studies.

Overall, comparing digital interventions to traditional paper-and-pencil approaches, our findings suggest similar levels of efficacy, particularly for poor readers. Meta-analyses of paper-and-pencil interventions targeting struggling readers have reported medium to large effects following various domain-specific interventions, with phonics-based interventions yielding effect sizes of $g=0.32$ to 0.52 (Galuschka et al., 2014) and other more encompassing domain-specific intervention yielding effect size of $g=0.49$ (Scammacca et al., 2015). For general readers, paper-and-pencil interventions tend to show smaller effects, with small to moderate effect sizes following domain-specific interventions such as training reading comprehension (Okkinga et al., 2018). Thus, digital tools achieve comparable reading outcomes to more traditional paper-and-pencil approaches while affording better adaptability and scalability of intervention.

Taken together, these findings underscore the importance of tailoring digital reading interventions to learner profiles and linguistic contexts. For poor readers, both domain-specific and domain-general approaches can be implemented with confidence: explicit reading instruction remains essential (Snowling & Hulme, 2011), but programs that also scaffold executive functions such as working memory and attention can provide added benefit (Cartwright, 2012). In practice, this means that shorter interventions can already deliver meaningful gains for struggling readers, given their larger margin for improvement, though extending training time is likely to maximize outcomes (Scammacca et al., 2015).

For general readers, by contrast, interventions should prioritize domain-specific components—particularly those that strengthen comprehension and fluency—since executive functions are less likely to represent a bottleneck once basic decoding is established (Perfetti, 2007). Developers and educators should also note that interventions may be more effective in syllabically simpler languages (Peters et al., 2019), although this remains a tentative finding in need of replication. Finally, for this group, programs should be designed with longer training durations (≥10 h) in mind, as shorter programs appear insufficient to produce reliable effects in learners with smaller gaps in reading proficiency.

Study Limitations and Future Directions for Digital Intervention

This study has several limitations that should be noted. The first set is common to many such meta-analytic works and concern the fact that, despite our best efforts, we were unable to include all possible studies within our inclusion criteria due to several general issues in the literature. First, despite contacting authors with work published in the field and other efforts, we were unable to collect any unpublished effect sizes or datasets apart from dissertations and theses. This is a notable drawback, as access to unpublished data is critical for accurately estimating effect sizes and capturing the full scope of the available evidence. Furthermore, data sharing practices play an important role in supporting replicability in research (Nuijten, 2019). Second, in some cases, we were unable to retrieve sufficient methodological information to determine the specific type of intervention used, even after contacting the original authors (see Table S4). Although our literature search was comprehensive—identifying studies even when technology use was not explicitly mentioned in the title or abstract—we cannot be certain that all eligible studies were captured. Third, it was not always clear whether participants had received an official diagnosis of developmental dyslexia or whether other comorbidities were present. In these cases, we relied on the authors' descriptions of their samples, though we recognize that a degree of ambiguity exists in classification practices across studies. To address this, we coded dyslexia-specific samples separately for descriptive purposes, but combined them with broader poor-reader groups in the main analyses given the limited number of dyslexia-only studies (m = 16) and probable overlap between groups.

Relatedly, the representation of orthographic systems across studies was uneven. Transparent alphabetic languages (e.g., Italian, Spanish, Finnish) were relatively more represented in studies of domain-general interventions, whereas opaque orthographies such as English were less common. This imbalance limits the generalizability of our conclusions to the full range of alphabetic systems and, in particular, to languages where decoding relies on different phonological and orthographic demands. Future research should therefore prioritize cross-linguistic validation across a wider set of writing systems, including both alphabetic and non-alphabetic scripts, to ensure that digital interventions can be appropriately tailored to diverse linguistic contexts (Ziegler & Goswami, 2005; Daniels & Share, 2018).

Another limitation concerns the relatively small sample sizes collected and reported in many studies, with a number of moderator analyses relying on fewer than five contributing studies (e.g., syllabic complexity in general readers, m = 3). While we used Hedges' g, which corrects for small-sample bias (Hedges, 1981), such limited numbers of studies—and the resulting low degrees of freedom in the RVE framework (often below 4)—still constrain the robustness and generalizability of moderator-specific findings, and may contribute to imprecise effect size estimates. This contrasts with paper-and-pencil interventions of reading that have been much better powered. This may be understandable as digital interventions first require investing in the design of a digital tool before testing its efficacy through a properly designed intervention study. High-quality intervention designs also require an active control group, and thus another contrasting digital tool being either available or developed. In the present reviewed studies, 29 studies did include such an active digital-based control, while 26 used only a passive, test-retest control group. Digital interventions require substantial investment in both digital design as well as in carrying out intervention studies. Yet, these two research areas require entirely different expertise. Although co-development between human-computer interface teams and reading experts has been increasing at a promising pace, the small number of participants found in existing studies so far mean there is much room to grow, with the current work not being necessarily representative of the entire population of interest (Munafò et al., 2017; Yarkoni, 2020). Hence, it is essential for future research to dedicate the resources and time to testing digital interventions in larger sample sizes.

Forthcoming research endeavors would also benefit from a more exhaustive examination of intervention programs, encompassing critical elements such as interface design, affordances, game mechanics, scaffolding techniques and the extent of human support integrated into these programs, as advocated by Green et al. in 2019). While we did our best to tabulate these different design choices, many papers lack detailed information on the design choices of their intervention programs (see SI Sect. 3.3). This gap is particularly relevant because much of the potential added value of digital tools—such as adaptivity, scalability, or enhanced engagement—depends on these very design choices. Yet, our analyses indicate that such affordances were neither systematically reported nor consistently implemented across the included studies. For instance, scaffolding techniques, particularly adaptive scaffolding, might play a crucial role in digital interventions by adjusting task difficulty based on user performance (Mayer, 2020). Our exploratory analysis (Figure S8) highlights the frequent use of adaptive scaffolding in digital interventions for poor readers, reflecting its importance in tailoring interventions to individual needs. In contrast, general readers showed a lower prevalence of adaptive approaches, with about half of the interventions using non-adaptive scaffolding. While our exploratory coding distinguished between adaptive (m = 42), non-adaptive (m = 10), and fixed-difficulty scaffolding (m = 6), the limited number of studies in the non-adaptive and fixed-difficulty categories precluded formal analysis of their relative impacts. It remains possible that the lower prevalence of adaptive scaffolding in studies with general readers contributes to their comparatively weaker effect sizes, but this hypothesis requires targeted testing in future work.

Similarly, the choice of devices—computers, tablets, and game consoles—may influence user engagement and accessibility, however, this remains an underexplored area warranting systematic study. Thus, although digital and traditional interventions currently show comparable efficacy at the outcome level, it remains that when it comes to scalability, digital tools offer scalability advantages that paper-and-pencil methods cannot, given their reliance on in-person delivery.

Beyond delivery features such as design and device choices, the diversity of targeted skills within interventions reflects another critical design dimension as variety in training is believed to facilitate transfer of learning across the different tasks reading entails (Deveau et al., 2015). Our analysis (see Figure S9) highlights that domain-specific interventions predominantly focus on a narrow set of linguistic and/or reading skills, while domain-general interventions target a broader range of skills such as executive functions, working memory, cognitive flexibility, and inhibitory control, either individually or in combination. This variety underscores the flexibility of digital tools to address a range of cognitive and reading-related skills. However, the heterogeneity in reported skills and tasks underscores the need for better standardization and reporting to enable more precise evaluations of the key design choices in intervention efficacy.

Such efforts will not only improve our understanding of what makes digital interventions effective but also provide actionable insights for optimizing their design and implementation in diverse educational, clinical, and home settings.

Conclusions

In summary, this meta-analytic work underscores the effectiveness of digital interventions in enhancing reading skills, addressing both poor readers, including those with dyslexia, and the general school population. Advancing educational equity and fostering inclusive learning requires overcoming the limitations of traditional approaches, such as reliance on specialized operators, lengthy durations, and limited adaptability (Denton et al., 2003; Vaughn & Fuchs, 2003). Personalized, scalable, and technology-enhanced interventions offer promising solutions to address diverse linguistic and educational needs.

Our findings resonate with the recent optimism expressed by intervention researchers regarding the growing adoption of digital interventions, which began with a few studies in the early 2000 s and gained sustained momentum after 2015 (Jamshidifarsani et al., 2019; Ostiz-Blanco et al., 2021; Peters et al., 2019). However, they also underscore persistent concerns about the limited number of high-quality studies that clearly link interventions to meaningful outcomes for children. To advance the field, there is an urgent need for rigorously designed studies that systematically examine the intersections of the moderating characteristics identified in this review. Given the resource-intensive nature of such work, now is a critical moment to establish international consortia dedicated to intervention research, extending beyond the scope of cross-national assessment alone. Building on the collaborative models of PIRLS (Mullis & Martin, 2019), PISA, and the Haskins Global Literacy Hub, such consortia should adopt shared protocols, open data practices, and harmonized reporting standards across languages and educational systems. By fostering coordinated, cross-linguistic research efforts, the field can generate robust evidence to inform the design of equitable, scalable, and effective digital reading interventions that serve diverse learners worldwide.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41465-025-00336-2.

Acknowledgements The authors are thankful to BZ and MM for their help with the initial stages of the literature search, review, and data extraction processes.

Author Contributions AP: Conceptualization, Validation, Formal analysis, Investigation, Data Curation, Visualization; Project Supervision; Writing - Original Draft; EC: Conceptualization, Validation, Formal analysis, Investigation, Data Curation, Visualization; Supervision; Writing - Original Draft; C.H. Investigation; Writing - Review & Editing; BB: Methodology, Formal Analysis, Data curation, Visualization; Writing - Original Draft; DB: Conceptualization, all other steps – Review and Advising, Writing – Review, Original text & Editing, Project supervision, Project administration, Project funding.

Funding Open access funding provided by University of Geneva. This work was supported by the NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888 as well as the Klaus J. Jacobs Foundation to D.B.

Data Availability The dataset as well as the code to reproduce the analyses reported in this work has been made publicly available via the MetaAna_app on ShinyApp:https://bbediou.shinyapps.io/MetaAna_app_reading .

Declarations

Competing interests D.B. is a founding member of Akili Interactive, Boston. A.P. and D.B. have filed a patent that protects some of the mechanics of the video game SOM used in Pasqualotto et al. (2022). The remaining authors declare no conflicts of interest.

References

References marked with an asterisk indicate studies included in the meta-analyses.

Alqahtani, A. Y., & Rajkhan, A. A. (2020). E-learning critical success factors during the COVID-19 pandemic: A comprehensive analysis of e-learning managerial perspectives. *Education Sciences, 10*(9), Article Article9.

Andreev, R., Terzieva, V., & Kademova-Katzarova, P. (2009). An approach to development of personalized e-learning environment for dyslexic pupils' acquisition of reading competence. *Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, 1–6. https://doi.org/1.1145/173174.1731829.

Barr, R., & Kirkorian, H. (2023). Reexamining models of early learning in the digital age: Applications for learning in the wild. *Journal of Applied Research in Memory and Cognition, 12*(4), 457. https://doi.org/10.1037/mac0000132

*Bertoni, S., Franceschini, S., Puccio, G., Mancarella, M., Gori, S., & Facoetti, A. (2021). Action video games enhance attentional control and phonological decoding in children with developmental dyslexia. *Brain Sciences, 11*(2), Article 171.

*Bertoni, S., Franceschini, S., Ronconi, L., Gori, S., & Facoetti, A. (2019). Is excessive visual crowding causally linked to developmental dyslexia? *Neuropsychologia, 130*, 107–117.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on english hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1117.

*Bonacina, S., Cancer, A., Lanzi, P. L., Lorusso, M. L., & Antonietti, A. (2015). Improving reading skills in students with dyslexia: The efficacy of a sublexical training with rhythmic background. *Frontiers in Psychology, 6*(OCT), 1–8. https://doi.org/10.3389/fpsyg.2015.01510

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-Analysis*, *2*, 221–235.

Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), Article 2. https://doi.org/1.1016/j.cognition.2006.05.009.

Bryant, B. R., Ok, M., Kang, E. Y., Kim, M. K., Lang, R., Bryant, D. P., & Pfannestiel, K. (2015). Performance of Fourth-Grade students with learning disabilities on multiplication facts comparing Teacher-Mediated and Technology-Mediated interventions: A preliminary investigation. *Journal of Behavioral Education*, *24*(2), 255–272. https://doi.org/1.1007/s10864-015-9218-z

Buckingham, J. (2020). Systematic phonics instruction belongs in evidence-based reading programs: A response to Bowers. *The Educational And Developmental Psychologist, 37*(2), Article Article2.

*Cancer, A., De Salvatore, M., Granocchio, E., Andreoli, L., Antonietti, A., & Sarti, D. (2022). The role of auditory and visual components in reading training: No additional effect of synchronized visual cue in a rhythm-based intervention for dyslexia. *Applied Sciences, 12*(7), Article 336.

Carneiro, R., & Gordon, J. (2013). Warranting our future: Literacy and literacies. *European Journal of Education*, *48*(4), Article4. https://doi.org/1.1111/ejed.12055

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144.

Cartwright, K. B. (2012). Insights From Cognitive Neuroscience: The Importance of Executive Function for Early Reading Development and Education. *Early Education and Development, 23*(1), 24–36. https://doi.org/10.1080/10409289.2011.615025

Cartwright, K. B., & Palian, S. R. (2024). Considering roles of executive functions in the science of reading: A meta-analysis highlighting promises and challenges of reading-specific executive functions. *Educational Psychologist*, *59*(4), 263–29. https://doi.org/1.1080/0046152.2024.2418392.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), Article1. https://doi.org/1.1177/1529100618772271

Cidrim, L., & Madeiro, F. (2017). Information and communication technology (ICT) applied to dyslexia: Literature review tecnologias Da Informação e Da Comunicação (TIC) Aplicadas à dislexia: Revisão de literatura. *Rev Cefac*, *19*, 99–10899.

Cirino, P. T., Miciak, J., Ahmed, Y., Barnes, M. A., Taylor, W. P., & Gerst, E. H. (2019). Executive function: Association with multiple reading skills. *Reading and Writing, 32*(7), Article 7.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates,.

Colenbrander, D., Kohnen, S., Beyersmann, E., Robidoux, S., Wegener, S., Arrow, T., Nation, K., & Castles, A. (2022). Teaching children to read irregular words: A comparison of three instructional methods. *Scientific Studies of Reading, 26*(6), 545–564.

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of quick fix interventions for children with reading disability? *Scientific Studies of Reading*, *18*(1), 55–73. https://doi.org/1.1080/10888438.2013.836200.

Corbett, A. (2001). Cognitive Computer Tutors: Solving the Two-Sigma Problem. In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.), *User Modeling 2001* (pp. 137–147). Springer. https://doi.org/1.1007/3-540-44566-8_14.

Dahlin, K. I. E. (2011). Effects of working memory training on reading in children with special needs. *Reading and Writing*, *24*(4), Article 4. https://doi.org/1.1007/s11145-010-9238-y.

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading, 22*(1), 101–116.

de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, *13*(4), 318–333. https://doi.org/1.1080/10888430903001308.

De Morton, N. A. (2009). The PEDro scale is a valid measure of the methodological quality of clinical trials: A demographic study. *Australian Journal of Physiotherapy, 55*(2), 129–133.

Dean, J., Pascoe, M., & le, R. J. (2021). Information and communication technology reading interventions: A scoping review. *Reading & Writing - Journal of the Reading Association of South Africa*, *12*(1):294. https://doi.org/10.4102/rw.v12i1.294.

Debeer, D., Vanbecelaere, S., Van Den Noortgate, W., Reynvoet, B., & Depaepe, F. (2021). The effect of adaptivity in digital learning technologies. Modelling learning efficiency using data from an educational game. *British Journal of Educational Technology, 52*(5), 1881–1897. https://doi.org/10.1111/bjet.13103.

Denton, C. A., Vaughn, S., & Fletcher, J. M. (2003). Bringing research-based practice in reading intervention to scale. *Learning Disabilities Research & Practice, 18*(3), 201–211. https://doi.org/10.1111/1540-5826.00075

Deveau, J., Jaeggi, S. M., Zordan, V., Phung, C., & Seitz, A. R. (2015). How to build better memory training games. Frontiers in Systems Neuroscience, Volume 8–2014. https://www.frontiersin.org/journals/systemsneuroscience/articles/10.3389/fnsys.2014.00243

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience, 18*, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Duke, N. K., Bennett-Armistead, V. S., and Roberts, E. (2003). "Bridging the gap between learning to read and reading to learn," in Literacy and Young Children: Research-Based Practices, eds D. M. Barone and L. M. Morrow (New York, NY: The Guildford Press), 226–242.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

*Ecalle, J., Kleinsz, N., & Magnan, A. (2013). Computer-assisted learning in young poor readers: The effect of graphosyllabic training on the development of word reading and reading comprehension. *Computers in Human Behavior, 29*(4), 1368–1376.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National reading panel's Meta-Analysis. *Review of Educational Research*, *71*(3), 393–447. https://doi.org/10.1080/02643290500483090

Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., Umiltà, C., & Mascetti, G. G. (2006). *The relationship between visuo-spatial attention and nonword reading in developmental dyslexia, 23*(6), 841–855.https://doi.org/10.1080/02643290500483090

Fisher, Z., Tipton, E., & Hou, Z. (2016). Robumeta: Robust variance meta-regression. R package (Version 1.8.)[Computer software]. *arXiv Preprint arXiv*, *1503*.

Foley, N. C., Teasell, R. W., Bhogal, S. K., & Speechley, M. R. (2003). Stroke rehabilitation evidence-based review: Methodology. *Topics in Stroke Rehabilitation, 10*(1), 1–7.

*Forné, S., López-Sala, A., Mateu-Estivill, R., Adan, A., Caldú, X., Rifà-Ros, X., & Serra-Grabulosa, J. M. (2022). Improving reading skills using a computerized phonological training program in early readers with reading difficulties. *International Journal of Environmental Research and Public Health, 19*(18), Article 11526.

Franceschini, S., & Bertoni, S. (2019). Improving action video games abilities increases the phonological decoding speed and phonological short-term memory in children with developmental dyslexia. *Neuropsychologia, 130*, 100–106. https://doi.org/10.1016/j.neuropsychologia.2018.1.023

*Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology*, *23*(6), 462–466. https://doi.org/10.1016/j.cub.2013.01.044

*Franceschini, S., Trevisan, P., Ronconi, L., Bertoni, S., Colmar, S., Double, K., Facoetti, A., & Gori, S. (2017). Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia. *Scientific Reports*, *7*(1), 1–12.https://doi.org/10.1038/s41598-017-05826-8

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *Plos One*, *9*(2), Article2. https://doi.org/1.1371/journal.pone.0089900

*Gebauer, D., Fink, A., Kargl, R., Reishofer, G., Koschutnig, K., Purgstaller, C., Fazekas, F., & Enzinger, C. (2012). Differences in brain function and changes with intervention in children with poor spelling and reading abilities. *PLoS One, 7*(5), e38201.

Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of Vision*, *15*(1), Article 1. https://doi.org/1.1167/15.1.8.

Grainger, J., Dufau, S., & Ziegler, J. C. (2016). Opinion A Vision of Reading. Trends in Cognitive Sciences, 20(3):3. https://doi.org/1.1016/j.tics.2015.12.008.

Green, C., Bavelier, D., Kramer, A., Vinogradov, S., Ansorge, U., Ball, K., Bingel, U., Chein, J., Colzato, L., Edwards, J., Facoetti, A., Gazzaley, A., Gathercole, S., Ghisletta, P., Gori, S., Granic, I., Hillman, C., Hommel, B., Jaeggi, S., & Witt, C. (2019). Improving Methodological Standards in Behavioral Interventions for Cognitive Enhancement. *Journal of Cognitive Enhancement, 3*,. https://doi.org/10.1007/s41465-018-0115-y

*Gustafson, S., Fälth, L., Svensson, I., Tjus, T., & Heimann, M. (2011). Effects of three interventions on the reading skills of children with reading disabilities in grade 2. *Journal of Learning Disabilities, 44*(2), 123–135. https://doi.org/10.1177/0022219410391187

Haft, S. L., Caballero, J. N., Tanaka, H., Zekelman, L., Cutting, L. E., Uchikoshi, Y., & Hoeft, F. (2019). Direct and indirect contributions of executive function to word decoding and reading comprehension in kindergarten. *Learning and Individual Differences, 76*, Article 101783.

Harrar-Eskinazi, K. L., De Cara, B., Leloup, G., Nothelier, J., Caci, H., Ziegler, J. C., & Faure, S. (2022). Multimodal intervention in 8-to 13-year-old French dyslexic readers: Study protocol for a randomized multicenter controlled crossover trial. *BMC Pediatrics*, *22*(1), Article1.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J., Savović, J., Page, M. J., Sterne, J. (2019). Revised Cochrane risk of bias tool for randomized trials (RoB 2). https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

*Hitchcock, C., & Westwell, M. S. (2017). A cluster-randomised, controlled trial of the impact of Cogmed Working Memory Training on both academic performance and regulation of social, emotional and behavioural challenges. *Journal of Child Psychology and Psychiatry, 58*(2), 140–150. https://doi.org/10.1111/jcpp.12638

*Hölz, A. (2015). Kurzzeitige effekte eines trainings der Leseflüssigkeit von leseschwachen kindern Mit isolierten silben und morphemen. *Empirische Sonderpädagogik, 7*(2), 153–171.

*Horne, J. K. (2017). Reading comprehension: A computerized intervention with primary-age poor readers. *Dyslexia, 23*(2), 119–114.

*Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annals of Dyslexia, 58*(2), 115–137. https://doi.org/10.1007/s11881-008-0017-2

*Hughes, J. A., Phillips, G., & Reed, P. (2013). Brief exposure to a self-paced computer-based reading programme and how it impacts reading ability and behaviour problems. *PLoS One, 8*(11), e77867.

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers and Education, 128*, 427–451. https://doi.org/10.1016/j.compedu.2018.1.003

*Jiménez, J. E., del Rosario Ortiz, M., Rodrigo, M., Hernández-Valle, I., Ramírez, G., Estévez, A., O'Shanahan, I., & de la Trabaue, L. M. (2003). Do the effects of computer-assisted practice differ for children with reading disabilities with and without IQ—Achievement discrepancy? *Journal of Learning Disabilities, 36*(1), 34–47.

*Jiménez, J. E., Hernández-Valle, I., Ramírez, G., del Rosario Ortiz, M., Rodrigo, M., Estévez, A., O'Shanahan, I., García, E., & de la Trabaue, L. M. (2007). Computer speech-based remediation for reading disabilities: The size of spelling-to-sound unit in a transparent orthography. *The Spanish Journal of Psychology, 10*(1), 52–67.

*Jiménez, J. E., & Muñetón, M. A. (2010). Effects of computer-assisted practice on reading and spelling in children with learning disabilities. *Psicothema, 22*(4), 813–821.

*Johann, V. E., & Karbach, J. (2020). Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children. *Developmental Science, 23*(4), e12866–e12866. https://doi.org/10.1111/desc.12866

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin, 147*(1), 1–15.

*Jones, J. S., Milton, F., Mostazir, M., & Adlam, A. R. (2020). The academic outcomes of working memory and metacognitive strategy training in children: A double-blind randomized controlled trial. *Developmental Science, 23*(4), e1287.

*Kamykowska, J., Haman, E., Richardson, U., Latvala, J. M., & Lyytinen, H. (2014). Developmental changes of early reading skills in six-year-old Polish children and graphogame as a computer-based intervention to support them. *L1-Educational Studies in Language and Literature*.

Karbach, J., Strobach, T., & Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology, 21*(3), Article 3.

*Karemaker, A., Pitchford, N. J., & O'Malley, C. (2010). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers & Education, 54*(1), 199–208.

Kim, J. S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis, 28*(4), 335–355. https://doi.org/10.3102/01623737028004335

Klahr, D., & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives, 5*(3), 196–204.

*Kloos, H., Sliemers, S., Cartwright, M., Mano, Q., & Stage, S. (2019). MindPlay virtual reading coach: Does it affect reading fluency in elementary school? *Frontiers in Education, 4*, 67.

*Konerding, M., Bergstroem, K., Lachmann, T., & Klatte, M. (2021). Effects of the computer-based grapho-phonological training lautarium in children with developmental dyslexia. *Praxis Der Kinderpsychologie Und Kinderpsychiatrie, 70*(4), 333–355.

*Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly, 48*(1), 61–76.

Landerl, K., Castles, A., & Parrila, R. (2021). Cognitive precursors of reading: A cross-linguistic perspective. *Scientific Studies Of Reading, 26*(2), 111–124. https://doi.org/10.1080/10888438.2021.1983820

Landerl, K., Freudenthaler, H. H., Heene, M., Jong, P. F. D., Manolitsis, G., Parrila, R., Georgiou, G. K., Landerl, K., Freudenthaler, H. H., Heene, M., & Jong, P. F. D. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading, 23*(3), 3. https://doi.org/10.1080/10888438.2018.1510936

Lassault, J., Sprenger-Charolles, L., Albrand, J. P., Alavoine, E., Richardson, U., Lyytinen, H., & Ziegler, J. C. (2022). Testing the effects of graphoGame against a computer-assisted math intervention in primary school. *Scientific Studies of Reading, 26*(6), 449–468. https://doi.org/10.1080/10888438.2022.2052884

*Lenhard, W., Baier, H., Endlich, D., Lenhard, A., Schneider, W., & Hoffmann, J. (2012). Computer Assisted reading comprehension instruction: Effects of automatically generated content feedback.

*Loeb, D. F., Gillam, R. B., Hoffman, L., Brandel, J., & Marquis, J. (2009). The effects of Fast ForWord Language on the phonemic awareness and reading skills of school-age children with language impairments and poor reading skills. *American Journal of Speech-Language Pathology.* https://doi.org/10.1044/1058-0360(2009/08-0067)

*Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, *18*(1), Article 1. https://doi.org/1.1080/09297049.2011.575772.

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, *4*(2), e11–e11. https://doi.org/1.2196/games.5888.

Łuniewska, M., Chyl, K., Dębska, A., Kacprzak, A., Plewko, J., Szczerbiński, M., Szewczyk, J., Grabowska, A., & Jednoróg, K. (2018). Neither action nor phonological video games make dyslexic children read better. *Scientific Reports, 8*(1), 1–11. https://doi.org/10.1038/s41598-017-18878-7

MacArthur, C. A., Ferretti, R. P., Okolo, C. M., & Cavalier, A. R. (2001). Technology applications for students with literacy problems: A critical review. *The Elementary School Journal, 101*(3), 273–301. https://doi.org/10.1086/499669

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy, 83*(8), 713–721.

Malouf, D. B. (1988). The effect of instructional computer games on continuing student motivation. *The Journal of Special Education*, *21*(4), 27–38. https://doi.org/1.1177/002246698802100406.

Mayer, R. E. (2020). *Cognitive foundations of game-based learning.*

*Marzen, L. M. (2021). Wirksamkeit des computerbasierten Trainingsprogramms Lautarium bei Kindern mit Lese-Rechtschreibschwierigkeiten: Eine randomisierte kontrollierte Studie.

McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., Larsen, L., Marinus, E., Wang, H. C., & Castles, A. (2012). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews, 12*, 12. https://doi.org/10.1002/14651858.CD009115.pub2

McArthur, G., Sheehan, Y., Badcock, N., Francis, D., Wang, H., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*, *11*, Article 11. https://doi.org/1.1002/14651858.CD009115.pub3.

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Cataldi, E. F., Mann, F. B., & Barmer, A. (2019). The Condition of Education 2019. NCES 2019 – 144. In *National Center for Education Statistics*. National Center for Education Statistics. https://eric.ed.gov/?id=ED594978

McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically reviewing graphogame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Reading Research Quarterly, 55*(1), 45–73.

*Mehringer, H., Fraga-González, G., Pleisch, G., Röthlisberger, M., Aepli, F., Keller, V., Karipidis, I. I., & Brem, S. (2020). (Swiss) GraphoLearn: An app-based tool to support beginning readers. *Research and Practice in Technology Enhanced Learning, 15*, 1–21.

Melby-Lervåg, M., & Lervåg, A. (2014). Effects of educational interventions targeting reading comprehension and underlying components. *Child Development Perspectives*, *8*(2), 96–10. https://doi.org/1.1111/cdep.12068.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of far transfer: Evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 4. https://doi.org/10.1177/1745691616635612

*Messer, D., & Nash, G. (2018). An evaluation of the effectiveness of a computer-assisted reading intervention. *Journal of Research in Reading, 41*(1), 140–158. https://doi.org/10.1111/1467-9817.12107

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J. F., Longeras, E., Valdois, S., George, F., & Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/1.1016/j.learninstruc.2013.09.003.

Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/1.1177/1094428106291059.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105.

Mullis, I. V., & Martin, M. O. (2019). *PIRLS 2021 assessment frameworks*. ERIC.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Du Sert, P., Simonsohn, N., Wagenmakers, U., Ware, E. J., J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), Article 1. https://doi.org/1.1038/s41562-016-0021.

*Murphy, C. F., & Schochat, E. (2011). Effect of nonlinguistic auditory training on phonological and reading skills. *Folia Phoniatrica Et Logopaedica, 63*(3), 147–153.

*Nagler, T., Zarić, J., Kachisi, F., Lindberg, S., & Ehm, J. H. (2021). Reading-impaired children improve through text-fading training: Analyses of comprehension, orthographic knowledge, and RAN. *Annals of Dyslexia, 71*(3), 458–482.

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (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.

Ni, N., Gathercole, S. E., Norris, D., & Saito, S. (2023). Asymmetric negative transfer effects of working memory training. *Memory & Cognition*, *51*(7), 1654–1669.

*Nicolson, R., Fawcett, A., & Nicolson, M. (2000). Evaluation of a computer-based reading intervention in infant and junior schools. *Journal of Research in Reading, 23*(2), 194–209.

*Niedo, J., Lee, Y.-L., Breznitz, Z., & Berninger, V. W. (2014). Computerized silent reading rate and strategy instruction for fourth graders at risk in silent reading rate. *Learning Disability Quarterly, 37*(2), 100–11.

Nuijten, M. B. (2019). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology*, *61*(5), 535–539.

*Nunnery, J. A., Ross, S. M., & McDonald, A. (2006). A randomized experimental evaluation of the impact of accelerated reader/reading renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk (JESPAR), 11*(1), 1–18.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*(4), 443–468.

OECD. (2019). PISA 2018 assessment and analytical framework. OECD publishing. https://doi.org/10.1787/b25efab8-en

Okkinga, M., van Steensel, R., van Gelderen, A. J. S. (2018). Effectiveness of Reading-Strategy Interventions in Whole Classrooms: a Meta-Analysis. *Educational Psychology Review, 30:*1215–1239 https://doi.org/1.1007/s10648-018-9445-7.

Ostiz-Blanco, M., Bernacer, J., Garcia-Arbizu, I., Diaz-Sanchez, P., Rello, L., Lallier, M., & Arrondo, G. (2021). Improving reading through videogames and digital apps: A systematic review. *Frontiers in Psychology, 12*, Article 652948.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology, 134*, 103–112.

Pape, K., Bjørngaard, J. H., Westin, S., Holmen, T. L., & Krokstad, S. (2011). Reading and writing difficulties in adolescence and later risk of welfare dependence. A ten year follow-up, the HUNT Study, Norway. *BMC Public Health*, *11*. https://doi.org/1.1186/1471-2458-11-718.

Parsons, S., & Bynner, J. (2007). *Research report over the lifecourse*. Issue November).

*Pasqualotto, A., Altarelli, I., De Angeli, A., Menestrina, Z., Bavelier, D., & Venuti, P. (2022). Enhancing reading skills through a video game mixing action mechanics and cognitive training. *Nature Human Behaviour*, *6*(4), Article 4. https://doi.org/1.1038/s41562-021-01254-x.

Pasqualotto, A., Parong, J., Green, C. S., & Bavelier, D. (2023). Video game design for learning to learn. *International Journal of Human-Computer Interaction, 39*(11), 2211–2228.

*Pasqualotto, A., & Venuti, P. (2020). A Multifactorial Model of Dyslexia: Evidence from Executive Functions and Phonological-based Treatments. *Learning Disabilities Research & Practice, 35*(3), Article 3. https://doi.org/10.1111/ldrp.12228

Pennington, B. F., Santerre–Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D. R., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), Article 1. https://doi.org/1.1037/a0025823.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37.

*Peters, J. L., Crewther, S. G., Murphy, M. J., & Bavin, E. L. (2021). Action video game training improves text reading accuracy, rate and comprehension in children with dyslexia: A randomized controlled trial. *Scientific Reports, 11*(1), 18584.

Peters, J. L., Losa, L. D., Bavin, E. L., & Crewther, S. G. (2019). Neuroscience and Biobehavioral Reviews Efficacy of dynamic visuo-attentional interventions for reading in dyslexic and neurotypical children: A systematic review. *Neuroscience and Biobehavioral Reviews*, *100*(February), Article February. https://doi.org/1.1016/j.neubiorev.2019.02.015.

Plass, J. L., Homer, B. D., Pawar, S., Brenner, C., & MacNamara, A. P. (2019). The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cognitive Development, 49*, 56–67.

Puccio, G., Gazzi, G., Bertoni, S., Franceschini, S., Mancarella, M., Gori, S., Ronconi, L., & Facoetti, A. (2023). Action video games training in children with developmental dyslexia: A Meta-Analysis. *International Journal of Human–Computer Interaction, 40*(22), 6829-6844.

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance Estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Pustejovsky, J., & Tipton, E. (2017). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections (Version.3. 5)[Computer software]*.

Richardson, U., & Lyytinen, H. (2014). The graphogame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*. https://doi.org/10.17011/ht/urn.201405281859

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods, 26*(2), 141.

*Ronimus, M., Eklund, K., Westerholm, J., Ketonen, R., & Lyytinen, H. (2020). A mobile game as a support tool for children with severe difficulties in reading and spelling. *Journal of Computer Assisted Learning, 36*(6), 1011–1025.

*Rosas, R., Escobar, J. P., Ramírez, M. P., Meneses, A., & Guajardo, A. (2017a). Impact of a computer-based intervention in Chilean children at risk of manifesting reading difficulties/Impacto de Una intervención Basada En Ordenador En niños Chilenos Con

Riesgo de Manifestar dificultades lectoras. *Infancia y Aprendizaje, 40*(1), 158–188.

Santoro, L. E., & Bishop, M. (2010). Selecting software with caution: An empirical evaluation of popular beginning reading software for children with early literacy difficulties. *Computers in the Schools*, *27*(2), 99–12.

*Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology, 105*(2), 310. https://doi.org/10.1037/a0031025

*Savage, R. S., Abrami, P., Hipps, G., & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1. *Journal Of Educational Psychology, 101*(3), 59.

*Savage, R. S., Erten, O., Abrami, P., Hipps, G., Comaskey, E., & van Lierop, D. (2010). ABRACADABRA in the hands of teachers: The effectiveness of a web-based literacy intervention in grade 1 language arts programs. *Computers & Education, 55*(2), 911–922.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities, 48*(4), 369–390. https://doi.org/10.1177/0022219413504995

*Schöfl, M. (2016). Delfino–Ein Online-Förderwerkzeug Bei LRS für Den einsatz in der klasse. Lernen Und Lernstörungen. https://doi.org/10.1024/2235-0977/a000132

*Serrano, M., Vidal-Abarca, E., & Ferrer, A. (2018). Teaching self-regulation strategies via an intelligent tutoring system (TuinLECweb): Effects for low-skilled comprehenders. *Journal of Computer Assisted Learning, 34*(5), 515–525.

Seymour, P. H. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., Lyytinen, H., Gombert, J. E., Normand, L., Schneider, M. T., Porpodas, W., Ragnarsdottir, C., Tressoldi, H., Vio, P., De Groot, C., Licht, A., Iønnessen, R., Castro, F. E., Cary, S. L., & Olofsson, L. (2003). Å. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), Article 2. https://doi.org/10.1348/000712603321661859.

*Shannon, L. C., Styers, M. K., Wilkerson, S. B., & Peery, E. (2015). Computer-assisted learning in elementary reading: A randomized control trial. *Computers in the Schools, 32*(1), 20–34.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an outlier orthography. *Psychological Bulletin, 134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Snowling, M. (2014). Dyslexia: A language learning impairment. *Journal of the British Academy, 2*(September 2013), Article September 2013. https://doi.org/1.5871/jba/002.043.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81*(1), 1. https://doi.org/10.1111/j.2044-8279.201.02014.x

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: Past, present and future. *Oxford Review of Education, 46*(4), 501–513. https://doi.org/10.1080/03054985.2020.1765756

*Solan, H. A., Shelley-Tremblay, J., Ficarra, A., Silverman, M., & Larson, S. (2003). Effect of attention therapy on reading comprehension. *Journal of Learning Disabilities, 36*(6), 556–563.

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203–219. https://doi.org/1.1080/01443410903509259.

Stanley, T. D. (2017). Limitations of PET-PEESE and Other Meta-Analysis Methods. *Social Psychological and Personality Science, 8*(5), 581–591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), Article 1. https://doi.org/1.1002/jrsm.1095.

*Studer-Luethi, B., Bauer, C., & Perrig, W. J. (2016). Working memory training in children: Effectiveness depends on temperament. *Mem Cogn, 44*, 171–186. https://doi.org/10.3758/s13421-015-0548-9

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*(6), 1556–1579. https://doi.org/10.1037/a0020612

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77–96. https://doi.org/10.1177/0022219414528540

*Szili, K., Kiss, R., Csapó, B., & Molnár, G. (2022). Computer-based development of reading skills to reduce dropout in uncertain times. *Journal of Intelligence, 10*(4), Article 89.

Talbott, E., Lloyd, J. W., & Tankersley, M. (1994). Effects of Reading Comprehension Interventions for Students with Learning Disabilities. *Learning Disability Quarterly, 17*(3), 223–232. https://doi.org/10.2307/1511075

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews On Evidence-Based Nursing, 1*(3), 176–184.

*Tijms, J. (2011). Effectiveness of computer-based treatment for dyslexia in a clinical care setting: Outcomes and moderators. *Educational Psychology, 31*(7), 873–896.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*(6), 604–634.

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods, 10*(2), 180–194.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research, 78*(6), 6. https://doi.org/10.1007/s00426-013-0537-1

Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2019). Phonics: Reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education, 34*(2), 208–238. https://doi.org/10.1080/02671522.2017.1420816

*Tressoldi, P. E., Vio, C., & Iozzino, R. (2007). Efficacy of an intervention to improve fluency in children with developmental dyslexia in a regular orthography. *Journal of Learning Disabilities, 40*(3), 203–209.

*Troia, G. A., & Whitney, S. D. (2003). A close look at the efficacy of fast forword language for children with academic weaknesses. *Contemporary Educational Psychology, 28*(4), 465–494.

Valdois, S., Roulin, J. L., & Line Bosse, M. (2019). Visual attention modulates reading acquisition. *Vision Research, 165*, 152–161. https://doi.org/10.1016/j.visres.2019.1.011

*van Uittert, A., Verhoeven, L., & Segers, E. (2022). Responsiveness to a game-based intervention to enhance reading efficiency in first graders. *Journal Of Computer Assisted Learning, 38*(1), 178–191.

Vaughn, S., & Fuchs, L. S. (2003). Redefining Learning Disabilities as Inadequate Response to Instruction: The Promise and Potential Problems. *Learning Disabilities Research & Practice, 18*(3), 137–146. https://doi.org/10.1111/1540-5826.00070

Verhoeven, L., & Perfetti, C. (2021). Universals in Learning to Read Across Languages and Writing Systems. *Scientific Studies of Reading, 26*(2), 150–164. https://doi.org/10.1080/10888438.2021.1938575

Verhoeven, L., Perfetti, C., & Pugh, K. (2019). Introduction: Developmental dyslexia-A cross-linguistic perspective. *Verhoeven, L.; Perfetti, C.; Pugh, K.(Ed.), Developmental Dyslexia across Languages and Writing Systems*, 1–22.

Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Ryckeghem, D. M. L. V. (2020). The effects of gamification on computerized cognitive training: Systematic review and meta-analysis. *JMIR Serious Games*, *8*(3), e18644. https://doi.org/1.2196/18644.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60*, 419–435.

*Vidal-Abarca, E., Gilabert, R., Ferrer, A., Ávila, V., Martínez, T., Mañá, A., Llorens, A. C., Gil, L., Cerdán, R., & Ramos, L. (2014). TuinLEC, an intelligent tutoring system to improve reading literacy skills/TuinLEC, Un tutor inteligente Para mejorar La competencia lectora. *Infancia y Aprendizaje, 37*(1), 25–56.

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: A deficit in visuospatial attention, not in phonological processing. *Trends in Cognitive Sciences, 14*(2):2. https://doi.org/1.1016/j.tics.2009.12.003.

Viechtbauer, W., & Viechtbauer, M. W. (2015). Package 'metafor.' *The Comprehensive R Archive Network. Package 'Metafor'. Http://Cran. r-Project.Org/Web/Packages/Metafor/Metafor. Pdf.*

Wanzek, J., Vaughn, S., Kim, A. H., & Cavanaugh, C. L. (2006). The Effects of Reading Interventions on Social Outcomes for Elementary Students with Reading Difficulties: A Synthesis. *Reading & Writing Quarterly, 22*(2), 121–138. https://doi.org/10.1080/10573560500242192

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28*, 551–576.

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/1.1017/S0140525X20001685.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross- language investigation. *Psychological Science*. https://doi.org/1.1177/0956797610363406.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.