

The Effects of Educational Technology Interventions on Literacy in Elementary School: A Meta-Analysis

Rebecca D. Silverman 
Kristin Keane 
Elena Darling-Hammond
Stanford University

Saurabh Khanna
University of Amsterdam

Educational technology is ubiquitous, and educational technology interventions often target literacy. However, there is insufficient information about the effects of these interventions on literacy outcomes in elementary school. Therefore, we conducted a systematic review and meta-analysis of quasi-experimental and experimental studies published between 2010 and 2023 that evaluated the effects of educational technology interventions on the literacy outcomes of K–5 students. The meta-analysis across 119 studies from 105 manuscripts revealed positive effects for decoding (0.33), language comprehension (0.30), reading comprehension (0.23), and writing proficiency (0.81). However, when isolating effects for standardized measures, effects for decoding (0.23), language comprehension (0.12), reading comprehension (0.14), and writing proficiency (0.34) were reduced. Moderator analyses suggested that effects can depend on outcome, intervention, and participant characteristics. In the discussion, we describe and contextualize findings and call for more research to further unpack moderator effects. We also discuss possible implications for practice and policy.

KEYWORDS: educational technology, literacy, elementary school

Investment in educational technology has grown exponentially over the past few decades (Corcoran, 2021; HolonIQ, 2020), and educational technology is becoming ever more present in classrooms across the United States and internationally (Escueta et al., 2020; Vega & Robb, 2019). For example, in a recent poll by Gallup (2019), 87% of the pre-kindergarten through 12th-grade U.S. public

school teachers surveyed indicated that they use educational technology to support student learning every day or a few days a week. In the same poll, 81% of teachers agreed there is value in using educational technology tools for teaching, but 73% of them noted that there is insufficient information about the effectiveness of these tools for supporting student learning. In fact, according to the international Organization for Economic Cooperation and Development (OECD; 2022, 2023), there is a critical need for up-to-date information on the effectiveness of educational technology to inform practitioners and policymakers worldwide who are making decisions about using educational technology interventions to improve learning across contexts.

Early literacy is an area in which educational technology interventions have proliferated (Common Sense Education, 2018), which makes sense given the integral role of early literacy to later success in school and beyond (International Literacy Association [ILA], 2015; National Institute of Child Health and Human Development [NICHD], 2010). To inform practice and policy and to guide future research on supporting early literacy, there is a need for current information on the effectiveness of educational technology interventions that target literacy skills in elementary school, an important period in education with a major focus on literacy development (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010). A seminal meta-analysis by Cheung and Slavin (2012) that examined the effects of educational technology interventions on K–12 reading achievement was published over a decade ago (Cheung & Slavin, 2012). Since then, the state of educational technology use has evolved. There is wider access to devices, the internet, and interactive educational tools and applications than ever before (Rideout & Robb, 2020). There have been several meta-analyses since Cheung and Slavin (2012) published their study (e.g., Takacs et al., 2015; Vanbecelaere et al., 2023; Xu et al., 2019), but none of them focuses specifically on elementary school and investigates a range of educational technology interventions across literacy outcomes. The present meta-analysis that reviews studies since 2010 is intended to meet this need. Below we review relevant theoretical background and previous research to contextualize our study.

Theoretical Background

The benefit of educational technology to student learning has been debated for decades (Dede, 2008). On one hand, educational technology could enhance learning by providing information through multiple modalities (Baddeley, 1992; Paivio, 1986) and offering increased opportunities for actively engaging in learning (Mayer, 2009; Wittrock, 1989). On the other hand, given humans' limited capacity to process large amounts and different types of information—much less to engage with it—educational technology could deter learning (Baddeley, 1992; Chandler & Sweller, 1991). Further, while some theorists have seen educational technology as just another vehicle to deliver instruction (R. E. Clark, 1994), others have seen educational technology as having an actual impact on learning (Kozma, 1994). In this view, the technology and the educational approach manifested through or embedded within that technology are inextricably linked.

Accordingly, Dede (2008) indicates that educational technologies may be informed by different pedagogical approaches (e.g., behaviorist, cognitivist, and constructivist). For example, educational technology that focuses on factual knowledge using direct instruction, repeated practice, and immediate evaluation and error correction would align with a behaviorist approach; educational technology that focuses on strategies to organize and process information effectively and efficiently would align with a cognitive approach; and educational technology that offers open-ended experiences and resources to facilitate independent or collaborative knowledge generation would align with a constructivist approach. Dede (2008) suggests that educational technologies that embody these pedagogical perspectives may be more or less appropriate for different instructional goals and objectives. Beyond pedagogy, others have proposed particular features included in some educational technology interventions that may support learning. For example, Bahari (2021) proposes that computer-mediated feedback can support language learning; Bernacki et al. (2021) suggest that personalization or adaptivity can improve academic performance; and Tobias et al. (2014) contend that gamification, which facilitates motivation and engagement, can facilitate learning of specific knowledge and skills. In alignment with Dede (2008), Mayer (2009) suggests that such features may be more or less important depending on the knowledge and skills addressed or assessed.

Extending these theories to early literacy, it could be that specific educational technology interventions or the pedagogy and features embedded in them could be more or less effective depending on the early literacy knowledge and skills under consideration. Though many models have been proposed and evaluated, a long-standing, empirically based model of reading is the Simple View (e.g., Florit et al., 2022; Gough & Tunmer, 1986), which proposes that the two primary components of reading comprehension are decoding and language comprehension. Although other factors are also important (e.g., executive functioning and motivation or engagement), there is consensus among researchers that decoding and language comprehension are both critical for reading comprehension across languages (Duke & Cartwright, 2021; Florit & Cain, 2011; Kim, 2020). In the original conception of the Simple View, decoding is defined as the ability to read words quickly and accurately (Gough & Tunmer, 1986), and decoding-related skills include phonological awareness, phonics, and fluency. Language comprehension is defined as the ability to understand language (Gough & Tunmer, 1986), and related skills include vocabulary knowledge, syntactic awareness, and inferencing ability. Ultimately, decoding and language comprehension skills are each necessary but insufficient as both are needed for reading comprehension (Snow & Matthews, 2016). Relatedly, the Simple View of Writing suggests that similar skills are needed for writing proficiency (Juel et al., 1986). Specifically, encoding (i.e., spelling) involves many of the same underlying skills as decoding and ideation (i.e., using language to create and communicate ideas) involves many of the same underlying skills as language comprehension. As with reading, other skills (e.g., handwriting or typing skills, executive functioning, motivation or engagement) are needed (Berninger et al., 2017; De La Paz & Butler, 2018), but research is clear that skills related to decoding and language comprehension are as critical

to writing proficiency as they are to reading comprehension (Kim & Graham, 2022).

Research suggests that the relative importance of instruction targeting decoding (i.e., code-focused instruction) or language comprehension (i.e., meaning-focused instruction) depends on what students need, and instruction to address these different aspects of reading comprehension and writing proficiency may require different instructional approaches (e.g., Bailey et al., 2020; Connor et al., 2011; Graham et al., 2012). For example, research suggests that direct instruction, which is behaviorally oriented, is effective for facilitating code-focused skills; and that strategy instruction, which is cognitively oriented, is effective for supporting meaning-focused skills (e.g., Suggate, 2016). Furthermore, research suggests that specific features of instruction (e.g., feedback) may be more important for students from particular populations (e.g., students with disabilities [Spear-Swerling, 2019] or English language learners [Zhang, 2022]). By extension, it could be that educational technology interventions have different effects depending on which aspects of reading and writing are targeted, which pedagogical approach or technology-based features are incorporated, or which populations of students are participating in the intervention. Thus, the present meta-analysis examines the effects of educational technology interventions across different literacy-related outcomes and examines whether intervention characteristics, such as the pedagogical orientation or technology-based features it includes, or participant characteristics moderate effects.

Previous Research

This study builds on previous meta-analyses on the effects of educational technology to support academic achievement. Kulik et al. (1985) published one of the first meta-analyses that suggested that educational technology interventions could have positive effects on language or reading outcomes in elementary school. The authors identified seven studies of computer-assisted instruction (CAI) that included language or reading outcomes and determined the average effect (Glass's Delta) across studies to be 0.42. Advances in educational technology and research from 1980 to 2010 were staggering. Thus, the 2012 meta-analysis by Cheung and Slavin that focused on studies conducted during those years was timely. These authors reviewed studies conducted across K–12 contexts to determine the effects of educational technology applications on reading outcomes, and they reported effects for kindergarten, elementary (i.e., Grades 1–6), and secondary students separately. Cheung and Slavin (2012) reviewed studies with a wide range of educational technologies such as computers, multimedia, or interactive whiteboards, but they only included experimental studies that lasted at least 12 weeks with at least two teachers per condition (i.e., treatment and control) and that included quantitative measures of general reading performance such as standardized tests. They excluded studies with other measures of reading-related skills such as phonemic awareness, vocabulary, and writing if there was no measure of general reading performance. Cheung and Slavin (2012) identified 84 studies and calculated effect sizes (ES) using the difference in posttest scores, adjusted for pretest scores and other covariates, divided by the unadjusted posttest pooled standard

deviation. Their analysis revealed positive effects of educational technology applications, compared to traditional approaches ($ES=0.16$).

Cheung and Slavin (2012) hypothesized that educational technology interventions may be differentially effective depending on intervention characteristics. They examined characteristics related to what Slavin (1995) termed the QAIT model, which suggested that the effectiveness of educational technology interventions may depend on four factors: quality, appropriate level, incentives, and time. They found what they termed “a marginally significant between-group effect” ($p \leq .10$) suggesting differences by intervention type (e.g., computer-managed learning systems, supplemental programs, comprehensive programs, and innovative technology applications). The authors also reported “marginally significant” differences ($p \leq .10$) by intensity (i.e., $>$ or <75 minutes per week) and significant differences ($p \leq .05$) by implementation level, which captured quality of implementation. The authors call for “further research . . . to determine how specific technology applications and specific teaching methods contribute to reading outcomes” (p. 212).

Cheung and Slavin (2012) also examined whether effects differed by participant characteristics. They found significant differences ($p = .01$) in effects across kindergarten ($k=8$, $ES=0.15$), elementary ($k=59$, $ES=0.10$), and secondary ($k=18$, $ES=0.31$) students, suggesting differences by grade range that may relate to how reading and writing change over time (Kim, 2023). Across K–12, they found “marginally significant” differences by ability level and significant differences by race, English language learner status, and socioeconomic status within studies. They did not find differences by gender or socioeconomic status across studies. In addition, Cheung and Slavin (2012) found significant differences depending on study design features (e.g., whether studies were published or unpublished and whether studies included a quasi-experimental design [QED] or randomized control trial [RCT]).

Educational technology has evolved rapidly since 2010, the last year included in Cheung and Slavin’s (2012) literature search. There have been several meta-analyses published since 2012, but none of them focuses specifically on elementary school and investigates a range of educational technology interventions across literacy outcomes. Some of these meta-analyses have been focused on particular types of educational technology interventions. For example, Takacs et al. (2015) conducted a meta-analysis examining the effects of technology-enhanced storybooks (i.e., electronic books or e-books). In their review of studies with preschool and elementary-age children (i.e., through Grade 6), results showed positive effects (Hedges’ g) on decoding-related skills (0.16), story comprehension (0.17), expressive vocabulary (0.20), and engagement and child-initiated communication during reading (0.26), but not receptive vocabulary (-0.08), suggesting that effects of technology-based interventions may differ according to the aspect of reading or writing under investigation. Moderator analyses suggested that multimedia features like animated pictures and sound effects were beneficial, but interactive elements like games, hotspots, pop-up questions, and clickable definitions were detrimental, particularly for students identified by the authors as “disadvantaged,” suggesting that the effect of educational technology interventions may differ by specific features and particular populations. Interestingly, the

conclusion that interactivity in general and gamification in particular could be associated with negative effects runs counter to findings from other systematic reviews. For example, T. Thomas et al. (2013), studying effects of computer-assisted instruction in elementary through higher education contexts, found positive effects associated with interactive versus didactic pedagogy; and D. B. Clark et al. (2016), reviewing studies on digital games in K–16 settings, found positive effects of gamification across domains (e.g., literacy, science, math).

In another meta-analysis that examined a particular type of educational technology intervention, Xu et al. (2019) focused on intelligent tutoring systems (ITSS), computer programs that provide students with targeted support and feedback. Reviewing studies conducted across K–12, Xu et al. (2019) found that ITSS had positive effects (Hedge's g) on reading comprehension when compared to traditional classroom instruction (0.60) and traditional tutoring (0.20). Unexpectedly, the authors found that ITSS that were strong in intensity (>75 minutes per week) and long in duration (>3 months) yielded a smaller ES (0.26) than programs that were weak in intensity and short in duration (0.78), suggesting that effects of educational technology interventions may differ according to intensity or duration. However, this difference faded when controlling for whether outcomes were from researcher-developed or standardized measures, indicating that findings may differ by the types of measures under study. Though not included in their meta-analysis, Xu et al. (2019) coded interventions for the inclusion of modeling, prediction, feedback, adaptivity, or scaffolding. In their discussion, Xu et al. (2019) suggest that further examination of whether effects differ according to these features is warranted. In fact, Major et al. (2021) conducted a meta-analysis focused on technology-enabled personalization that included adaptivity and feedback and found that, across studies conducted with 6- to 15-year-olds in low- and middle-income counties, greater personalization was related to higher math and literacy outcomes with a standardized ES of 0.18.

Other meta-analyses have focused on specific populations of students. For example, Lee et al. (2020) conducted a meta-analysis on the effects of technology-integrated instruction, compared to traditional instruction, for English language learners in K–12 classroom settings. These authors found positive effects ($ES=0.47$) on literacy outcomes. Weng et al. (2014) implemented a meta-analysis on the effects of what they called cognitive-skills based computer assisted instruction for PK–12 students with disabilities. These authors found a positive weighted average effect (0.35) for group-comparison design studies across outcomes (e.g., decoding, language comprehension, math), though results were not disaggregated by the outcome. Shamir and Korat (2015) published a review of studies of the effects of electronic storybooks on emergent literacy outcomes for kindergarten students they referred to as “at-risk for reading difficulties,” which included students with disabilities and students from low socioeconomic backgrounds. Though the review was not a meta-analysis per se, the authors found positive effects on emergent literacy outcomes for both groups of children. These publications add to the evidence that the effects of educational technology interventions on literacy outcomes may depend on participant characteristics.

Recently, Vanbecelaere et al. (2023) examined the effects of a range of digital interventions for early reading and examined intervention and participant

characteristics, but these authors only included studies that were focused on decoding. These authors explicitly excluded studies in which 50% or more of the intervention content targeted spelling, reading comprehension, or writing. Synthesizing effects across 72 studies with participants from 3 to 12 years old, Vanbecelaere et al. (2023) found that, compared to traditional teaching methods, digital technology interventions had a positive effect (Hedge's $g=0.37$) and that this effect did not depend on whether outcomes were cognitive or noncognitive, near or far transfer, or immediate or delayed. Additionally, these authors found that effects did not differ by intervention characteristics such as the duration of the intervention, the type of technology used, or whether gamification or adaptation was included. Additionally, though Savva et al. (2022) found that adult mediation facilitated language and literacy in a study on the effects of electronic storybooks with students in PK–2 and Archer et al. (2014) found that teacher training and support led to greater effects in a tertiary meta-analysis on the effectiveness of information communication technology (ICT) interventions in K–12 settings, Vanbecelaere et al. (2023) found that whether the teacher was actively involved and whether the teacher received coaching did not moderate effects. Furthermore, counter to Takacs et al. (2015), the effects did not depend on the socioeconomic status of the participants. Importantly, though, effects differed according to the mean age of participants such that studies with older children tended to have smaller effects than those with younger children.

While the meta-analysis of Vanbecelaere et al. (2023) comes close to meeting the need for an up-to-date meta-analysis focused on the effects of educational technology interventions across literacy outcomes in elementary school, it misses the mark because it does not disaggregate results for pre-kindergarten and elementary school and because it does not include studies focused on other aspects of literacy besides decoding. Given that literacy is a critical focus area in elementary school that includes decoding, language comprehension, reading comprehension, and writing (NGA & CCSSO, 2010), an up-to-date meta-analysis on the effects of various educational technology interventions that (a) focuses specifically on interventions implemented with elementary age or grade students; (b) examines effects across literacy-related outcomes; and (c) explores outcome, intervention, and participant characteristics as potential moderators of effects is warranted.

The Present Study

Building on the theory and research discussed above, we set out to conduct a meta-analysis of recent research conducted with elementary age or grade students to answer the following questions: (a) What are the effects of educational technology interventions on literacy-related outcomes including decoding, language comprehension, reading comprehension, and writing proficiency? (b) Do the effects differ depending on outcome, intervention, or participant characteristics? Though not central to our research questions, we also planned to examine whether effects differ by study design characteristics. For outcome characteristics, as in Xu et al. (2019) and Vanbecelaere et al. (2023), we intended to explore whether effects differed across literacy-related outcomes for proximal and distal researcher-developed and standardized measures. For intervention characteristics, we

examined differences by (a) pedagogical orientation as in T. Thomas et al. (2013); (b) features including gamification, adaptation, or feedback as in D. B. Clark et al. (2016), Major et al. (2021), and Takacs et al. (2015); (c) the type of technology, whether the intervention included non-technology-based instruction, and whether teacher training was included as in Vanbecelaere et al. (2023); and (d) intervention duration as in Xu et al. (2019). For participant characteristics, we explored grade range as in Cheung and Slavin (2012) and whether the studies included children from low socioeconomic backgrounds, children who were learning in a second or other language (L2+) from their primary language, and/or children with disabilities as in Shamir and Korat (2015), Lee et al. (2020), and Weng et al. (2014). For design characteristics, following Cheung and Slavin (2012) who found that effects depended on whether a study was quasi-experimental or randomized, though Takacs et al. (2015) and Vanbecelaere et al. (2023) did not, we explored differences by design type; and, following Xu et al. (2019) who found that effects depended on whether the comparison group could be characterized as business as usual or an alternative intervention, we analyzed differences by type of comparison group. In this meta-analysis, we followed guidelines established by Cooper et al. (2009) to identify studies; we used methods proposed by Lipsey and Wilson (2001) for data analysis; and we used recommendations by Alexander (2020) and Pigott and Polanin (2020) to guide our reporting.

Methods

Literature Search and Study Selection Criteria

First, we conducted a systematic search using the ProQuest ERIC and PsychINFO databases, the EBSCO Academic Search Premier database, and the Web of Science Core Collection. These databases include a wide range of social science journals and are highly relevant to the topic of educational technology. Additionally, we searched the ProQuest Dissertations and Theses Global database and conducted a manual search of the tables of contents of 15 journals across education and technology. We limited our search to peer-reviewed articles and dissertations published in English between 2010 and 2023. See Appendix A in Supplemental Materials, in the online version of the journal, for search terms and dates; and refer to Appendix B in Supplemental Materials, in the online version of the journal, for a list of manually-searched journals.

As shown in Figure 1, we engaged in a two-stage screening process. In the first stage, the first three authors reviewed titles and abstracts and excluded records if it was clear that the studies described in the records did not meet *one* of the following criteria: (a) the study includes educational technology; (b) the study focuses on supporting literacy; (c) the study includes literacy outcomes; (d) the study investigates an intervention (not an assessment) applicable in school settings; (e) the study involves K–5 students or children ages 5 to 11; and (f) the study uses a quasi-experimental or experimental design. If we could not tell from the title and abstract if a study should be excluded, we advanced the record to the second stage of the screening process. For the first stage of the screening process, we coscreened 20% of the corpus with 91.7% agreement. Discrepancies were

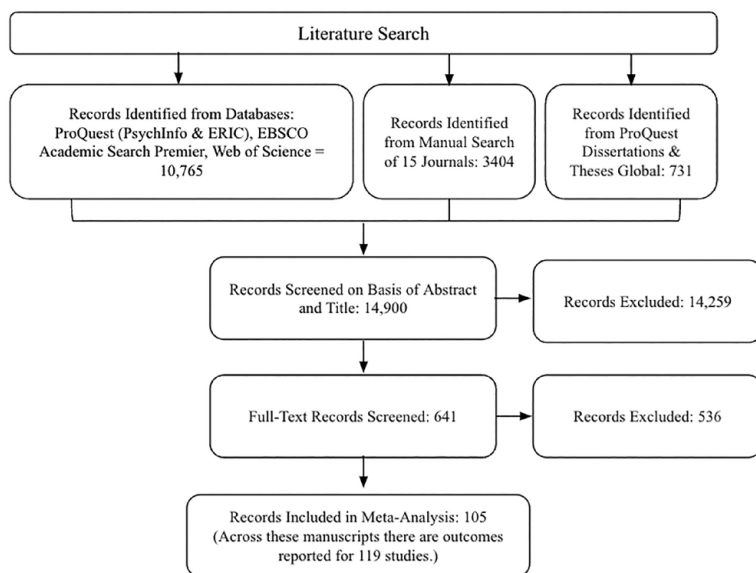


FIGURE 1. *Literature search flow diagram.*

discussed and adjudicated as a group. We screened the remainder of the records individually.

For the second stage, we refined the criteria above and reviewed full manuscripts, excluding any studies that did not meet *all* of the following criteria:

1. *The study evaluates an intervention involving educational technology.* We included interventions with e-books, online games, computer programs, or videos accessed via a computer, mobile phone, tablet, interactive whiteboard, or other electronic device. We included studies that investigated educational technology interventions with or without additional instructional support, but we excluded studies that used educational technology only for professional development and coaching of teachers (i.e., students did not use the technology).
2. *The study includes an educational technology intervention focused on supporting literacy.* We excluded studies that focused only on supporting general cognitive skills (e.g., working memory), handwriting or typing, or other content areas.
3. *The study includes quantitative literacy measures and presents pre- and posttest means and standard deviations by condition OR sufficient information to be able to extract or calculate effect sizes and standard errors.* We included studies with measures that assess decoding-related skills (i.e., phonological awareness, phonics, spelling, fluency), language comprehension-focused skills (i.e., vocabulary, listening comprehension, oral

language), reading comprehension, or writing proficiency. Note that, based on Takacs et al. (2015) and considering the reciprocal relationship between motivation or engagement and literacy (Hebbecker et al., 2019), we originally set out to investigate effects on motivation or engagement in addition to decoding and language comprehension skills. However, we found few studies with motivation or engagement outcomes. Therefore, we dropped these outcomes from our analyses. We will return to this decision in the discussion. We excluded studies in which the only outcome was from an informal reading inventory given the questionable reliability and validity of this type of measure (e.g., Klingbeil et al., 2015; D. Thomas & Dyches, 2019).

4. *The study includes an educational technology intervention with multiple (>1) sessions that could be applied in a school-based setting (e.g., school, after school, summer school).* We excluded studies with only one session because we wanted to focus on ongoing interventions relevant to school-based settings. However, we did not have a hypothesis about the minimum number of sessions, so we included any studies with more than one session.
5. *The study focuses on students in K–5 or children ages 5 to 11.* If the study included other grades or ages, at least 70% of the participants in the study had to be in K–5 or ages 5 to 11.
6. *The study uses a between-subjects quasi-experimental or experimental design with at least one treatment and one comparison or control group.* We included only studies in which the comparison or control group did not include educational technology or the educational technology included was not focused on literacy. If students were nested within teachers, classrooms, or schools, there had to be at least two teachers, classrooms, or schools per condition in the study to avoid confounding of treatment with teacher, classroom, or school effects. We excluded studies in which the level of assignment to condition was inconsistent (e.g., in some cases schools were assigned to condition and in some cases classrooms within a school were assigned to condition). We also excluded studies if the treatment and control groups were from different populations (e.g., students with and without disabilities).

At this stage, we coscreened all of the manuscripts with 92.3% agreement and adjudicated discrepancies as a group. See Appendix C in the Supplemental Materials, in the online version of the journal, for examples of studies that were excluded based on the criteria noted above.

Coding Procedures

For each manuscript, we recorded the overall sample size as well as the size of the sample by condition and coded studies for the outcome, intervention, participant, and design characteristics described below. If information was not available, we contacted the author(s) to obtain missing information. All manuscripts were double coded with 94.3% agreement, and all discrepancies were resolved through deliberation.

Outcome Characteristics

We coded for whether studies included measures of decoding, language comprehension, reading comprehension, or writing proficiency. We classified measures of phonological awareness, word identification, oral reading fluency, and spelling as measures of decoding; and we classified measures of vocabulary, listening comprehension, or oral expression as measures of language comprehension. We also coded for whether measures were standardized or researcher-designed. We defined standardized measures, using guidelines from the Institute of Education Sciences, as measures that are administered uniformly, following established administration and scoring procedures, and have evidence of reliability and validity across studies (What Works Clearinghouse, 2022). We further categorized researcher-developed measures as proximal or distal. These terms have been used in various ways across studies, but, in this study, we defined proximal measures as measures that capture what was taught in the intervention (e.g., being able to decode or determine the meaning of a word explicitly taught in the intervention), and we defined distal measures as measures that assess taught skills or strategies through untaught content or context (e.g., being able to decode a word not specifically taught in the intervention or being able to figure out the meaning of a word not explicitly covered in the intervention). For example, in one study, participants were given “two lists of trained word material” to read aloud (Görge et al., 2020, p. 6), and we coded that measure as proximal. In another study, participants were given “items that required a transfer of syntactic-morphological rules that were not explicitly included” in the intervention (Ritterfeld & Luke, 2021, p. 65), and we coded that measure as distal. Note that effects associated with proximal measures should be viewed with skepticism. As indicated in the What Works Clearinghouse (2022) *Procedures and Standards Handbook*, Version 5.0., “When outcome measures are closely aligned with or tailored to the intervention condition, the study findings may not be an accurate indication of the effect of the intervention (p. 27).” In Appendix D in the Supplemental Materials, in the online version of the journal, we provide examples of how we categorized measures across literacy-related outcomes. We also coded for whether studies included engagement or motivation outcome measures, whether studies included a delayed posttest to follow up on whether effects were sustained over time, and whether the outcome measures were administered in English or in another language, though we did not use these codes in our moderation analyses.

Intervention Characteristics

We categorized the educational technology interventions by the learning theory that they seemed to embody: (a) behavioral (i.e., focused on factual knowledge using direct instruction, repeated practice, and corrective feedback), (b) cognitive (i.e., focused on strategies to organize and process information), or (c) constructivist (i.e., focused on providing open-ended experiences and resources to support knowledge generation). Note that while some educational technology interventions included more than one pedagogical approach, we decided to classify each study according to the approach that was most predominant. Furthermore, we coded for specific features of the educational technology intervention including gamification, adaptivity, and feedback. We coded an intervention as including gamification if there were games that conferred points, badges, or other rewards

for completing or doing well on specific tasks; we coded it as adaptive if it tailored the level of difficulty or challenge based on assessment or performance; and we coded it as including feedback if there was some form of communication to users that indicated whether responses were correct or incorrect or provided suggestions for improvement. We took a conservative approach to coding for these variables. If studies explicitly stated that there was some form of adaptivity, gamification, or feedback, we coded “yes” for these features. It could be that some of the educational technology interventions included these features, but they are coded as “no” in our dataset because this was not explicit in the manuscript. Additionally, we coded for whether the technology could be classified as computer-assisted instruction, including intelligent tutoring systems, an e-book, or some other type of educational technology (e.g., interactive whiteboard), whether the technology was used by itself or in conjunction with other non-technology-based instruction, and whether or not the authors indicated that teacher training was provided. We also estimated the amount of time users participated in the intervention. This was challenging because in some cases authors provided information on the amount of time the researchers intended for the students to participate in the educational technology intervention, in other cases authors provided the exact amount of time students actually participated or a range across students, and in still other cases authors were vague when discussing duration (e.g., 12 weeks). We tried to infer the average number of hours in each study based on the information provided, but we know this data is not precise or consistent across studies.

Participant Characteristics

Given findings about differences by grade or age (Cheung & Slavin, 2012; Vanbecelaere et al., 2023), we coded for and investigated differences by grade level bands (i.e., K–2 [or ages 5–7] or 3–5 [or ages 8–11]) for analyses. In addition, we coded for whether the study included children from specific subgroups: (a) children from low socioeconomic backgrounds, (b) children who were learning in a second or other language (L2+) from their primary language, and/or (c) children with disabilities. Specifically, we determined whether at least 33% of the sample in a study was from the particular subgroup in question. Though this was an arbitrary decision, it was intended to ensure that there were substantial numbers of the given subgroup in the study. Unfortunately, in many studies, it was hard to determine whether children from particular subgroups were included or the exact number of children or the percentage of the sample from those subgroups. Therefore, it could be that some of the studies included participants from these populations, but they are coded as not including them in our dataset. Very few studies disaggregated results for specific populations. Therefore, findings related to whether these populations were included should be interpreted cautiously. Though not a focus of analyses, we also coded for whether the study was conducted in the United States or internationally.

Design Characteristics

We coded for whether studies used an RCT or a QED. We also noted whether the control or comparison condition(s) were business as usual (BAU) or an alternative treatment (ALT).

Analytic Procedures

For analyses, we created a database of the studies in our corpus that included information needed to derive effect sizes by domain (i.e., decoding, language comprehension, reading comprehension, and writing), type of measure (i.e., researcher-developed proximal or distal or standardized), and condition (e.g., treatment or control). We also included in this database categorical variables corresponding to our codes for study characteristics. To calculate effect sizes across studies, we used Hedge's g with the following small sample bias correction: $1 - (3/[4\{n_T + n_C\} - 9])$ (where n_T and n_C refer to treatment and control group sizes, respectively) (Hedges, 1982).¹ For studies that included sample sizes, means, and standard deviations for pretest and posttest, we calculated the standardized mean change using raw score standardization (Becker, 1988). In other words, we calculated the effect sizes as the mean pre–post change in the control group subtracted from the mean pre–post change in the treatment group, divided by the pooled pretest standard deviation (Becker, 1988; Morris, 2000). If baseline equivalence was established within a particular domain (e.g., decoding), we included posttest only measures within that domain and calculated the standardized mean difference (i.e., the difference between the treatment and control group means divided by the pooled standard deviation) (Hedges, 1981). In some studies, the authors did not report full information on pretest and posttest means and standard deviations by condition, but they included sufficient information to be able to extract or calculate effect sizes and standard errors. For these studies, we used information from reported effect sizes and standard errors (e.g., Cohen's d , Hedges' g , Glass' Δ) or calculated effect sizes using other available test statistics, such as F -statistics reported with analyses of variance. For studies that used a hierarchical design with clustering at the classroom or school level, we first calculated student-level statistics to generate effect sizes and effect size variances.² We then corrected effect size variances (EV) using a design effect adjustment (i.e., $EV_{\text{corrected}} = EV_{\text{original}} \times [1 + \{\text{average cluster size} - 1\} \times ICC]$) assuming an intra-class correlation coefficient (ICC) of .175, a midpoint of ICCs between .10 and .25 typically reported in educational research (Graham et al., 2020; Hedges & Hedberg, 2007; Heo et al., 2010; Kish, 1965; Särndal et al., 2003).³ Note that, in many cases, the researchers use a different method for deriving effect sizes; thus, the effect sizes we report may differ from those in the source studies. See Supplemental Materials Appendix E, in the online version of the journal, for more information.

When studies included multiple results for a type of measure within a domain, we aggregated effect sizes. Including each result when there are multiple results for a type of measure within a domain would violate the assumption of independence and lead to spurious heavier weights being assigned to studies with multiple results. While other meta-analyses have tackled this challenge by either selecting a single effect size from those available or by using simple unweighted means for aggregation to the study level, we followed Borenstein (2009) and simultaneously aggregated all within-study effect sizes for a type of measure within a domain while taking into account the correlations among the within-study results. We also conducted sensitivity analyses for these within-study correlations ($r = .2-.8$)

before assigning one effect size for each type of measure within each domain (Borenstein, 2009; Wampold et al., 1997). After obtaining a single effect size for each type of measure within each domain, we synthesized effect sizes based on a random-effects model and estimated model parameters using the restricted maximum likelihood (REML) estimator. One key advantage of REML over other estimation methods (e.g., maximum likelihood estimation) is that it tends to provide less biased estimates of the variance components, especially when there is significant heterogeneity among studies or there are few studies in the analysis (Langan et al., 2019).

Once effects were determined, we examined the heterogeneity of effect sizes and conducted moderator analyses for outcomes that had high heterogeneity (Higgins & Green, 2022). In order to avoid identification of spurious moderators through Type I error, we chose moderators based on theory and previous meta-analyses as discussed above. Moderator analyses were conducted via mixed-effects modeling, where we used random-effects meta-regressions to model moderating effects. Meta-regression results form an intuitive analogy to the family of linear regression models, as the slope in meta-regressions can be interpreted as the change in standardized mean difference when the moderating variable changes by one unit, controlling for other predictors (Thompson & Higgins, 2002). In addition to investigating each moderator individually, we ran meta-regressions with multiple moderators for each outcome. Many of the potential moderators under investigation in this study are related. Therefore, we carefully chose moderators to include in meta-regression analyses with multiple predictors to avoid substantial confounding. Datasets and analytic code are available on the Open Science Framework website.

Results

The literature search and study selection process yielded 14,187 peer-reviewed articles and 731 dissertations for a total of 14,900 records. We excluded 14,259 records in the first stage and 536 in the second stage of the screening process. We identified a total of 105 manuscripts (101 peer-reviewed articles and 4 dissertations) meeting the inclusion criteria. Fourteen manuscripts included more than one study or sample. For example, L. V. Lysenko and Abrami (2014) report on two studies that were conducted in different academic years, and Wijekumar et al. (2019) report on one study with separate analyses for students in Grades 4 and 5. When two samples were examined separately, we considered each analysis as a separate study. Thus, the 105 manuscripts yielded a total of 119 studies. References for all included studies are provided in Appendix F in the Supplemental Materials in the online version of the journal.

Descriptive Information on Included Studies

Tables S1 through S3 in the Supplemental Materials that appear in the online version of the journal provide descriptive information for each of the 119 studies included in the meta-analyses.

Outcome Characteristics

Outcomes were coded by domain (i.e., decoding, language comprehension, reading comprehension, and writing) and type of measure (i.e., researcher-developed proximal or distal or standardized). While some studies ($k=51$) contributed an effect size for only one type of measure within only one domain (e.g., Bos et al. [2016] contributed one effect size that was from a standardized measure of reading comprehension), other studies ($k=68$) contributed multiple effect sizes within and across domains (e.g., Fogarty et al. [2020] contributed effect sizes towards the analyses of researcher-developed proximal decoding measures, researcher-developed proximal language comprehension measures, standardized language comprehension measures, researcher-developed distal reading comprehension measures, and standardized reading comprehension measures). Across the corpus of studies, there were 221 outcomes. Decoding outcomes ($k=86$) were more frequent than language comprehension outcomes ($k=64$), and reading comprehension outcomes ($k=65$) were more frequent than writing proficiency outcomes ($k=6$). There was a higher prevalence of distal researcher-developed measures ($k=29$) compared to proximal researcher-developed measures for decoding ($k=12$), but there was higher prevalence of proximal researcher-developed measures ($k=29$) compared to distal researcher-developed measures ($k=7$) for language comprehension. No studies included proximal researcher-developed measures of reading comprehension or writing proficiency. Though not a focus of analyses, only 10 studies included a motivation or engagement measure, and only 19 studies included follow-up assessment. Roughly a third ($k=39$) of the studies in the corpus included outcome assessments that were administered in a language other than English (e.g., Hebrew, German, Spanish, Turkish). Given that the Simple View of Reading, which we used to theoretically ground this study, has been validated across languages (e.g., Florit et al., 2022; Kim & Park, 2019; Peng et al., 2021; Sparks, 2021), we determined that there was no need to separate out these studies in analyses.

Intervention Characteristics

The educational technology interventions across studies were more commonly aligned with behavioral ($k=50$) or cognitivist ($k=58$) pedagogy than with constructivist pedagogy ($k=11$). Many interventions included gamification ($k=49$), adaptivity ($k=39$), or feedback ($k=88$). One study included one treatment with gamification and one without (Boggio et al., 2023), and one study included one treatment with adaptation and one without (Solheim et al., 2018). There were more interventions with additional non-technology-based instruction ($k=64$) than without this instruction ($k=53$). Two studies included two treatments and compared effects with and without technology (Connor et al., 2019; Segal-Drori et al., 2010). Most interventions could be characterized as providing computer-assisted instruction ($k=98$), while relatively fewer included e-books ($k=7$) or other technologies ($k=14$). Over half of the studies ($k=68$) mentioned that teacher training was included. The average number of hours spent using the educational technologies across studies was 21.68 ($SD=23.62$), though that number varied widely from less than one hour of intervention to as many as 120 hours of intervention. The distribution for duration was negatively skewed so we determined not to

include duration as a continuous variable in analyses. Instead, we grouped studies by terciles to represent whether interventions were relatively short (<8 hours), medium (8–22 hours), or long (>22 hours).

Participant Characteristics

Roughly half of the studies ($k=52$) included participants in K–2 or ages 5 to 8 and the other half ($k=56$) included participants in Grades 3 to 5 or ages 8 to 11. (Note that some studies [$k=11$] included students in K–2 and 3–5). In 44 studies, at least one-third of the sample had a low socioeconomic background; in 28 studies, at least one-third of the sample was L2+; and in 17 studies, at least one-third of the sample was identified as having a disability. Though not a substantive focus in this study, we found that over half (60%) of the studies in the corpus were conducted outside of the United States (e.g., in Canada, China, Finland, India, Turkey, and Zambia).

Design Characteristics

In this corpus of studies, 84 studies were RCT rather than QED studies, and 92 included a business-as-usual control group rather than an alternative treatment.

Effects of Educational Technology Interventions on Literacy-Related Outcomes

The 119 studies in this meta-analysis yielded 730 individual effect sizes. Aggregation within each study and outcome (decoding, language comprehension, reading comprehension, and writing proficiency) resulted in 221 effect sizes used. Table 1 shows effect sizes (g) across literacy outcomes. See Figures F1 through F4 in the Supplemental Materials, in the online version of the journal, for forest plots for each outcome. For decoding, the average effect was $g=0.33$ ($k=86$, 95% confidence interval [CI] [0.24, 0.43], $p<.01$). For language comprehension, the average effect was $g=0.30$ ($k=64$, 95% CI [0.21, 0.39], $p<.01$). The average effect for reading comprehension was $g=0.23$ ($k=65$, 95% CI [0.16, 0.30], $p<.01$), and the average effect for writing proficiency was $g=0.81$ ($k=6$, 95% confidence interval [CI] [0.40, 1.23], $p<.01$). Each of these outcomes showed significant heterogeneity: $Q(85)=234.46$, $p<.01$, $I^2=75.65$ for decoding; $Q(63)=160.62$, $p<.01$, $I^2=68.66$ for language comprehension; $Q(65)=209.63$, $p<.01$, $I^2=69.82$ for reading comprehension; and $Q(5)=11.27$, $p<.05$, $I^2=56.36$ for writing proficiency.

Effects by Outcome, Intervention, Participant, or Design Characteristics

Given the limited number of studies with writing outcomes, we did not analyze the full set of potential moderators for this outcome. We examined only outcome type and found differences in effects on distal researcher-developed measures ($g=1.05$, $k=4$, 95% CI [0.58, 1.51], $p<.01$) and standardized measures ($g=0.34$, $k=2$, 95% CI [−0.12, 0.80], $p<.01$). A pairwise comparison suggested this difference was significant ($\Delta g=0.38$, $p<.05$). For the other outcomes, we examined differences by outcome characteristics in two ways. First, as seen in Table 1, we executed separate meta-regressions for each outcome by measurement type.⁴ In this analysis, we found the effects of proximal, distal, and standardized measures to be $g=0.39$ ($k=12$), $g=0.49$ ($k=29$), and $g=0.23$ ($k=45$) for decoding; and

TABLE 1
Effects across literacy outcomes by domain and type of measure

Outcomes	<i>k</i>	<i>g</i> [95% CI]	<i>p</i>	Heterogeneity
Decoding (overall)	86	0.33 [0.24, 0.43]	<.01	$Q=234.46$; $df=85$; $p<.01$; $T^2=0.12$; $I^2=75.65$
Decoding (proximal RD)	12	0.39 [0.18, 0.61]	<.01	$Q=24.40$; $df=11$; $p=.01$; $T^2=0.07$; $I^2=57.59$
Decoding (distal RD)	29	0.49 [0.25, 0.72]	<.01	$Q=114.35$; $df=28$; $p<.01$; $T^2=0.30$; $I^2=83.89$
Decoding (standardized)	45	0.23 [0.14, 0.32]	<.01	$Q=87.58$; $df=44$; $p<.01$; $T^2=0.04$; $I^2=54.42$
Language comprehension (overall)	64	0.30 [0.21, 0.39]	<.01	$Q=160.62$; $df=63$; $p<.01$; $T^2=0.07$; $I^2=68.66$
Language comprehension (proximal RD)	29	0.58 [0.41, 0.75]	<.01	$Q=85.41$; $df=28$; $p<.01$; $T^2=0.14$; $I^2=73.04$
Language comprehension (distal RD)	7	0.21 [0.06, 0.36]	<.01	$Q=12.43$; $df=6$; $p=.05$; $T^2=0.02$; $I^2=44.14$
Language comprehension (standardized)	28	0.12 [0.06, 0.18]	<.01	$Q=11.10$; $df=27$; $p=.9970$; $T^2<0$; $I^2=0$
Reading comprehension (overall)	65	0.23 [0.16, 0.30]	<.01	$Q=209.62$; $df=64$; $p<.01$; $T^2=0.04$; $I^2=69.82$
Reading comprehension (distal RD)	20	0.37 [0.25, 0.50]	<.01	$Q=39.82$; $df=19$; $p<.01$; $T^2=0.03$; $I^2=59.16$
Reading comprehension (standardized)	45	0.14 [0.08, 0.21]	<.01	$Q=86.25$; $df=44$; $p<.01$; $T^2=0.02$; $I^2=47.63$
Writing proficiency (overall)	6	0.81 [0.40, 1.23]	<.01	$Q=11.27$; $df=5$; $p<.05$; $T^2=0.15$; $I^2=56.36$
Writing proficiency (distal RD)	4	1.05 [0.58, 1.51]	<.01	$Q=5.94$; $df=3$; $p=.11$; $T^2=0.11$; $I^2=48.02$
Writing proficiency (standardized)	2	0.34 [-.12, 0.80]	0.15	$Q=0.01$; $df=1$; $p=.9184$; $T^2=0$; $I^2=0$

Note. RD=researcher developed.

TABLE 2
Estimates by individual outcome, intervention, participant, or design characteristics

		Decoding			Language Comprehension			Reading Comprehension		
		<i>k</i>	β	<i>SE</i>	<i>p</i>	<i>k</i>	β	<i>SE</i>	<i>p</i>	<i>p</i>
Outcome characteristics	RD proximal	12	.16	.15	.26	29	.42	.07	<.01**	
	RD distal	29	.21	.11	.06	7	.11	.10	.27	
	Standardized	45				28				<.01**
Intervention characteristics	Cognitivist	20	-.10	.11	.41	31	.08	.08	.30	.88
	Constructivist	6	.13	.20	.52	8	.52	.13	<.01**	
	Behaviorist	60				25				
Technology use	Tech plus	31	.01	.10	.93	29	-.18	.09	.04*	.02*
	Tech only	52				33				
	e-book	5	.16	.20	.43	6	.66	.15	<.01**	
Technology type	Other	2	-.04	.50	.94	14	.17	.09	.07	.40
	CAI	79				44				
	Yes	53	.01	.10	.90	24	-.16	.09	.09	.44
Gamification	No	30				40				
	Yes	40	-.04	.10	.71	11	-.20	.12	.09	.71
	No	43				53				
Adaptivity	Yes	68	-.08	.12	.52	38	-.18	.09	<.05*	.17
	No	18				26				

(continued)

TABLE 2. (continued)

		Decoding			Language Comprehension			Reading Comprehension		
		<i>k</i>	β	<i>SE</i>	<i>p</i>	<i>k</i>	β	<i>SE</i>	<i>p</i>	<i>p</i>
Duration	Medium	38	.04	.11	.73	22	-.10	.10	.32	.32
	High	22	.23	.11	.04*	13	-.10	.12	.40	.74
	Low	26				29				
Teacher training	Yes	41	-.04	.10	.67	39	-.12	.09	.20	.50
	No	45				25				
Participant characteristics										
Grade level	K-2	62	.09	.11	.43	31	.04	.09	.67	.35
	3-5	24				33				
Low SES*	Yes	22	.26	.11	.02*	24	.06	.09	.50	.45
	No	64				40				
L2+ learners*	Yes	17	-.08	.12	.49	19	-.05	.10	.66	.42
	No	69				45				
Students w/ disabilities*	Yes	25	.01	.11	.93	8	.10	.13	.46	.28
	No	61				56				
Design characteristics										
Design	RCT	72	.26	.13	.05	41	.04	.10	.68	.68
	QED	14				23				
Control or comparison	BAU	62	.09	.11	.41	50	.10	.12	.37	.12
	ALT	21				12				

Note. The comparison category is at the bottom of each category. Studies were classified as including students from low socioeconomic status (SES) backgrounds, students from L2+ language backgrounds, and students with disabilities if 33% or more of the sample was from that population.

* $p < .05$, ** $p < .01$.

TABLE 3
Estimates from meta-regression models with multiple moderators

		Decoding	Language Comprehension	Reading Comprehension
Intercept		−.09 .15	−.24* .11	.21 .13
Outcome characteristics				
Type	RD proximal	.26 .17	.50** .09	
	RD distal	.27* .11	.21* .10	.20** .07
		Standardized		
Intervention characteristics				
Pedagogy	Cognitivist	−.20 .15	.13 .07	−.06 .09
	Constructivist	.17 .23	.44** .13	
		Behaviorist		
Duration	Medium	.20 .12	.26** .09	−.01 .12
	High	.44** .15	.27* .11	.10 .14
		Low		
Tech use	Tech plus	.05 .14	−.06 .07	−.15 .08
	Tech only			
Participant Characteristics				
Grade level	K−2	.09 .11	.04 .07	.00 .08
Low SES	3−5			
	Yes	.24* .11	.06 .07	.04 .07
	No			
df		76	54	57
Tau squared		.10	.01	.02
Tau squared SE		.03	.01	.01
Akaike information criterion (AIC)		117.07	49.05	46.14
Bayesian information criterion (BIC)		142.71	7.93	64.52

Note. Standard errors in parentheses.
* $p < .05$, ** $p < .01$.

$g = 0.58$ ($k = 29$), $g = 0.21$ ($k = 7$), and $g = 0.12$ ($k = 28$) for language comprehension. For reading comprehension, there were no researcher-developed proximal measures; effects for distal and standardized measures were $g = 0.37$ ($k = 20$) and $g = 0.14$ ($k = 45$), respectively. Next, we conducted meta-regression analyses across the outcome types. See Table 2. Pairwise comparisons suggested no significant differences for decoding, but significant differences between proximal and distal ($\Delta g = 0.34$, $p < .01$) and proximal and standardized ($\Delta g = 0.42$, $p < .01$)

measures for language comprehension. (There was no difference between distal and standardized measures for language comprehension.) For reading comprehension, the effect for distal measures was significantly higher than for standardized measures ($\Delta g = 0.22, p < .01$).

As shown in Table 2, we found few significant differences associated with intervention characteristics across outcomes. For decoding, we found a significant difference between interventions of high and low duration ($\Delta g = 0.23, p = .04$) and between studies that did or did not include substantial numbers of students from a low socioeconomic background ($\Delta g = 0.26, p = .02$). For language comprehension, we found significant differences by pedagogical orientation, technology type, technology use (i.e., whether technology was used alone or with additional instruction) and feedback. Interventions that were constructivist in orientation were associated with higher effects than those that were behavioral in orientation ($\Delta g = 0.52, p < .01$); interventions with e-books were associated with higher effects than those with computer-assisted instruction ($\Delta g = 0.66, p < .01$); and interventions that included non-technology-based instruction ($\Delta g = -0.18, p < .05$) and feedback ($\Delta g = -0.18, p < .05$) were associated with lower effects than interventions without non-technology-based instruction and feedback. For reading comprehension, we found significant differences by technology use. Interventions that included non-technology-based instruction ($\Delta g = -0.16, p < .05$) were associated with lower effects than interventions without non-technology based instruction. We did not find significant differences associated with the design characteristics under investigation.

In addition to examining each potential moderator individually, we selected a few variables to include in multiple moderator analyses. Specifically, we included the type of measure, pedagogical orientation, duration, technology use, and socioeconomic status, which were significant in at least one of the individual meta-regression analyses noted above. Given the finding in Vanbecelaere et al. (2023) suggesting that effects were associated with age, we also included grade level in the multiple moderator analyses. See Table 3. To avoid confounding with pedagogical orientation, we did not include technology type and feedback in the multiple moderator analyses even though these variables were significant in some of the individual moderator analyses above. For decoding, there was a significant difference between distal and standardized measures ($\Delta g = 0.27, p < .01$), high and low duration interventions ($\Delta g = 0.44, p < .01$), and interventions conducted with substantial numbers of students from low socioeconomic backgrounds ($\Delta g = 0.24, p < .01$). For language comprehension, there was a significant difference between proximal and standardized measures ($\Delta g = 0.50, p < .01$) and distal and standardized measures ($\Delta g = 0.21, p < .01$). There were also differences between constructivist and behavioral interventions ($\Delta g = 0.44, p < .01$), medium and low duration interventions ($\Delta g = 0.26, p < .01$), and high and low duration interventions ($\Delta g = 0.27, p < .01$). For reading comprehension, there was a significant difference between distal and standardized measures ($\Delta g = 0.20, p < .01$).

Bias and Sensitivity Analyses

To determine the robustness of our findings, we examined the possibility of publication and sample size bias in our corpus of studies. Studies reporting strong

effects are more likely to be published, and studies with small samples are at greater risk of producing statistically nonsignificant results and, therefore, not being published. For each outcome—decoding, language comprehension, reading comprehension, and writing proficiency—we plotted all independent effects, residualized for variance attributable to measurement type (researcher-developed proximal versus researcher-developed distal versus standardized). A visual analysis of the funnel plots in Figures S5 through S8 in the Supplemental Materials, in the online version of the journal, shows that most estimates lie within prediction limits on the basis of sampling error. We considered the effects of studies that showed larger positive effects and lower effect size variances than expected on overall effect sizes using sensitivity analyses involving resynthesizing effect sizes after winsorizing such outliers and found the effects robust to these changes. Further, we conducted a *p*-curve analysis that addressed any biases induced by *p*-hacking (Simonsohn et al., 2014). (We excluded the writing proficiency outcome since there were too few studies to have a meaningful *p*-curve.) As can be seen in Figures S9 through S11 in the Supplemental Materials, the *p*-curves for each construct show a significant right skew and pass the evidential value test, suggesting the presence of true intervention effects for each construct under consideration. Finally, we conducted three sets of sensitivity analyses to verify the robustness of our synthesized effect sizes. First, we resynthesized effect sizes by excluding studies with QED designs since selection bias is more likely in QEDs compared to RCTs. Second, we winsorized effect sizes to reduce the effects of outliers, and then resynthesized effect sizes. Third, we resynthesized effect sizes using a correlated effects metaregression model that allows for varying levels of dependence between effect sizes. Table S4 in the Supplemental Materials shows that effect sizes and statistical significance across approaches mirror the trends outlined above.

Discussion

As educational technology tools to support literacy flood the market, policy-makers and practitioners need information on effectiveness, and developers and researchers need information on promising methods and gaps to fill. Building on previous meta-analyses (e.g., Cheung & Slavin, 2012; Kulik et al., 1985; Vanbecelaere et al., 2023), this study presents findings from a meta-analysis of studies on the effects of educational technology interventions on literacy-related outcomes in elementary school. The study focuses on research conducted since 2010, a period in which there has been rapid growth in the educational technology sector (Fittes, 2022; HolonIQ, 2020; Vega & Robb, 2019). The study makes a unique contribution by examining effects and potential moderators of effects across decoding, language comprehension, reading comprehension, and writing outcomes for students at the elementary age or grade level. Below we describe and contextualize our findings and suggest future directions.

Effects on Literacy-Related Outcomes

Results from the present meta-analysis suggest that educational technology interventions have positive effects across decoding ($k=86$, $g=0.33$), language comprehension ($k=64$, $g=0.30$), reading comprehension ($k=65$, $g=0.23$), and

writing proficiency ($k=6$, $g=0.81$) in elementary school. Using benchmarks from Kraft (2020)⁵ these are medium to large effect sizes that compare favorably to effect sizes across studies of educational interventions. For example, in one meta-analysis, Gersten et al. (2020) found that the mean effect size of reading interventions across reading outcomes for students in Grades 1 to 3 was 0.39 (Hedge's g). These authors also found that effect sizes differed for decoding (0.41) and reading comprehension (0.32) outcomes. In another meta-analysis, Neitzel et al. (2022) found that the average effect size (standardized mean difference) of reading programs across standardized measures of reading achievement for students they described as "struggling readers" in elementary school was 0.23. These authors also found differences for decoding (0.28) and reading comprehension (0.14). Interestingly, these authors found that the effect size for technology-supported adaptive instruction was only 0.09, which was nonsignificant. However, in a meta-analysis of language comprehension-focused interventions on researcher-developed and standardized measures in elementary school by Silverman et al. (2020), the average effect across interventions on reading comprehension was 0.19 (Hedge's g), and technology-mediated interventions tended to have greater effects than non-technology-mediated interventions.

Differences across meta-analyses make comparisons fraught, but the effects found in the present study are consistent with prior meta-analyses that suggest that educational technology interventions can have positive effects on elementary school age students' literacy (e.g., Cheung & Slavin, 2012; Kulik et al., 1985; Lee et al., 2020; Takacs et al., 2015; Vanbecelaere et al., 2023). Importantly, whereas previous meta-analyses showed positive effects for specific aspects of literacy (e.g., Vanbecelaere et al., 2023), particular types of educational technology interventions (e.g., Takacs et al., 2015), or specific populations of students (e.g. Lee et al., 2020), the present meta-analysis examined effects across literacy outcomes, types of educational technology, and student populations. These results suggest that investment in educational technology to support literacy is warranted (Fittes, 2022; HolonIQ, 2020), and the prevalence of educational technology in schools is justified (Gray & Lewis, 2021; Vega & Robb, 2019).

Considering the corpus of studies located and reviewed for this meta-analysis yields a few noteworthy observations about the state of research on the effects of educational technology interventions on literacy-related outcomes in elementary school. First, there were relatively fewer studies that included language comprehension outcomes and reading comprehension outcomes than studies that included decoding outcomes, suggesting that more research may be needed on these other critical aspects of reading. Second, there were only a few studies with writing proficiency outcomes ($k=6$). Writing proficiency is essential for academic success throughout school (Graham & Perin, 2007); therefore, much more research on educational technology interventions with writing proficiency outcomes is needed. Though there were few studies with writing proficiency outcomes, the average effect across these studies is impressive. Future research could build on studies in this meta-analysis that showed positive effects of a digital storytelling approach (Sarica & Usluel, 2016), a computer-based graphic organizer and self-regulation strategy intervention (Boykin et al., 2019), and a digital writing workshop model (Aktaş & Akyol, 2020). Third, few studies included motivation or

engagement outcomes ($k=10$) and few studies included follow-up assessment ($k=19$). Given the reciprocal relationship between motivation or engagement and literacy (Hebbecker et al., 2019), future research should examine direct effects of educational technology interventions on motivation and engagement outcomes and consider indirect effects on literacy outcomes through motivation and engagement outcomes. Furthermore, since effects of educational technology interventions could fade over time (Bailey et al. 2020), more studies with follow-up assessments (e.g., months or even years after the end of the intervention) are needed to determine whether effects associated with educational technology interventions are lasting.

Importantly, analyses revealed wide heterogeneity across studies with decoding, language comprehension, reading comprehension, and writing proficiency outcomes, suggesting that different educational technology interventions yield different effects on these outcomes. Therefore, we conducted moderator analyses to examine whether effects depended on particular characteristics indicated by theory and previous research. Findings from moderator analyses focused on outcome, intervention, and participant characteristics are discussed below. (Given the limited number of studies with writing proficiency outcomes and the relatively limited heterogeneity across these studies, we focused only on outcome characteristics for that domain.)

Effects by Outcome Characteristics

Similar to Xu et al. (2019) but counter to Vanbecelaere et al. (2023), effects on decoding, language comprehension, and reading comprehension differed by how the outcomes were measured. In general, effects on standardized measures were lower than for researcher-developed measures across outcomes. For language comprehension, effects on proximal measures, many of which assessed knowledge of specific vocabulary words taught in the intervention, were significantly higher than effects for distal and standardized measures. As indicated earlier, effect sizes from researcher-developed proximal measures, which we defined as measures that capture what was taught in the intervention, should be viewed with skepticism because they may inflate effect sizes by “teaching to the test,” particularly when the content is not taught in the control or comparison condition (Marulis & Neuman, 2013).

For decoding, language comprehension, reading comprehension, and writing proficiency, effects on distal measures were significantly higher than effects for standardized measures. Researcher-developed distal measures, which we defined as measures that assess taught skills or strategies through untaught content or context, are useful in research because they typically capture skills outside of the specifically taught content in a way that is sensitive to incremental change over shorter periods of time. However, effects on such measures may not reflect the same level of impact as effects on standardized measures, which are typically designed to measure change over the course of an entire academic year or developmental stage. Looking across outcomes, while the effects for distal measures for decoding, language comprehension, and reading comprehension and the effect for standardized measures of decoding and writing would be considered large, according to Kraft (2020), the effect size for standardized measures of language

comprehension and reading comprehension would be considered medium. In other words, the effects on standardized measures of language comprehension and reading comprehension are relatively modest.

As previously discussed, reading comprehension is considered the product of decoding and language comprehension (Gough & Tunmer, 1986). According to Paris (2005), while decoding is considered *constrained* in that it encompasses a narrow set of skills (e.g., in English students learn 26 letters and 44 phonemes that map onto those letters) and can typically be acquired within a relatively condensed developmental window (e.g., kindergarten through 3rd grade), language comprehension skills are considered *unconstrained* in that the set of skills included is not bound (e.g., there are hundreds of thousands of words in the English language) and develop across the lifespan (e.g., vocabulary knowledge develops well beyond elementary school in adolescence and adulthood). The unconstrained nature of language comprehension makes it particularly challenging to teach (Snow & Matthews, 2016). Nevertheless, while decoding is highly predictive of reading comprehension in lower elementary school, language comprehension becomes increasingly important as students become more fluent with decoding and students begin to tackle more complex text across content areas on their own in upper elementary school and into secondary school (Catts et al., 2005; Foorman et al., 2018). Thus, the language comprehension skills that underly reading comprehension should be supported throughout elementary school (Snow & Matthews, 2016).

Given the relatively smaller effect on standardized measures of language comprehension and reading comprehension found in this study, policymakers and educators should have realistic expectations of what effects currently available educational technology interventions may have on standardized measures in these domains. Additionally, researchers and developers should consider how to optimize educational technology interventions to better support generalized language comprehension and reading comprehension skills for elementary school students. Educational technology interventions may need to go beyond focusing on teaching students a limited set of words, as in Korat (2010), which had the largest effect on researcher-developed proximal measures for language comprehension, and instead support students in developing and applying strategies for broad meaning-making, as in L. Lysenko et al. (2019), which had the largest effect on standardized language comprehension measures. Also, the strategy-based approach used in intelligent tutoring interventions had large effects on reading comprehension across studies (e.g., Wijekumar et al., 2013, 2018, 2019) and may provide a useful model for interventions targeting language comprehension as well.

Effects by Intervention Characteristics

Guided by theory and research, we identified particular intervention characteristics to examine as potential moderators of effects of educational technology interventions on literacy-related outcomes. Specifically, we examined effects by the pedagogical orientation of the intervention; whether the intervention included features such as gamification, adaptation, or feedback; the type of technology included in the intervention; whether the intervention included non-technology-based instruction; and whether the intervention included teacher training.

Remarkably, none of these characteristics moderated effects on reading comprehension, and there were few indications that these characteristics moderated effects for decoding and language comprehension. However, several descriptive observations and a few findings from inferential analyses call for discussion.

Descriptively, more studies of educational technology interventions that included decoding outcomes were aligned with a behavioral pedagogical orientation ($k=60$) than a cognitivist ($k=20$) or constructivist ($k=6$) pedagogical orientation. Though effects for these pedagogical orientations were not statistically different for decoding, the prevalence of educational technology interventions that included decoding outcomes with a behavioral pedagogical orientation aligns with non-technology-based research that suggests that constrained skills such as phonics are best taught through explicit and systematic instruction (Al Otaiba et al., 2019; Petscher et al., 2020). Notably, several interventions with large effects on standardized measures of decoding skills are behaviorally-oriented. These interventions include the Headsprout Early Reading (HER) intervention (Storey et al., 2020; Tyler et al., 2015) and an intervention called GraphoGame (Saine et al., 2011). These programs focus heavily on phonics through direct instruction, repeated practice, and corrective feedback. By the same measure, more studies of educational technology interventions that included reading comprehension outcomes were aligned with a cognitive pedagogical orientation ($k=45$) than a behavioral ($k=20$) or constructivist ($k=0$) pedagogical orientation, and the prevalence of educational technology interventions that included reading comprehension outcomes with a cognitive pedagogical orientation aligns with non-technology-based research that suggests that reading comprehension may be best taught through a strategy-based approach (Suggate, 2016). This approach is evident in the numerous studies of ITSS by Wijekumar et al. (e.g., Wijekumar et al., 2013, 2018, 2019) that focus on using text structure-related strategies to support comprehension.

For language comprehension, results from the meta-analysis showed that educational technology programs aligned with a constructivist approach were more effective than those aligned with a behaviorist approach, suggesting that educational technology interventions that use the behaviorist approach may be less effective for supporting language comprehension. (Note that this finding was found in both the individual and multiple meta-regression analyses.) This finding is likely related to another finding from the meta-analysis that indicated that, for language comprehension, educational technology interventions with e-books were more effective than those using computer-assisted instruction. In this study, since they provide open-ended experiences and resources to support learning, most interventions with e-books were classified as constructivist and five of the eight studies with language comprehension outcomes that were categorized as constructivist included e-books. (One intervention that included e-books and embedded learning strategies instruction was classified as cognitive [i.e., Connor et al., 2019].) This finding is consistent with previous meta-analyses (e.g., Savva et al., 2022; Takacs et al., 2015; Zucker et al., 2009) that have found positive effects of e-books on language comprehension-related skills.

In the present meta-analysis, examples of interventions with language comprehension outcomes that were categorized as constructivist and having e-books

include studies by Korat (2010) and Korat et al. (2017). These interventions, with large effect sizes (i.e., above 1.00) on researcher-developed proximal measures, included digital texts with visual and auditory support that students could use to facilitate meaning-making. Importantly, though, these studies did not include standardized measures for this outcome. Interestingly, three studies with the largest effects (i.e., ≥ 0.20) on standardized measures of language comprehension were all categorized as cognitive in orientation. These studies evaluated a program called ABRACADABRA (an acronym that stands for A Balanced Reading Approach for Canadians [now, Children] Designed to Achieve Best Results for All) (Abrami et al., 2016 [0.27]; L. Lysenko et al., 2019 [0.53]; Lysenko & Abrami (2014) [0.21]), which focuses on decoding as well as language and comprehension strategies such as using context clues to figure out word and text meaning and includes scaffolds to support receptive and expressive language development. Note, however, that most studies with standardized measures of language comprehension, including those referenced above, had effect sizes with confidence intervals that included negative values. Only one study (Alvarez-Marinelli et al., 2014), which evaluated a Computer Assisted Language Learning (CALL) program that was classified as cognitive in orientation because it included language and comprehension-building strategies (e.g., question-answer strategies), had an effect size with a confidence interval that did not include 0 ($g=0.21$, 95% CI [0.03, 0.47]).

These results suggest that, as proposed by Dede (2008), different educational technology interventions that align with different pedagogical approaches may be more or less effective depending on the intended outcome. In this meta-analysis, we found few studies that compared multiple interventions within the same sample and context, and the interventions in these studies were typically similar pedagogically. (For example, Ecalle et al. [2013] examined two different types of computer-assisted instruction that were both cognitively oriented.) To guide policymakers and educators seeking to fund or adopt educational technology interventions, research comparing programs with different pedagogical orientations within the same sample and context is needed. Furthermore, interventions that we classified as having similar pedagogies were often very different (e.g., consider the ABRACADABRA and the CALL programs mentioned above). Therefore, we looked beyond pedagogy to specific features of educational technology interventions that could, theoretically, drive differences across programs.

Specifically, we examined whether gamification, adaptation, or feedback moderated effects. There was no evidence for these features as moderators of effects on decoding or reading comprehension. However, feedback was negatively associated with effects on language comprehension. Gamification and adaptation were also negatively associated with effects on language comprehension, though these relationships were not statistically significant. Considering previous meta-analyses that suggest educational technology interventions with feedback, adaptivity, and gamification promote learning (D. B. Clark et al., 2016; Major et al., 2021), we were surprised that these features were not positively and significantly associated with the outcomes we studied. However, Takacs et al. (2015) found that interactive features, including games, hotspots, pop-up questions, and clickable definitions, negatively affected vocabulary and comprehension in their

meta-analysis of electronic storybook interventions implemented with preschool and elementary-age children. Takacs et al. (2015) suggest that interactive features may be distracting and result in cognitive overload. Interestingly, though, Vanbecelaere et al. (2023) found that gamification and adaptivity did not moderate effects of digital reading interventions implemented with children ages 3 to 12. It could be that the role of such features as feedback, adaptivity, and gamification in educational technology interventions differs depending on the age and/or grade of the students, the focus of the intervention (e.g., decoding, language comprehension, etc.), the format of the technology (e.g., e-book or computer-assisted instruction), etc. Also, in this study, while these features were either not correlated or negatively correlated with the cognitive and constructivist pedagogies ($-0.11, p > .05$ to $-0.47, p < .01$), they were positively correlated with the behavioral pedagogy ($0.25, p < .01$ to $0.52, p < .01$). In order to tease out the relative importance of pedagogy or specific features under various conditions, there needs to be much more research on educational technology interventions, including research that systematically varies intervention characteristics. Only two studies in the present meta-analysis included two treatments that varied on feedback, adaptivity, or gamification. In a study with students they identified as “at risk,” Solheim et al. (2018) included two CAI treatments, one with a fixed progression and one with an adaptive progression of games, and found no difference between the treatments. The researchers expressed uncertainty about whether adaptation did not add value due to the content and context of the study or whether adaptation was not at an optimal level to support learning. Boggio et al. (2023) studied a spelling application with and without gamification that included dynamic feedback features and found what they identified as a “marginally positive effect of gamification that is greater for the weakest students” (p. 1332). These researchers highlight the need for more research investigating various features with different types of learners.

Few other intervention characteristics in this meta-analysis revealed significant differences. For decoding, interventions lasting >22 hours were associated with higher effects than those lasting <8 hours; and, in the meta-regression with multiple moderators analysis, interventions of both medium and high duration had greater effects than for those with low duration. With the difficulty we had in determining duration across studies, we view results related to this variable with extreme caution. Only one study in our corpus compared two treatments that differed in duration. In that study, Alvarez-Marinelli et al. (2014) compared two computer-based language learning interventions, one that lasted roughly 28 hours and another that lasted about 53 hours. These authors found higher effects for the shorter treatment on language comprehension outcomes, but the interventions also varied in content. A recent meta-analysis on intervention dosage by Roberts et al. (2022) suggests that intervention effect sizes follow a concave parabolic shape such that effects are positively associated with intervention dosage up to a point after which there are diminishing returns. However, Roberts et al. excluded interventions administered via a computer without a live instructor. More research on educational technology interventions needs to systematically manipulate duration, and future meta-analyses should explore the nonlinear approach used in Roberts et al.

For language comprehension and reading comprehension, studies that included non-technology-based instruction tended to have lower effects than studies without non-technology-based instruction; however, these effects did not hold in the meta-regression models with multiple moderators. Also, since whether non-technology-based instruction was included was correlated with other potential moderators, it could be that one of these relationships was underlying the initial moderating effect of non-technology-based instruction. In light of research suggesting that comprehensive interventions or interventions that include non-technology-based instruction in addition to technology-based instruction are more effective (Cheung & Slavin, 2012; Savva et al., 2022; T. Thomas et al., 2013), more research is needed on the role of non-technology-based instruction in the context of educational technology interventions. Finally, contrary to Archer et al. (2014), we did not find that effects were associated with whether teacher training was discussed, though it could be that many studies in our corpus included teacher training even though it was not described. In order to determine whether teacher training to support implementation of educational technology interventions is important and, if it is, what it should look like, future research should include detailed descriptions of teacher training and studies should systematically vary features of teacher training. Unlike Cheung and Slavin (2012), we did not explore implementation. Again, this information was often absent in manuscripts, but the quality and consistency of implementation should be considered in future meta-analyses.

Effects by Participant Characteristics

The only participant characteristic that surfaced as a moderator of effects in this meta-analysis was socioeconomic status. For decoding, studies with a substantial percentage ($>33\%$) of students from low socioeconomic backgrounds tended to have larger effects compared to other studies. This finding is in line with previous reviews suggesting that the effects of educational technology interventions could differ for students from different socioeconomic backgrounds (e.g., Cheung & Slavin, 2012; Takacs et al., 2015), but it differs from the review by Vanbecelaere et al. (2023) in which socioeconomic status did not moderate effects. Given the inconsistency in reporting on participant characteristics, we recognize the need for much more research to determine whether effects of educational technology interventions have greater effects for students from low socioeconomic backgrounds and whether effects differ by language status and disability status. Few studies in our corpus disaggregated findings for students from these backgrounds. Some studies explored effects of educational technology interventions with specific populations (e.g., Esteves et al. [2010] worked with students with reading disabilities, Lange [2019] worked with students from low-income backgrounds, and Wijekumar et al. [2019] worked with English learners), but there were not enough studies focused on each population to run analyses separately for these studies. Given the heterogeneity of students in schools (Digital Promise, 2016), future research on educational technology interventions needs to be clear about student demographics and evaluate whether interventions are equally effective for supporting the literacy development of elementary school students from different backgrounds. Of note, though not a substantive focus in this study, we

did not find differences according to whether studies were conducted in the United States ($k=50$) or internationally ($k=70$).

Effects by Design Characteristics

Though Cheung and Slavin (2012) found that effects differed by whether the study included random assignment or not), Xu et al. (2019) found that effects differed by the nature of the comparison group (e.g., whether the comparison group participated in business as usual or other instruction), and Vanbecelaere et al. (2023) found differences by study quality, a measure that included design characteristics, we did not find differences according to design characteristics. Theoretically, these characteristics could influence effect size estimates (e.g., Kraft, 2020; Lipsey & Wilson, 2001), but they have not been shown to play a moderating role in other recent meta-analyses investigating literacy interventions (e.g., Gersten et al., 2020; Hall et al., 2023; Neitzel et al., 2022; Silverman et al., 2020).

Limitations

An inherent limitation of meta-analysis is that findings are generalized from the existing body of research. There are numerous educational technology interventions that have not been studied by researchers who disseminate findings through academic journals or dissertations. Research on a wider range of educational technology interventions is needed to better understand how such interventions affect literacy. Studies disseminated through other avenues that have not undergone peer review by independent researchers are much less trustworthy. Another ingrained limitation is that moderator analyses within the context of meta-analyses are merely exploratory in that, in most studies, moderators have not been systematically manipulated. Within the scope of studies we identified, effects on literacy-related skills were likely confounded with the specific interventions, participants, measures, and methods used. For example, there were positive effects of constructivist-oriented educational technology interventions as well as e-book-based interventions on language comprehension. Since e-book interventions were typically classified as constructivist because they offer open-ended opportunities for children to experience and learn from text, it is impossible to know whether the effect on language comprehension is driven by the pedagogy or technology type. Additionally, some studies that included e-books (Connor et al., 2019; Korat, 2010) only examined effects on proximal researcher-developed measures, confounding intervention characteristics with measurement type.

This meta-analysis is further limited because we only examined selected intervention and participant characteristics to explore in moderator analyses. Though our choice of potential moderators was theoretically and empirically driven, there are many other characteristics of educational technology interventions that may be relevant (Xie et al., 2018). Furthermore, for each characteristic, we kept categories mutually exclusive and, when educational technology interventions included features aligned with more than one category, we chose the category that was the best match. For example, though the e-books that Korat (2010) studied included some direct instruction, which could be categorized as behaviorally-oriented, we determined that the open-ended nature of the e-books

made the constructivist orientation the best match for this educational technology intervention. As Dede (2008) noted, many educational technology interventions do not neatly align with particular characteristics. By using mutually exclusive categories, we ignored ways in which the interventions could have aligned with more than one category. To fully understand the impact of these characteristics on the effectiveness of educational technology interventions, much more fine-grained analyses of intervention characteristics are needed. Such analyses will require a corpus of studies with sufficient variation across potential moderators.

Additionally, this meta-analysis was limited by the information provided in the manuscripts we identified. As we note, descriptions of interventions and participant characteristics were often sketchy or unclear. This was particularly problematic for participant characteristics. Many studies did not report on key participant characteristics (e.g., race or gender) or did not clarify how characteristics were defined (e.g., socioeconomic background and language or ability status). In order to determine whether educational technology interventions can accelerate literacy development for marginalized or under-served groups, research with clearly defined participant samples and comparisons by subgroups within samples is needed.

Conclusion

Educational technology interventions that claim to support literacy are increasingly prevalent. This meta-analysis can help guide policymakers and educators in making decisions about funding and adopting these interventions. However, while the present meta-analysis suggests positive effects for educational technology interventions and provides indications of potential moderators of these effects for literacy-related outcomes in elementary school, it also calls attention to the fact that more research with clear descriptions of intervention and study characteristics is needed across literacy outcomes.

ORCID iDs

Rebecca D. Silverman  <https://orcid.org/0000-0002-9785-0313>

Kristin Keane  <https://orcid.org/0000-0002-4300-2749>

Notes

¹ We conducted analyses using R (R Core Team, 2023) and RStudio (RStudio Team, 2023) and the following statistical packages in our effect size analyses: metafor, MAD, dmetar, effectsize, and robumeta (Ben-Shachar *et al.*, 2020; Fisher & Tipton, 2015; Harrer *et al.*, 2019; Hoyt ACDR&WT, 2014; van Lissa, 2017; Viechtbauer, 2010).

² In some cases, authors did not provide student-level data in the manuscript. In these cases, we contacted the authors to access the student-level data for our analyses.

³ This design effect adjustment reflects the increased uncertainty due to the clustered nature of the data (Heo *et al.*, 2010; Särndal *et al.*, 2003). Using the average cluster size is useful here as it serves as a conservative upper bound of the exact design effect inflation in case of studies where cluster sizes are not identical (e.g., slightly uneven classroom sizes) (Gabler *et al.*, 1999). The average cluster size could be that of the classroom or that of the school, based on the level of cluster randomization defined in each study.

⁴ Since we aggregated effects across treatments for the purposes of the meta-analysis, we excluded studies from the analysis of a particular moderator when the study included two treatments that differed according to this moderator.

⁵ Kraft (2020) proposed the following benchmarks for PK–12 education intervention effect sizes: <0.05 is small, 0.05 to <0.20 is medium, and ≥ 0.20 is large.

References

- Abrami, P. C., Wade, C. A., Lysenko, L., Marsh, J., & Gioko, A. (2016). Using educational technology to develop early literacy skills in Sub-Saharan Africa. *Education and Information Technologies*, 21(4), 945–964. <https://doi.org/10.1007/s10639-014-9362-4>
- Aktaş, N., & Akyol, H. (2020). Effect of digital writing workshop activities on writing motivation and development of story writing skills. *International Journal of Progressive Education*, 16(3), 270–287. <https://doi.org/10.29329/ijpe.2020.248.20>
- Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, 90(1), 6–23. <https://doi.org/10.3102/0034654319854352>
- Al Otaiba, S., Allor, J. H., Baker, K., Conner, C., Stewart, J., & de la Cruz, V. M. (2019). Teaching phonemic awareness and word reading skills focusing on explicit and systematic approaches. *Perspectives on Language and Literacy*, 45(3), 11–16. <https://files.eric.ed.gov/fulltext/ED604475.pdf>
- Alvarez-Marinelli, H., Blanco, M., Lara-Alecio, R., Irby, B. J., Tong, F., Stanley, K., & Fan, Y. (2014). Computer assisted English language learning in Costa Rican elementary schools: An experimental study. *Computer Assisted Language Learning*, 29(1), 103–126. <https://doi.org/10.1080/09588221.2014.903977>
- Archer, K., Savage, R., Sanghera-Sidhu, S., Wood, E., Gottardo, A., & Chen, V. (2014). Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis. *Computers and Education*, 78, 140–149. <https://doi.org/10.1016/j.compedu.2014.06.001>
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. <https://doi.org/10.1126/science.1736359>
- Bahari, A. (2021). Computer-mediated feedback for L2 learners: Challenges versus affordances. *Journal of Computer Assisted Learning*, 37(1), 24–38. <https://doi.org/10.1111/jcal.12481>
- Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41(2), 257–278. <https://doi.org/10.1111/j.2044-8317.1988.tb00901.x>
- Ben-Shachar, M. S., Lüdtke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review*, 33(4), 1675–1715. <https://doi.org/10.1007/s10648-021-09615-8>

- Berninger, V., Abbott, R., Cook, C. R., & Nagy, W. (2017). Relationships of attention and executive functions to oral language, reading, and writing skills and systems in middle childhood and early adolescence. *Journal of Learning Disabilities*, 50(4), 434–449. <https://doi.org/10.1177/0022219415617167>
- Boggio, C., Zaher, A., & Bosse, M.-L. (2023). ECRIMO, an app to train first graders' spelling: Effectiveness and comparison between different designs. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, 54(5), 1332–1350. <https://doi.org/10.1111/bjet.13354>
- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). Russell Publishing Sage Foundation.
- Bos, L. T., De Koning, B. B., Wassenburg, S. I., & van der Schoot, M. (2016). Training inference making skills using a situation model approach improves reading comprehension. *Frontiers in Psychology*, 7, 116. <https://doi.org/10.3389/fpsyg.2016.00116>
- Boykin, A., Evmenova, A. S., Regan, K., & Mastropieri, M. (2019). The impact of a computer-based graphic organizer with embedded self-regulated learning strategies on the argumentative writing of students in inclusive cross-curricula settings. *Computers and Education*, 137, 78–90. <https://doi.org/10.1016/j.compedu.2019.03.008>
- Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities; The connections between language and reading disabilities* (pp. 25–40). Lawrence Erlbaum Associates Publishers.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332. https://doi.org/10.1207/s1532690xc0804_2
- Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3), 198–215. <https://doi.org/10.1016/j.edurev.2012.05.002>
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. <https://doi.org/10.3102/0034654315582065>
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29. <https://doi.org/10.1007/BF02299088>
- Common Sense Education. (2018). *Apps and websites to build literacy skills: Top picks from Common Sense Education*. <https://www.readingrockets.org/article/apps-and-websites-build-literacy-skills-top-picks-common-sense-education>
- Connor, C. M. D., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., & Hwang, J. K. (2019). Building word knowledge, learning strategies, and metacognition with the word-knowledge e-book. *Computers and Education*, 128, 284–311. <https://doi.org/10.1016/j.compedu.2018.09.016>
- Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, 4(3), 173–207. <https://doi.org/10.1080/19345747.2010.510179>
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Publishing Sage Foundation.

- Corcoran, B. (2021, July 30). The next wave of EdTech will be very, very níg- and global. *Edsurge*. <https://www.edsurge.com/news/2021-07-30-the-next-wave-of-edtech-will-be-very-very-big-and-global>
- Dede, C. (2008). Theoretical perspectives influencing the use of information technology in teaching and learning. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 43–62). Springer.
- De La Paz, S., & Butler, C. (2018). Promoting motivated writers: Suggestions for teaching and conducting research with students with learning disabilities and struggling learners. *Learning Disabilities: A Multidisciplinary Journal*, 23(2), 56–69. <https://doi.org/10.18666/LDMJ-2018-V23-I2-9064>
- Digital Promise. (2016). *The growing diversity in today's classroom*. https://digitalpromise.org/wp-content/uploads/2016/09/lps-growing_diversity_FINAL-1.pdf
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, 56(S1), S25–S44. <https://doi.org/10.1002/rrq.411>
- Ecalle, J., Kleinsz, N., & Magnan, A. (2013). Computer-assisted learning in young poor readers: The effect of grapho-syllabic training on the development of word reading and reading comprehension. *Computers in Human Behavior*, 29(4), 1368–1376. <https://doi.org/10.1016/j.chb.2013.01.041>
- Escueta, M., Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4), 897–996. <https://doi.org/10.1257/jel.20191507>
- Esteves, K. J., & Whitten, E. (2010). Assisted reading with digital audiobooks for students with reading disabilities. *Reading Horizons*, 51(1), 21–40. https://scholarworks.wmich.edu/reading_horizons/vol51/iss1/4
- Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. *ArXiv/abs/1503.02220*.
- Fittes, E. K. (2022, January 10). Global venture capital investment in ed-tech soared in 2021, hitting \$20B. *EdWeek Market Brief*. <https://marketbrief.edweek.org/market-place-k-12/global-venture-capital-investment-ed-tech-soared-2021-hitting-20b/>
- Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, 23(4), 553–576. <https://doi.org/10.1007/s10648-011-9175-6>
- Florit, E., Roch, M., Dicaldo, R., & Levorato, M. C. (2022). The Simple View of Reading in Italian beginner readers: Converging evidence and open debates on the role of the main components. *Learning and Individual Differences*, 93, 101961. <https://doi.org/10.1016/j.lindif.2020.101961>
- Fogarty, M., Coyne, M. D., Simmons, L. E., Simmons, D. C., Henri, M., Kwok, O. M., Ware, S. M., Dalton, K., Williams, K. A., & Wang, H. (2020). Effects of technology-mediated vocabulary intervention for third-grade students with reading difficulties. *Journal of Research on Educational Effectiveness*, 13(2), 271–297. <https://doi.org/10.1080/19345747.2019.1698086>
- Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences*, 63, 12–23. <https://doi.org/10.1016/j.lindif.2018.02.011>
- Gabler, S., Häder, S., & Lahiri, P. (1999). A model based justification of Kish's formula for design effects for weighting and clustering. *Survey Methodology*, 25(1), 105–106.

- https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999001/article/4718-eng.pdf?st=Q5j_vJLB
- Gallup. (2019). *Education technology use in schools: Student and educator perspectives*. <http://www.newschools.org/wp-content/uploads/2019/09/Gallup-Ed-Tech-Use-in-Schools-2.pdf>
- Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, 13(2), 401–427. <https://doi.org/10.1080/19345747.2019.1689591>
- Görge, R., Huemer, S., Schulte-Körne, G., & Moll, K. (2020). Evaluation of a digital game-based reading training for German children with reading disorder. *Computers and Education*, 150, 103834. <https://doi.org/10.1016/j.compedu.2020.103834>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. <https://doi.org/10.1177/074193258600700104>
- Graham, S., Kiuahara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, 90(2), 179–226. <https://doi.org/10.3102/0034654320914744>
- Graham, S., McKeown, D., Kiuahara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104(4), 879–896. <https://doi.org/10.1037/a0029185>
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. Carnegie Corporation of New York. <https://www.carnegie.org/publications/writing-next-effective-strategies-to-improve-writing-of-adolescents-in-middle-and-high-schools/>
- Gray, L., & Lewis, L. (2021). *Use of educational technology for instruction in public schools: 2019–20* (NCES 2021–017). United States Department of Education, National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021017>
- Hall, C., Dahl-Leonard, K., Cho, E., Solari, E. J., Capin, P., Conner, C. L., Henry, A. R., Cook, L., Hayes, L., Vargas, I., Richmond, C. L., & Kehoe, K. F. (2023). Forty years of reading intervention research for elementary students with or at risk for dyslexia: A systematic review and meta-analysis. *Reading Research Quarterly*, 58(2), 285–312. <https://doi.org/10.1002/rrq.477>
- Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *Dmetar: Companion R Package for the guide “Doing meta-analysis in R”*. R package version 0.1.0. <http://dmetar.protectlab.org/>
- Hebbecke, K., Förster, N., & Souvignier, E. (2019). Reciprocal effects between reading achievement and intrinsic and extrinsic reading motivation. *Scientific Studies of Reading*, 23(5), 419–436. <https://doi.org/10.1080/10888438.2019.1598413>
- Hedges, L. V. (1981). Distribution theory for Glass’s estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490–499. <https://doi.org/10.1037/0033-2909.92.2.490>
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>

- Heo, M., Kim, Y., Xue, X., & Kim, M. Y. (2010). Sample size requirement to detect an intervention effect at the end of follow-up in a longitudinal cluster randomized trial. *Statistics in Medicine*, 29(3), 382–390. <https://doi.org/10.1002/sim.3806>
- Higgins, J. P. T., & Green, S. (Eds.) (2022). *Cochrane handbook for systematic reviews of interventions* (version 6.3). <https://training.cochrane.org/handbook/archive/v5.0.0/>
- HolonIQ. (2020, January 28). *\$87bn+ of global edtech funding predicted through 2030. \$32bn last decade*. Author. <https://www.holoniq.com/notes/87bn-of-global-edtech-funding-predicted-to-2030>
- Hoyt ACDR&WT. (2014). *MAd: Meta-analysis with mean differences*. <https://CRAN.R-project.org/package=MAd>
- International Literacy Association (ILA). (2015). *The cost of illiteracy*. <https://www.literacyworldwide.org/docs/default-source/resource-documents/ila-take-action-costs.pdf>
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78(4), 243–255. <https://doi.org/10.1037/0022-0663.78.4.243>
- Kim, Y. G. (2020). Toward integrative reading science: The direct and indirect effects model of reading. *Journal of Learning Disabilities*, 53(6), 469–491. <https://doi.org/10.1177/0022219420908239>
- Kim, Y. G. (2023). Learning to read and write. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds), *International encyclopedia of education* (Vol. 6, 4th ed., pp. 335–343). Elsevier.
- Kim, Y. G., & Graham, S. (2022). Expanding the direct and indirect effects model of writing (DIEW): Reading-writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, 114(2), 215–238. <https://doi.org/10.1037/edu0000564>
- Kim, Y. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing*, 32, 1319–1343. <https://doi.org/10.1007/s11145-018-9913-y>
- Kish, L. (1965). *Survey sampling*. John Wiley and Sons Inc.
- Klingbeil, D. A., McComas, J. J., Burns, M. K., & Helman, L. (2015). Comparison of predictive validity and diagnostic accuracy of screening measures of reading skills. *Psychology in the Schools*, 52(5), 500–514. <https://doi.org/10.1002/pits.21839>
- Korat, O. (2010). Reading electronic books as a support for vocabulary, story comprehension and word reading in kindergarten and first grade. *Computers & Education*, 55(1), 24–31. <https://doi.org/10.1016/j.compedu.2009.11.014>
- Korat, O., Kozlov-Peretz, O., & Segal-Drori, O. (2017). Repeated e-book reading and its contribution to learning new words among kindergartners. *Journal of Education and Training Studies*, 5(7), 60–72. <https://doi.org/10.11114/jets.v5i7.2498>
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19. <https://doi.org/10.1007/BF02299087>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, 1(1), 59–74. [https://doi.org/10.1016/0747-5632\(85\)90007-X](https://doi.org/10.1016/0747-5632(85)90007-X)

- Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods, 10*(1), 83–98. <https://doi.org/10.1002/jrsm.1316>
- Lange, A. A. (2019). Technology, instructional methods, and the systemic messiness of innovation: Improving reading fluency for low socio-economic elementary school students. *Educational Technology Research and Development, 67*(5), 1333–1350. <https://doi.org/10.1007/s11423-019-09675-2>
- Lee, S., Kuo, L., Xu, Z., & Hu, X. (2020). The effects of technology-integrated classroom instruction on k-12 English language learners' literacy development: A meta-analysis. *Computer Assisted Language Learning, 35*(5–6), 1106–1137. <https://doi.org/10.1080/09588221.2020.1774612>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.
- Lysenko, L. V., & Abrami, P. C. (2014). Promoting reading comprehension with the use of technology. *Computers and Education, 75*, 162–172. <https://doi.org/10.1016/j.compedu.2014.01.010>
- Lysenko, L., Abrami, P. C., Wade, C. A., Marsh, J. P., WaGioko, M., & Kiforo, E. (2019). Promoting young Kenyans' growth in literacy with educational technology: A tale of two years of implementation. *International Journal of Educational Research, 95*, 176–189. <https://doi.org/10.1016/j.ijer.2019.02.013>
- Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology, 52*(5), 1935–1964. <https://doi.org/10.1111/bjet.13116>
- Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(3), 223–262. <https://doi.org/10.1080/19345747.2012.755591>
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology, 53*(1), 17–29. <https://doi.org/10.1348/000711000159150>
- National Governors Association Center for Best Practices & Council of Chief State School Officers (NGA & CCSSO). (2010). *Common core state standards*. Authors.
- National Institute of Child Health and Human Development (NICHD). (2010). *Developing early literacy: Report of the national early literacy panel*. US Government Printing Office. <https://www.nichd.nih.gov/publications/product/346>
- Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly, 57*(1), 149–179. <https://doi.org/10.1002/rrq.379>
- Organization for Economic Development and Cooperation (OECD). (2022). *Education at a glance 2022: OECD indicators*. Author. <https://doi.org/10.1787/3197152b-en>
- Organization for Economic Development and Cooperation (OECD). (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. Author. <https://doi.org/10.1787/c74f03de-en>
- Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202. <https://doi.org/10.1598/RRQ.40.2.3>

- Peng, P., Lee, K., Luo, J., Li, S., Joshi, R. M., & Tao, S. (2021). Simple view of reading in Chinese: A one-stage meta-analytic structural equation modeling. *Review of Educational Research*, 91(1), 3–33. <https://doi.org/10.3102/0034654320964198>
- Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., Lonigan, C. J., Phillips, B. M., Schatschneider, C., Steacy, L. M., Terry, N. P., & Wagner, R. K. (2020). How the science of reading informs 21st-century education. *Reading Research Quarterly*, 55(S1), S267–S282. <https://doi.org/10.1002/rrq.352>
- Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, 90(1), 24–46. <https://doi.org/10.3102/0034654319877153>
- R Core Team. (2023). *R: A language and environment for statistical computing*. The R Foundation. <https://www.r-project.org/>
- Rideout, V., & Robb, M. B. (2020, November 17). The common sense census: Media use by kids age zero to eight, 2020. *Common Sense Media*. <https://www.common-sensemedia.org/research/the-common-sense-census-media-use-by-kids-age-zero-to-eight-2020>
- Ritterfeld, U., & Lüke, T. (2021). Audio stories as incidental language teachers: A compensatory approach for migrant and low-SES children in Germany. *Journal of Media Psychology: Theories, Methods, and Applications*, 33(2), 60–71. <https://doi.org/10.1027/1864-1105/a000281>
- Roberts, G. J., Dumas, D. G., McNeish, D., & Coté, B. (2022). Understanding the dynamics of dosage response: A nonlinear meta-analysis of recent reading interventions. *Review of Educational Research*, 92(2), 209–248. <https://doi.org/10.3102/00346543211051423>
- RStudio Team. (2023). *RStudio: Integrated development for R*. R. Studio. PBC, MA URL.
- Saine, N. L., Lerkkanen, M. K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Development*, 82(3), 1013–1028. <https://doi.org/10.1111/j.1467-8624.2011.01580.x>
- Sarica, H. Ç., & Usluel, Y. K. (2016). The effect of digital storytelling on visual memory and writing skills. *Computers and Education*, 94, 298–309. <https://doi.org/10.1016/j.compedu.2015.11.016>
- Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.
- Savva, M., Higgins, S., & Beckmann, N. (2022). Meta-analysis examining the effects of electronic storybooks on language and literacy outcomes for children in grades Pre-K to grade 2. *Journal of Computer Assisted Learning*, 38(2), 526–564. <https://doi.org/10.1111/jcal.12623>
- Segal-Drori, O., Korat, O., Shamir, A., & Klein, P. S. (2010). Reading electronic and printed books with and without adult instruction: Effects on emergent reading. *Reading and Writing*, 23(8), 913–930. <https://doi.org/10.1007/s11145-009-9182-x>
- Shamir, A., & Korat, O. (2015). Educational electronic books for supporting emergent literacy of kindergarteners at-risk for reading difficulties—What do we know so far? *Computers in the Schools*, 32(2), 105–121. <https://doi.org/10.1080/07380569.2015.1027868>
- Silverman, R. D., Johnson, E., Keane, K., & Khanna, S. (2020). Beyond decoding: A meta-analysis of the effects of language comprehension interventions on K–5

- students' language and literacy outcomes. *Reading Research Quarterly*, 55(S1), S207–S233. <https://doi.org/10.1002/rrq.346>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681. <https://doi.org/10.1177/1745691614553988>
- Solheim, O. J., Frijters, J. C., Lundetræ, K., & Uppstad, P. H. (2018). Effectiveness of an early reading intervention in a semi-transparent orthography: A group randomised controlled trial. *Learning and Instruction*, 58, 65–79. <https://doi.org/10.1016/j.learninstruc.2018.05.004>
- Slavin, R. E. (1995). A model of effective instruction. *Educational Forum*, 59(2), 166–176. <https://www.proquest.com/scholarly-journals/model-effective-instruction/docview/62747035/se-2> <https://doi.org/10.1080/00131729509336383>
- Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *Future of Children*, 26(2), 57–74. <https://doi.org/10.1353/foc.2016.0012>
- Sparks, R. L. (2021). Identification and characteristics of strong, average, and weak foreign language readers: The simple view of reading model. *Modern Language Journal*, 105(2), 507–525. <https://doi.org/10.1111/modl.12711>
- Spear-Swerling, L. (2019). Structured literacy and typical literacy practices: Understanding differences to create instructional opportunities. *TEACHING Exceptional Children*, 51(3), 201–211. <https://doi.org/10.1177/0040059917750160>
- Storey, C., McDowell, C., & Leslie, J. C. (2020). Headsprout early reading for specific literacy difficulty: A comparison study. *Journal of Behavioral Education*, 29(3), 619–633. <https://doi.org/10.1007/s10864-019-09336-7>
- 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>
- Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research*, 85(4), 698–739. <https://doi.org/10.3102/0034654314566989>
- Thomas, D., & Dyches, J. (2019). The hidden curriculum of reading intervention: A critical content analysis of Fountas and Pinnell's leveled literacy intervention. *Journal of Curriculum Studies*, 51(5), 601–618. <https://doi.org/10.1080/00220272.2019.1616116>
- Thomas, T., Alexander, K., Jackson, R., & Abrami, P. C. (2013). The differential effects of interactive versus didactic pedagogy using computer-assisted instruction. *Journal of Educational Computing Research*, 49(4), 403–436. <https://doi.org/10.2190/EC.49.4.a>
- Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559–1573. <https://doi.org/10.1002/sim.1187>
- Tobias, S., Fletcher, J. D., Bediou, B., Wind, A. P., & Chen, F. (2014). Multimedia learning with computer games. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 762–784). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.037>
- Tyler, E. J., Hughes, J. C., Beverley, M., & Hastings, R. P. (2015). Improving early reading skills for beginning readers using an online programme as supplementary instruction. *European Journal of Psychology of Education*, 30(3), 281–294. <https://doi.org/10.1007/s10212-014-0240-7>

- Vanbecelaere, S., Said-Metwaly, S., Van den Noortgate, W., Reynvoet, B., & Depaep, F. (2023). The effectiveness of Tier 1 digital interventions for early reading: A meta-analysis. *British Journal of Educational Technology*, 54(5), 1397–1432. <https://doi.org/10.1111/bjet.13351>
- Van Lissa, C. J. (2017, September 29). *MetaForest: Exploring heterogeneity in meta-analysis using random forests*. <https://doi.org/10.31234/osf.io/myg6s>
- Vega, V., & Robb, M. B. (2019, April 18). The common sense census: Inside the 21st-century classroom. *Common Sense Media*. <https://www.commonsensemedia.org/research/the-common-sense-census-inside-the-21st-century-classroom>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, “all must have prizes.” *Psychological Bulletin*, 122(3), 203–215. <https://doi.org/10.1037/0033-2909.122.3.203>
- Weng, P., Maeda, Y., & Bouck, E. C. (2014). Effectiveness of cognitive skills-based computer-assisted instruction for students with disabilities: A synthesis. *Remedial and Special Education*, 35(3), 167–180. <https://doi.org/10.1177/0741932513514858>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE).
- Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2013). High-fidelity implementation of web-based intelligent tutoring system improves fourth and fifth graders content area reading comprehension. *Computers and Education*, 68, 366–379. <https://doi.org/10.1016/j.compedu.2013.05.021>
- Wijekumar, K. K., Meyer, B. J. F., Lei, P., Beerwinkle, A. L., & Joshi, M. (2019). Supplementing teacher knowledge using web-based intelligent tutoring system for the text structure strategy to improve content area reading comprehension with fourth- and fifth-grade struggling readers. *Dyslexia: An International Journal of Research and Practice*, 26(2), 120–136. <https://doi.org/10.1002/dys.1634>
- Wijekumar, K. K., Meyer, B. J. F., Lei, P., Hernandez, A. C., & August, D. L. (2018). Improving content area reading comprehension of Spanish speaking English learners in grades 4 and 5 using web-based text structure instruction. *Reading and Writing*, 31(9), 1969–1996. <https://doi.org/10.1007/s11145-017-9802-9>
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2
- Xie, K., Di Tosto, G., Chen, S., & Vongkulluksn, V. W. (2018). A systematic review of design and technology components of educational digital resources. *Computers and Education*, 127, 90–106. <https://doi.org/10.1016/j.compedu.2018.08.011>
- Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology*, 50(6), 3119–3137. <https://doi.org/10.1111/bjet.12758>
- Zhang, Y. (2022). Incorporating peer response with teacher feedback in teaching writing to EFL learners: A literature review. *English Language Teaching*, 15(3), 48–53. <https://www.proquest.com/scholarly-journals/incorporating-peer-response-with-teacher-feedback/docview/2674222379/se-2>

Zucker, T., Moody, A. K., & McKenna, M. C. (2009). The effects of electronic books on pre-kindergarten-to-grade 5 students' literacy and language outcomes: A research synthesis. *Journal of Educational Computing Research*, 40(1), 47–87. <https://doi.org/10.2190/EC.40.1.c>

Authors

REBECCA D. SILVERMAN (corresponding author) is the Judy Koch Professor of Education at the Graduate School of Education at Stanford University, 485 Lasuen Mall, Stanford, CA 94305, USA; e-mail rdsilver@stanford.edu. She conducts research on the role of language comprehension in reading development throughout childhood.

KRISTIN KEANE is a postdoctoral researcher at the Graduate School of Education at Stanford University, California, USA; e-mail knkeane@stanford.edu. She uses mixed-methods approaches to studying literacy development and instruction in elementary school.

ELENA DARLING-HAMMOND is a doctoral candidate in the Graduate School of Education at Stanford University, California, USA; e-mail elenadh@stanford.edu. She is interested in the role of technology in supporting the literacy development of diverse learners.

SAURABH KHANNA is an assistant professor of communication science at the University of Amsterdam; e-mail saurabh.khanna@uva.nl. He works on developing methods to quantify the diversity of information on the internet and is leveraging quantitative methods to inform educational policy.