Reading comprehension on handheld devices vs. on paper: A narrative review and meta-analysis of the medium effect and its moderators

Ladislao Salmerón [1], Lidia Altamura [1], Pablo Delgado [2], Anastasia Karagiorgi [3], Cristina Vargas [1]

[1] University of Valencia (Spain)

[2] University of Seville (Spain)

[3] Julius-Maximilians-Universität Würzburg (Germany)

This is a pre-print version. Please check the final version published at Journal of Educational Psychology © https://dx.doi.org/10.1037/edu0000830

Author Note:

Funding for this research was provided by Grant PID2020-118512GB-I00 funded by MCIN/AEI/ 10.13039/501100011033, financed by the Spanish Ministry of Science and Innovation.

Correspondence concerning this article should be addressed to Ladislao Salmerón, University of Valencia.

Contact: Ladislao.Salmeron@valencia.edu

As handheld devices, such as tablets, become a common tool in schools, a critical and urgent question for the research community is to assess their potential impact on educational outcomes. Previous meta-analytic research has evidenced the 'screen inferiority effect': readers tend to understand texts slightly worse when reading on-screen than when reading the same text in print. Most primary studies from those meta-analyses used computers as on-screen reading devices. Accordingly, the extent to which handheld devices, which provide a reading experience closer to books than computers, are affected by the screen inferiority effect remains an open question. To address this issue, we reviewed relevant literature regarding potential moderating factors for the screen inferiority effect through the lenses of the Reading for Understanding framework. We then performed two meta-analyses aimed at examining the differences in reading comprehension when reading on handheld devices, as compared to print. Results from the two multilevel random-effect meta-analyses, which included primary studies that used either between-participants (k = 38, g = -.113) or within-participants (k = 21, g = -.103) designs consistently showed a significant small size effect favoring print text comprehension. Moderator analyses helped to partially clarify the results, indicating in some cases a higher screen inferiority effect for undergraduate students (as compared to primary and secondary school students) and for participants that were assessed individually (as opposed to in groups). We discuss the need to continue fostering print reading in schools while developing effective ways to incorporate handheld devices for reading purposes.

Educational impact and implications statement: Handheld devices, such as tablets, are widely used in schools as reading devices. Our synthesis of existing studies indicates that readers comprehend slightly better when they read a text in print as

opposed to on a handheld device. In-print reading can be considered as an efficient way to promote students’ text comprehension.

Keywords: meta-analysis, reading comprehension, reading media differences, on-screen reading, in-print reading

The question whether readers understand similarly when reading on screen or in print has been the focus of theoretical discussions and empirical research during the last three decades (Baron, 2015, 2021; Dillon, 1992; Kingston, 2008; Noyes & Garland, 2008; Singer & Alexander, 2017; Wang et al., 2007; Wolf, 2018). Although the evidence is still mixed regarding the role of the reading medium in text comprehension, five recent meta-analyses have identified the screen inferiority effect: readers across a wide range of ages tend to understand texts slightly better in print than on screen (Clinton, 2019; Delgado et al., 2018; Furenes et al., 2021; Kong et al., 2018; Öztop & Nayci, 2021). The effect is more pronounced when texts are expository rather than narrative (Clinton, 2019; Delgado et al., 2018; Schwabe et al., 2022), and when reading time is limited (Delgado et al, 2018). Notwithstanding, these meta-analyses also show that the heterogeneity among effect sizes is rather high, meaning that there are still many facets of the role of the reading medium in comprehension that we still fail to understand. Since the publication of those reviews and meta-analyses there has been a surge in the number of empirical studies that aimed not just to test the screen inferiority effect, but rather to understand it by testing potential underlying mechanisms. Accordingly, the first goal of our study is to provide an updated narrative review of relevant characteristics for the effects of reading medium on comprehension, based on the well known reading for understanding framework (Snow et al., 2002).

Since tablets were created in the early 2010s they have increasingly become popular in schools, where they coexist with computers. In a review of 43 studies that documented the introduction of tablets in classrooms, Boon et al. (2021) concluded that a majority of teachers and students were positively disposed towards using such devices in educational settings. Nevertheless, such enthusiasm is not supported by solid empirical research, as recent reviews noted mixed evidence for the effects of tablets on several competences and discipline areas, including reading comprehension (Boon et al., 2021; Haßler et al., 2016). Accordingly, the second goal of our study is to fill this gap, by synthesizing studies that specifically compared the comprehension outcomes after reading on tablets or e-readers vs. printed texts.

A narrative review of the reading medium effect and its potential moderators

In the following, we shall start by describing previous reviews on the association between handheld devices (mostly tablets) and comprehension and learning. Next, we review relevant literature that has analyzed potential moderating factors for the screen inferiority effect, including reader, text, activity, and context related factors. Finally, we report the two meta-analyses performed and discuss the results, emphasizing the extent to which they can explain the screen-inferiority effect on handheld devices.

Previous research syntheses on the effects of handheld devices on text comprehension and learning

Previous meta-analyses on the effects of reading medium on comprehension analyzed studies that used a variety of digital devices, including desktop computers with CRT or LED screens, laptop computers, PDAs, smartphones, tablets, and e-readers.

Two meta-analyses evaluated the extent to which the type of device moderated the screen inferiority effect, by grouping studies that used either computers or handheld devices (Delgado et al., 2018; Öztop & Nayci, 2021). In both cases, the effects were not significant, indicating that the screen inferiority effect did not differ as a function of digital device. However, these results must be taken with caution, as most of the studies that have compared print vs. on-screen reading used computers as a digital device. Indeed, the number of studies included in the above-mentioned meta-analyses that used handheld devices was limited (n = 12 [14 effect sizes] in Delgado et al., 2018; n = 2 in Öztop & Nayci, 2021). In addition, among the 12 studies analyzed by Delgado et al.

(2018) there were studies that used handheld devices other than tablets, such as cell phones or PDAs. Similarly, in a recent meta-analysis comparing print and on-screen reading of narrative texts, Schwabe et al. (2022) also tested the moderator effect of the different digital reading devices. Following the previous trends, they included more computers (n = 20) than tablets (n = 6), or e-readers (n = 9). Paired comparisons between the three devices did not reveal any significant effect**.**

Few relevant systematic reviews have been conducted to explore the relationships between tablet use and learning outcomes. Haßler et al. (2016) reviewed 23 studies reporting associations between tablet use in the classroom and learning across different disciplines. While most studies reported positive effects of using tablets, the results were mixed. More importantly, the authors identified severe limitations in the research methods used in the original studies, which made it difficult to draw firm conclusions. In a later review, Boon et al. (2021) synthesized 43 studies that analyzed the effect of iPads on learning in children aged 9-14 years old. Contrary to the authors' expectations, the analysis revealed that the use of tablets for major school learning areas did not consistently enhance academic outcomes. In addition, the authors found that most studies employed exploratory research designs with no control group, which limits their validity and generalizability.

Other studies have analyzed the relation between tablet use in classrooms and student reading comprehension achievement. Using a large dataset from OECD PISA 2018, Bryant et al. (2020) examined the relationship between the use of tablets, laptops, and desktop computers in the classroom and students' reading comprehension scores. After controlling for student socioeconomic status (SES), school type, and country, the authors found that students' use of tablets in the classroom was negatively associated with performance (-21 points, approx. half of a grade level learning). When tablets were used in class only by teachers, but not by students, the relationship became positive, a pattern that emphasizes students' use as a major responsible factor for the devices' effects.

In sum, previous evidence is not conclusive regarding the effects of tablets on comprehension and learning across disciplines. Nevertheless, as reviewed, recent correlational evidence suggests a negative association for reading comprehension achievement.

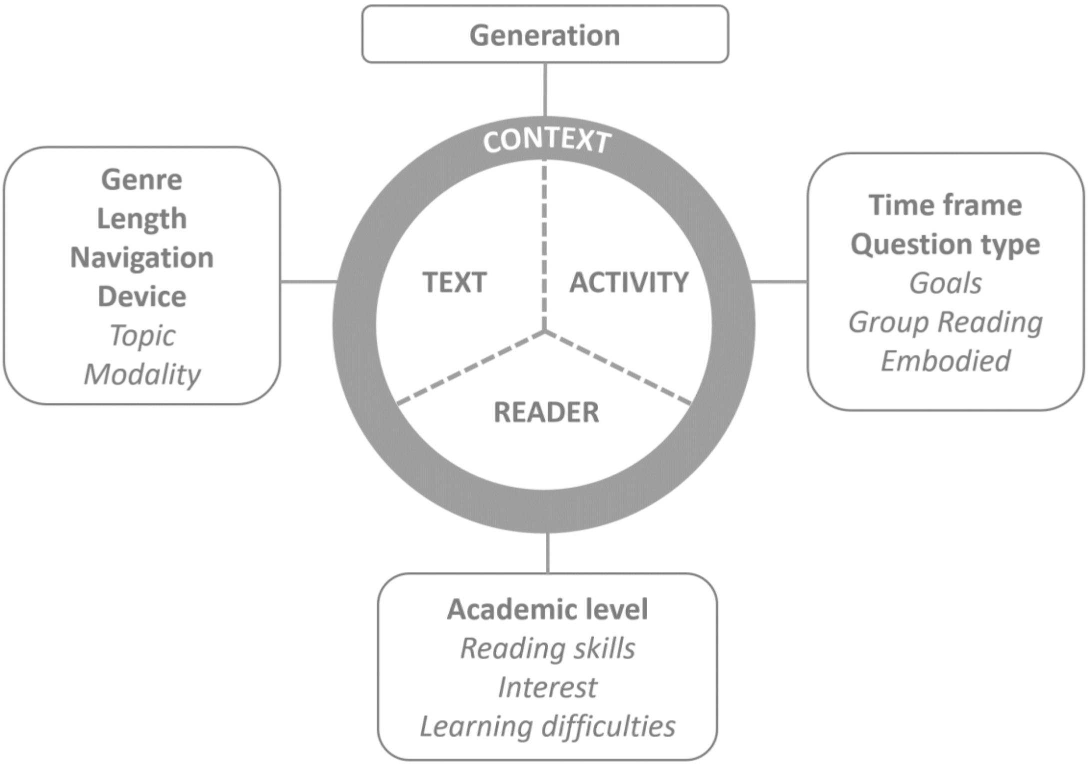
Potential moderators for the screen inferiority effect

Since the publication of the first systematic reviews and meta-analyses reporting the screen inferiority effect in 2017-2018 (Delgado et al., 2018; Kong et al., 2018; Singer & Alexander, 2017), several researchers in literacy studies have been involved in efforts to understand and explain the underlying mechanisms for this phenomenon. In our view, this has been a major shift in the literature, which previously focused, in most cases, on testing potential differences between reading media rather than testing potential explanations. One of the most influential theoretical accounts for the screen inferiority effect is the shallowing hypothesis (Annisette & Lafreniere, 2017), which originally aimed to describe how the current pattern of rapid and constant consumption of information through digital technology impacts users' cognition. The shallowing hypothesis assumes that typical interactions with digital devices involve quick episodes with short pieces of texts, which favors a mindset aiming to browse as much information as possible, without in-depth elaboration. Accordingly, as people engage in reading on a digital device, they may activate such a shallow mindset. This may cause them to be less efficient in allocating cognitive resources during reading, a situation that may be particularly visible in challenging contexts that demand readers to efficiently self-regulate their reading processing (Delgado & Salmerón, 2021). Previous research has identified different elements where the potential negative effects of such shallower processing may be particularly salient, thus causing greater comprehension problems when reading on screen.

In order to synthesize those factors into a coherent model we relied on the reading for understanding framework (Snow et al., 2002), which conceives text comprehension as the process of extracting and constructing knowledge. Comprehension occurs as the interaction of three main elements: reader (individual abilities, knowledge and attitudes), text (any form of print or digital text), and activity (purpose, processes, and consequences of reading), that take place within a particular socio-cultural context. For each of those four elements we discuss factors for which there is substantial evidence, as well as features for which we only identified emerging evidence. As we discuss next, most previous research exploring the effects of reading medium on comprehension have focused on the text and the activity levels, and to a lesser extent at the reader and socio-cultural context (see Figure 1).

Figure 1

Potential moderators for the screen inferiority effect on handheld devices, organized by reader, text, activity, and socio-cultural context



Note. Moderators with only emerging evidence are identified in italics.

Reader characteristics

The most researched reader factor within the literature about reading media and comprehension is academic level (Clinton, 2019; Delgado et al., 2018; Öztop & Nayci, 2021). Other relevant reader factors, for which we only identified emerging evidence, are topic interest (Clinton-Lisell, 2022; Delgado & Salmerón, 2021; Latini et al., 2020), reading comprehension skills (Salmerón et al., 2021; Støle et al., 2020), and learning difficulties (Ben-Yehudah & Brann, 2019; Stern & Shalev, 2013).

Through the lens of the shallowing hypothesis, higher frequency of social media interactions would result in a higher tendency to adopt a shallow mindset when reading on screens. Given that teenagers use social media and web-browsing more often than younger children (Rideout & Robb, 2019), we should expect that screen inferiority emerges in older students. The moderating effect of student academic level on the screen inferiority effect has been analyzed in previous meta-analyses (Clinton, 2019; Delgado et al., 2018; Öztop, & Nayci, 2021). Delgado et al. (2018) grouped the effect sizes in three broad categories: grades 1 to 6 (*n* = 8), grades 7 to 12 (*n* = 8), and undergraduates (*n* = 36). They found null effects, indicating that the screen inferiority effect on comprehension was similar across educational levels. A similar null pattern was obtained in the meta-analysis by Clinton (2019), who compared studies that sampled adult participants (*n* = 26) or children (*n* = 7). Nevertheless, as most studies analyzed in both meta-analyses sampled undergraduate students, these results must be taken with caution. Öztop and Nayci (2021), in a meta-analysis of studies that used texts written in Turkish, compared specific primary school grades: fourth (*n* = 6), fifth (*n* = 15), sixth (*n* = 3), and eighth (*n* = 2) grades. Although the students' grade was not a significant moderator, it is worth noting that the effect size for fourth grade students (*g* = -0.84) doubled that of fifth and sixth grades (*g*s = -0.40 and -0.41), contradicting the predictions derived from the shallowing hypothesis. Recently, Salmerón et al. (2023) found a similar pattern regarding the negative relation between the amount of time spent reading on digital devices in English classes and reading comprehension achievement in two representative samples of US fourth and eighth graders (unstandardized estimates = -4.03 vs -2.07). In sum, while the shallowing hypothesis predicts a higher screen inferiority effect for older students, who are more used to interacting with social media, existing evidence does not support such assumption. Nevertheless, as previous meta-analysis included mostly undergraduate samples (Clinton, 2019; Delgado et al., 2018), or only analyzed primary school students (Öztop & Nayci, 2021), the moderating effect of academic level on the relation between reading medium and comprehension is still an open issue.

A first reader factor with emerging evidence is topic interest (Clinton-Lisell, 2022; Delgado & Salmerón, 2021; Latini et al., 2020). Higher interest on a to be learnt topic could boost readers' attention towards the text, which could in turn overcome the expected shallow processing of digital texts. Contrary to this expectation, recent evidence with undergraduate samples have reported no significant interactions between topic interest and comprehension across reading media (Clinton-Lisell, 2022; Delgado & Salmerón, 2021; Latini et al., 2020). Alternatively, the expected shallow processing of digital texts could lower the chances that the text could capture readers' situational interest. Against this, Clinton-Lisell (2022) reported a non-significant interaction between situational interest (measured after the reading session) and comprehension across reading media.

A second reading factor with emerging evidence is students' reading comprehension skills (Salmerón et al., 2021; Støle et al., 2020). Readers with higher comprehension levels at the end of primary school are characterized by having better self-monitoring of comprehension, inference and integration skills, and knowledge and use of story structure (Oakhill & Cain, 2012). Against this background, Salmerón et al. (2021) proposed the comprehension levels hypothesis, which suggests that the higher order skills of good comprehenders could help them minimize the distractions elicited by tablets during reading tasks. Tests for this hypothesis are scarce, and evidence is mixed. Støle et al. (2020) evaluated a large sample of ten-year-old Norwegian students, who completed a reading test either in print or on a computer. They used the combined scores of those tests to group students by reading comprehension skill, with low, medium, and high performers. Overall, results replicated the screen inferiority effect, with lower scores for the test taken on a computer. However, comparisons by reading comprehension level revealed that the effect was substantial and similar across all three groups. In a follow up study, Salmerón et al. (2021) evaluated a sample of 10- to 13-year-old students, who read two expository texts either in print or on tablet, with or without time pressure. Participants were split into two groups of high and low comprehension skills, respectively, according to their scores on a different test assessing baseline reading comprehension. Results indicated that students with low reading comprehension skills comprehended better in print than on tablet when reading under time pressure. Those with higher skills were not affected by the reading medium.

Lastly, we identified emerging evidence on how cognitive or learning disabilities could moderate the screen inferiority effect. The reported meta-analyses (Clinton, 2019; Delgado et al., 2018; Öztop, & Nayci, 2021) did not include empirical studies that analyzed samples from special populations. Nevertheless, few studies have experimentally examined the reading medium effect on text comprehension in students with low attentional capacity: one in upper secondary students (Stern & Shalev, 2013) and another in undergraduate students (Ben-Yehudah & Brann, 2019). Contrary to the shallowing hypothesis, in Stern and Shalev's (2013) study, all participants performed better in the computer condition. Participants with low sustained attention levels (mostly students with diagnosed ADHD) performed better on screen than in print when the line-spacing in the texts was increased in the on-screen texts, whereas participants with high sustained attention levels performed even better with standard spacing on screen. In a more recent study, Ben-Yehudah and Brann (2019) found contrary results, supporting the screen inferiority effect. In this case, a sample of undergraduate students self-paced the reading of a longer text (1,200 words) and were not allowed to look back at it when answering comprehension questions. The results showed that typically developing students attained similar text comprehension in both media, whereas the students with ADHD comprehended poorer on screen than in print.

In sum, research remains largely inconclusive regarding the role of readers' characteristics on text comprehension across reading media. With respect to academic level, a major limitation of previous meta-analyses is that they mostly included original studies with undergraduate samples. By incorporating recent studies with younger populations, we will be able to shed further light on this issue. As for features with emerging evidence (interest, reading comprehension skills and learning disabilities), more research will be needed in the following years.

Text characteristics

Text characteristics have a significant impact on comprehension. Research looking at how textual characteristics could interact with reading medium has substantially focused on linguistic features (genre and length), as well as format (navigation through pages and type of handheld digital device). We identified two additional text features with emerging evidence: topic (Haddock et al., 2020) and multimodality (Latini et al., 2020).

Regarding text genre, meta-analytic evidence has revealed that narrations are easier to comprehend than expository texts (Clinton et al., 2020), as the latter tend to use more specialized vocabulary and complex rhetorical structures. Building on the assumptions of the shallowing hypothesis, Delgado et al. (2018) proposed that more challenging expository texts produce a higher screen inferiority effect than narratives, a claim that has been also supported by Clinton's (2019) meta-analysis. In addition, although in Öztop and Nayci's (2021) meta-study the moderator analyses did not reach statistical significance, the authors pointed towards the same direction as the other two referred meta-analyses. In the same line, the recent meta-analysis by Schwabe et al.

(2022) synthesized studies comparing comprehension of narrations presented either in print or on screen, yielding comparable results across both media. To our knowledge, no previous study has directly compared the effects of text genre and reading medium using exclusively handheld devices as the digital reading device.

The second textual factor that could moderate the reading medium effect is text length. According to the shallow hypothesis, common quick and scan patterns when reading digital media might induce the adoption of a discontinuous attentional style. Consequently, readers of digital texts may struggle to keep focus on a particular digital task for a prolonged period such as when facing lengthy texts. In their review, Singer & Alexander (2017) identified better comprehension scores in print (vs. digital reading) when the texts contained more than 500 words or the information was presented in more than one page, as was the case for 91.67% of the final 36 reviewed studies. Furthermore, Delgado et al. (2018) analyzed this factor as a potential moderator in their meta-analysis. Effect sizes were divided into those that used short (less than 1000 words, n = 22) or long texts (1000 words or more, n = 26) and found that text length did not moderate the reading medium effect.

Third, previous research has addressed the way in which readers navigate through text pages as an influential factor of text processing and comprehension. Text pages provide fixed cues for spatiotemporal orientation, but such cues are lost when digital documents use scrolling to navigate through the text (Haverkamp et al., 2023). Accordingly, scrolling may increase the negative effect of on-screen reading, as compared to those in which pages are fully presented on screen and navigated one by one. Delgado et al. (2018) analyzed the need for scrolling as a potential moderator of the mean effect size of the reading medium and found that the screen inferiority effect was larger when the studies used scrollable texts, but the difference was also non-significant. Regarding the issue of how navigation plays a role specifically with handheld devices, the design of tablets but especially that of e-readers are thought to resemble traditional printed books, while maintaining updated technological advantages. Nevertheless, the specific differences between swiping full pages (horizontally, as in printed books but also possible on digital devices) or scrolling (vertically, characteristic of digital devices) on reading comprehension have not been examined for handheld devices. Based on Delgado et al.'s (2018) results and the idea of print-texts resemblance, we could expect a diminishing screen inferiority effect for page swiping compared to scrolling while reading on handheld devices.

Fourth, research has investigated the type of handheld digital device. The shallowing hypothesis' assumption that digital devices promote quick interactions and frequent attention shifts makes sense for tablets, as they are designed to support a wide variety of applications and can distract users with frequent notifications or with the temptation to engage in off-task activities (Baron et al., 2017). However, such assumption is difficult to defend when it comes to e-readers, as they do not serve any other purpose than reading, as printed books do. A recent qualitative study (Bon & Burke, 2022) gathers readers' perceptions of e-reader screens as 'clear', 'easy to read', 'more natural', and 'less tiring', as compared to tablets, laptops, and smartphones. Accordingly, we may expect that the screen inferiority effect increases when reading on tablets, as opposed to e-readers. However, the few studies that have compared reading comprehension outcomes for the same texts displayed either on a tablet or on an e-reader (Baker, 2010; Kretzschmar et al, 2013; Norman, & Furnes, 2016, Exp. 1; Wang & Lin, 2016) reported similar comprehension outcomes between the two types of devices, a pattern of results which does not support the shallowing hypothesis when it comes to reading on handheld devices.

Within text factors, emerging research exists for topic (Haddock et al., 2020) and multimodality (Latini et al., 2020). Building on media-content matching effects in persuasion research, Haddock et al. (2020) proposed that text content (e.g., traditional vs. modern issues) could interact with the reading medium (traditional paper vs. modern digital devices) in shaping readers' comprehension. Text comprehension should be higher when content and medium matches, such as when a student reads a text about space ships on a tablet, or a text about Ancient Rome on paper. Haddock and colleagues tested these assumptions in two studies that yielded mixed evidence. The expected matching effect was found only in one of the four text-medium comparisons analyzed. Finally, Latini et al. (2020) tested if the screen inferiority effect was moderated by multimodality. Authors tested a sample of undergraduate students that read an expository document composed of text and graphics supplementing the text content. Integrated processing of texts and graphics was higher in print than on screen, as identified by eye-tracking measures. Nevertheless, integrated comprehension did not differ across reading media.

In sum, previous research has identified text genre as a relevant moderator of the screen inferiority effect, with stronger negative effects for expository than narrative texts. Evidence for length, navigation, and type of handheld digital device remains inconclusive, with prior meta-analytical evidence revealing non-significant trends. Including evidence from additional original studies in our meta-analyses could help to clarify their effects. Regarding text features with emerging evidence, as text topic naturally varies across studies, we could study for the first time the matching hypothesis (Haddock et al., 2020) by means of a meta-analysis. Other text factors, such as multimodality, should be further addressed in future research, as only a minority of studies exploring reading medium effects included texts and graphics.

Activity factors

Most research exploring activity factors in the context of reading medium effects have focused on reading time frame and question type (Clinton et al., 2019; Delgado et al., 2018). We identified four other activity factors with emerging evidence: goals, individual vs. group settings, and embodied interactions.

Regarding time frame, meta-analytic evidence shows that the screen inferiority effect increases when time to complete the reading assignment is limited (Delgado et al., 2018). On the contrary, in contexts where readers self-pace their reading, they may overcome the distractions associated with their on-screen reading mindset by investing more time rereading the text (Bowman et al., 2010). Supporting this claim, in a meta-analysis of studies comparing print and screen-based reading Clinton (2019) reported that reading times tend to be longer on digital texts, as compared to equivalent printed texts. However, recent empirical studies that have systematically varied the time frame while comparing readers' comprehension on tablets and in print resulted in null effects (Delgado & Salmerón, 2022; Salmerón et al., 2021). In Delgado and Salmerón's (2022) study, eye-movement data indicated that participants did not differ in terms of initial or subsequent rereading of text paragraphs as a function of timeframe. Accordingly, this new evidence suggests that tablets may provide a reading experience closer to printed texts, rather than that provided by computers, minimizing the potential harmful effect of time pressure.

The type of reading comprehension questions used has been analyzed extensively in regards to the medium effects, yielding null effects (Clinton, 2019; Delgado et al., 2018). Text comprehension models assume that readers represent the text both as a collection of ideas stated in the text, as well as the situation inferred from it (Kintsch, 1998). That latter representation is assumed to be more challenging than the former, as readers must go beyond what is explicitly stated in the text by connecting isolated ideas from the text together or with their existing knowledge. A shallower processing induced by digital texts would be particularly detrimental for comprehending texts at a more challenging level (i.e., situation model representation). Contrary to this assumption, previous meta-analytical evidence showed that the screen inferiority effect is not moderated by the type of questions (textbase or inferential) included in the original studies (Clinton, 2019; Delgado et al., 2018). Accordingly, we could expect that text comprehension measures would not moderate the screen inferiority effect in handheld devices.

Although it remains a major activity factor in the general literature of reading comprehension (Britt et al., 2018), it has only recently that goals attracted the attention of researchers exploring reading medium effects. A shallow processing in digital text may be eliminated if readers set high standards for the quality of their comprehension, as when they are requested to read to prepare for an exam (van den Broek et al., 1995). Recent attempts to explore if reading medium interact with goals (reading for pleasure versus to study for an exam [Latini et al., 2020], or reading to prepare a report to inform or to persuade [Valenzuela & Castillo, 2023]), yield no significant interactions. A second factor that has recently been identified as relevant for reading medium effects is group settings. In discussing the lack of comprehension differences between paper and tablet reading in their study, Latini and Bråten (2022) proposed that, as screen reading may demand for increased attention to perform similarly to print reading, a context that adds further distraction, such as group reading, may particularly potentiate the screen inferiority effect. Notably, meta-analytic evidence has proved a small-size negative effect of auditory distractions, such as the ones we may expect while reading in a shared room, on text comprehension (Vasilev et al., 2018). A third relevant activity factor with only limited evidence is embodied interactions with digital tools. According to an embodied perspective of cognition, digital devices differ in the extent to which their physicality and affordances support readers' comprehension (Mangen & Van der Weel, 2016; Mangen & Pirhonen, 2022). In this regard, a relevant feature of reading media is the extent to which readers can hold the device. Empirical research on reading medium preferences indicates that readers value handheld devices for their tactile nature and the ability to hold them, in contrast to reading on computers (Mangen & Pirhonen, 2022). Consequently, we could expect a reduced screen inferiority effect on handheld digital devices, particularly when readers are able to hold the device in their hands.

To conclude, meta-analytical evidence shows that time frame is a major activity moderator of the screen inferiority effect, with larger effects for tasks in which time is limited (Clinton, 2019; Delgado et al., 2018). Type of comprehension question, on the contrary, doesn't moderate the effects of reading medium. The effect of some of the activity factors with emerging evidence could be analyzed in our metaanalysis, as they point to characteristics that naturally vary across original studies (individual vs. group activity, embodied interactions). On the contrary, as a majority of previous works exploring reading medium effects are based on studies that used a goal of reading to prepare for a comprehension test (Latini et al., 2020), this factor cannot be measured by means of a meta-analysis, and should be addressed in future research.

Socio-cultural context

We identified generational cohorts as the only socio-cultural context factor that has been studied in the context of reading medium and comprehension. The great technological evolution experienced in recent decades has modified the way people interact with digital reading devices. Before the emergence of Web 2.0 in the mid-2000s (Stephens, 2007), reading on digital devices resembled more traditional print reading, at least in terms of text length and lack of social interactive features that could distract reading. For example, the “like button” on Facebook was introduced in 2010, six years after the platform was launched. The emergence of new forms of on-screen reading brought by Web 2.0, capitalized by social media and web forums, imposes a quick rhythm of reading that is frequently interrupted by the need to interact with texts, as readers are expected to comment on news or blog posts, or to rate or reply to other users’ posts on social media. Accordingly, detrimental effects of on-screen reading should be closely linked to the rise of Web 2.0. Initial evidence for this claim comes from the meta-analysis by Delgado et al. (2018), which regressed the effect size of the difference between print vs. on-screen reading comprehension in the year of publication of each study, from 2000 to 2017. Results showed a small significant positive effect, indicating that over the years the difference in favor of print increased, in contrast to what would have been expected if the ‘screen inferiority effect’ was due to readers’ lack of practice with digital tools.

Other relevant socio-cultural context factors, such as students’ SES and degree of technology inclusion in schools, could affect students’ familiarity with digital texts (OECD, 2015), substantially contributing to differences in reading medium comprehension across media. In a previous meta-analysis, Delgado et al. (2018) signaled that information regarding those factors was seldom reported in original studies. Accordingly, future research should specifically address the role of those socio-cultural context factors in analyzing the effects of reading medium on comprehension.

The present study

Our review relied on the reading for understanding framework (Snow et al., 2002) to identify evidence for potential moderators for the screen inferiority effect, including reader, text, activity, and socio-cultural characteristics. We aimed to empirically test this scheme in a meta-analysis of studies that specifically compared comprehension on handheld devices to on print text.

Specifically, we had two objectives. First, we aimed to examine whether reading on handheld digital devices would result in lower reading comprehension outcomes than reading on paper (screen inferiority effect). Second, we wanted to identify moderating factors of the effect of the reading medium on reading comprehension outcomes. Specifically, based on the literature reviewed, we expected that the following moderator factors would affect the screen inferiority in the following directions:

Reader factors

1.1) Educational stage: null effect.

Text factors

2.1) Expository texts: larger effect than narrative ones.

2.2) Long texts: larger effect than short texts.

2.3) Topic novelty and reading medium are not matched: larger effect.

2.4) Scrolling through the text pages: larger effect than horizontal swiping.

2.5) Tablets: larger effect than using e-readers.

Activity factors

3.1) Imposed time limits: larger effect than self-paced reading time.

3.2) Comprehension questions: null effect.

3.3) Group setting: larger effect than individual settings.

3.4) Readers could not hold the tablet/e-reader: larger effect than when they were allowed to.

Socio-cultural context factors

4.1) Participants from recent years: larger effect than on those from previous years.

Method

Selection criteria

To filter the reports retrieved in all the search phases, we selected those studies that met the following inclusion criteria:

1. The study used an experimental design in which text comprehension is compared between reading in print and on digital devices. Accordingly, studies that used several digital devices (e.g., audiobooks and e-readers) but no print texts were not included (e.g., Rogowsky et al., 2015).
2. The digital reading devices are either a tablet or an e-reader. For instance, studies that only included computers (e.g., Lenhard et al., 2017) or mobile phones (e.g., Schneps et al., 2013) were discarded.
3. Participants must be able to read autonomously. For example, studies that employed shared-reading experiences or parent-reading were not included (e.g., Rvachew et al., 2017).
4. The study makes an empirical contribution that includes the results of the comparison (i.e., the paper is not a review or an opinion). Other meta-analyses and reviews were consulted but not included in our meta-analysis (e.g., Delgado et al., 2018).
5. The study was published or presented from 2010 onwards. Formal publication was not required.
6. The report is written in English, Spanish, German, French, or Greek.
7. The report includes sufficient statistical information to calculate the effect size or was provided by the authors following a personal request.

Search procedure

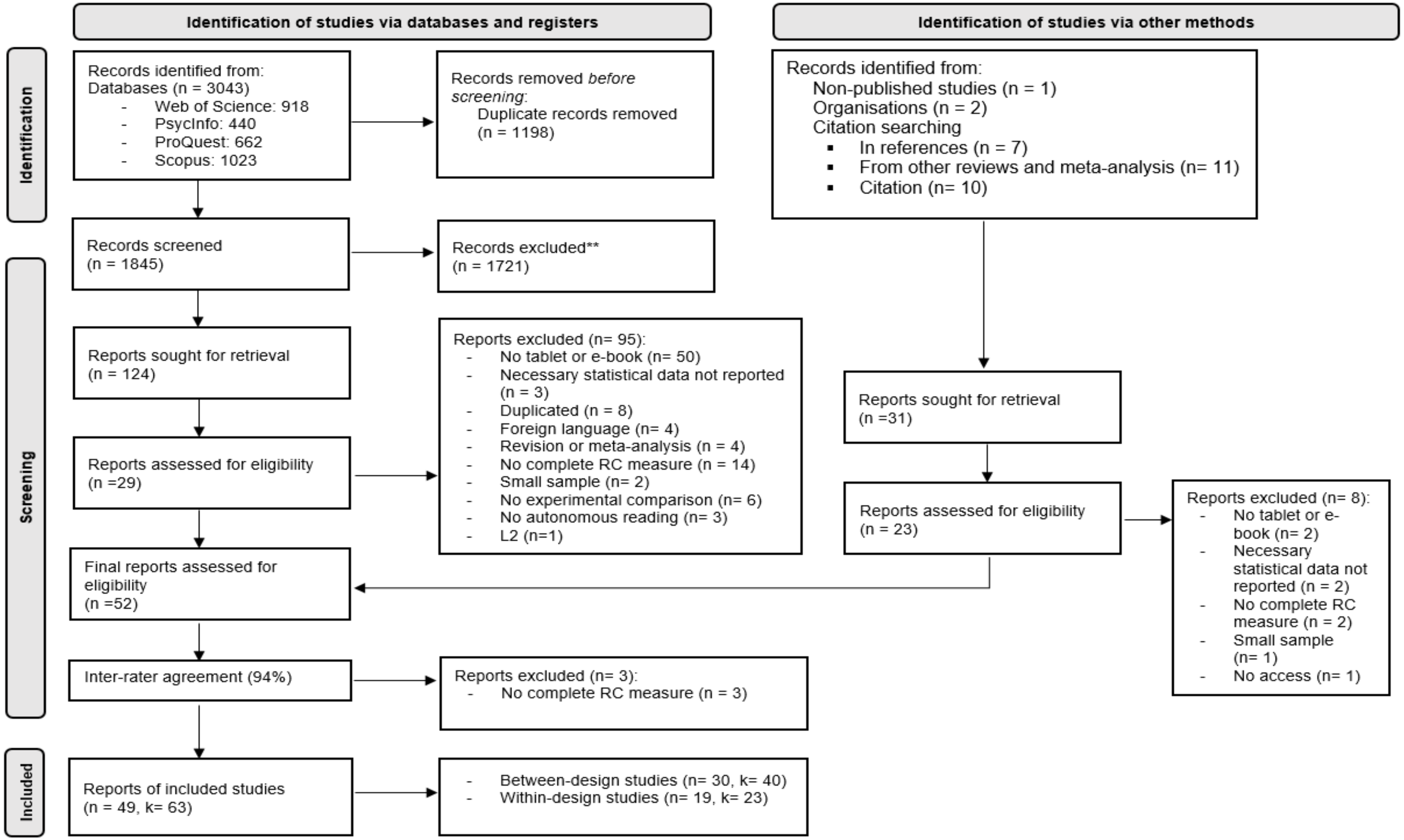
We initially conducted a systematic search through the academic databases Web of Science, Scopus, PsycInfo, and ProQuest. These databases also contain gray literature, such as dissertations. We entered the following search function: (“tablet reading” OR "iPad reading" OR “computer reading” OR “online reading” OR “screen reading” OR “digital reading” OR “print reading” OR “paper versus screen” OR “differential test” OR “computer based testing” OR “computerized testing” OR “computer assisted testing” OR “electronic book” OR “electronic text” OR “reading medium” OR “mode effect”) AND (“comprehension” OR “retention” OR “test performance” OR “learning”). The publication years were restricted from 2010 to 2021. The database search ended in September 2021. These terms were searched as the title, abstract, or keywords.

As recommended by Card (2012), we complemented the search with additional strategies. Therefore, in the second phase, we examined references included in eleven previous reviews and meta-analyses. From those studies, 16 were already identified in the initial database search and 10 were incorporated in a later phase. In total, our meta-analysis includes 24 studies that have not been analyzed in previous reviews (see Table 1 and Table 2 in the online supplemental material for more details). To incorporate gray literature, we contacted authors of original studies through email to request additional information and specifically asked them for any unpublished studies. Additionally, we reached out to other experts in this field, making a similar request. Additionally, a forward search was performed using Web of Science, ProQuest, and Google Scholar to find studies that cited the selected works. This search was performed in May 2022. Lastly, references from the selected studies were also retrieved and examined. The search and filter process of all the reports collected through the different search-strategies ended in June 2022.

The search described above yielded 1,845 records. Following the criteria exposed, two of the authors selected the eligible studies, yielding a set of 52 studies. After this first selection, one of the other authors acted as a third coder and assessed the selected studies. The coders agreed on 94% of the initial selection. Disagreements were discussed and 3 studies were removed. The final sample consisted of 49 studies (30 between-participants design, 19 within-participants design). As some studies included information from independent subsamples, the effect sizes analyzed were 63 in total (40 between-participants design, 23 within-participants design). A flow diagram based on PRISMA guidelines (Page et al., 2021) outlines this process in Figure 2.

Figure 2

PRISMA flow diagram of the search and selection process



Coding the studies

Based on the four main factors identified through the narrative review (reader, activity, text, and socio-cultural context factors), descriptive information and relevant substantive variables were gathered across the studies with the intention of using moderator analysis to explain the heterogeneity among the reported effect sizes. When the necessary information for a particular variable could not be retrieved from the report, it was coded as “Not reported” (NR).

The first factor, reader characteristics, was represented by the corresponding educational stage of the participants at the moment the study was conducted. This variable was coded based on the US educational system to homogenize the categorization (0: No student sample, 1: Pre-school, 2: Primary school, 3: Secondary school, 4: Undergraduate students, 5: Mixed sample, i.e., when the sample is large, and it ranges over several stages).

The second factor, textual characteristics, comprised five variables. First, we coded text genre (categorized as 0: Narrative, 1: Expository or 2: Mixed). Second, text length was measured as the number of words per text (Total word count). Because some studies used more than one text, we additionally coded Average word count. Third, we coded how readers navigated through the texts (0: Paging, swiping pages horizontally, 1: Scroll, swiping vertically). Fourth, referring to the digital devices employed, we differentiated between 0: Tablets (portable devices that could be used for several purposes, such as iPad, tablet, tablet computer, Chromebook tablet…) and 1: E-readers (portable devices that could only be used to read, such as Kindle, e-reader, digital book, electronic textbook, nook book…). In cases in which the study provided information about any other device, we only selected the information about the tablets and/or e-readers (e.g., Baker, 2010). Lastly, one of our variables of interest regarding text characteristics was text topic, in order to test the hypothesis of topic-media congruence (Haddock et al., 2020). Unfortunately, it was too complex to categorize this variable from the perspective of this hypothesis (i.e., modern vs. traditional), as most of the studies used non-fiction expository texts (e.g., scientific topics) and the topics of narrative texts were not clearly described in most of the studies.

For the third factor, referring to the activity, we focused on four variables. Firstly, the allotted time for reading (0: Self-paced, no time limit, 1: Time-limit). Secondly, regarding the reading comprehension instrument employed, we classified the studies according to the reading comprehension dimensions assessed (0: Literal, 1: Inferential, 2: Mixed). Thirdly, we specified the session modality – if the experimental task took place in 0: Group sessions, or in 1: Individual sessions. Lastly, regarding embodied interactions, we coded whether readers were allowed to freely manipulate the handheld device (1: Possibility to hold) or not (0: No possibility to hold).

Finally, the fourth factor, about socio-cultural context, was coded as the year in which the study was published or presented.

Furthermore, regarding information that is not framed within the four main factors, we gathered some additional sociodemographic information such as the average age of participants and the percentage of females and males for descriptive purposes. Besides, to test the methodological quality of the studies, we included the following methodological variables: total sample size; sampling procedure (0: Non-probabilistic-convenience sampling, 1: Probabilistic-random sample selection); and group allocation to the medium conditions (0: Random, 1: Quasi-random, 2: Non-random but matched or controlled, 3: Non-random and not controlled). Related to the reading comprehension instrument, we classified whether it was part of a 0: Standardized assessment or 1: Non-standardized test (i.e., created by the researchers). We also categorized the studies depending on the reading comprehension instrument reliability (0: Questionable reliability, when reliability indexes, as Cronbach alpha or Omega, were inferior to .70 or were not reported, 1: Acceptable reliability, when reliability indexes were above .70 or the test was used in large scale assessments. Lastly, as an extrinsic variable to control for possible publication bias we coded the corresponding publication status (0: Published in peer-review journal, or 1: Not-published).

Reliability in Coding

The coding of moderators was performed by two of the authors. In order to assess the reliability of the codification described above, an independent research assistant additionally coded all the moderator variables of 42 out of the 49 selected articles. The small percentage left out (k = 7) served as training. We next describe the inter-rater reliability separately for between- and within-participants design studies.

Inter-rater reliability was adequate. For categorical variables (n = 13) Cohen's Kappa coefficient was used and the average reliability for between-participants studies was .87 ($minimum_{Sampling\ procedure}$ = .48). Because the Kappa coefficient takes into account the frequency of category appearance, even though there may be high agreement rates, in these cases the coefficient tends to be low (Warrens, 2010). Therefore, we also report the mean percentage of inter-rater agreement = 95.60% (minimum = 86.11%, maximum = 100%). For the within-participants studies the Cohen's Kappa coefficient average was .89 ($minimum_{Standardized\ test}$ = .74) and the percentage of inter-rater agreement = 95.00% (minimum = 85%, maximum = 100%).

For continuous variables (n = 7) the intra-class correlation (95% CI) mean was 1, meaning perfect inter-rater coding for between-participants studies. Regarding within-participants studies, the intra-class correlation average coefficient was equal to .99 ($minimum_{Total\ sample\ size}$ = .98). All disagreements were discussed and, if needed, the other authors were consulted. For transparency and objectivity, a coding manual was developed and is available upon request. A descriptive overview of the studies' characteristics is given in the Results section and can be found in detail in the Supplemental Tables 3 and 4. Next, the variables coded across the studies are described.

Statistical analysis

Two meta-analyses were performed in two datasets, respectively: one including the studies that used a between-participants design, and the other including those with within-participants design. Because some threats to internal validity are different in each design and the combination can confuse the interpretation of the outcome, we find it important to conduct two separate meta-analyses according to the study design (Borenstein et al., 2009).

Effect size index

For each comparison included in the datasets, the effect size was calculated using means, standard deviations, and sample sizes (Borenstein et al., 2009) as reported in the scientific articles or provided by the authors upon request. When the studies used a between-participants design, the standardized mean difference Hedges' *g* was used as the effect size index. This index was defined as the difference between the on-screen reading (treatment) and in-print reading (control) groups' means on the post-test, divided by a pooled within-group standard deviation (Cohen, 1988). A positive Hedges' *g* indicates better comprehension outcomes for on-screen reading, and vice versa. In addition, to estimate unbiased effect sizes, the correction factor for small sample sizes proposed by Hedges and Olkin (1985) was used.

To estimate the effect size from the studies that used a within-participants design (i.e., all participants read both in print and digitally), we used the standardized mean change index, d_c. This index is calculated by subtracting the mean of the treatment group from the mean of the control group, whose result is divided by the standard deviation of change scores. Also, the correction factor for small sample sizes was applied (Botella & Sánchez-Meca, 2015). In this case, to keep the interpretation of the direction of the mean effect size constant across both datasets (i.e., positive values indicating better reading outcomes for the condition and vice versa), we used the on-screen condition as the control group. None of the primary studies reported the correlation coefficients, and hence, all values were imputed for a correlation value of .104 based on the within-participants estimate in a previous meta-analysis by Delgado et al. (2018).

A three-level random-effect meta-analysis was applied to address the fact that some effect sizes (*k*) were nested within studies (*n*). In this model, the first level represents a within-effect size model, the second level shows variation between the effect sizes within the same study, and the third one represents variation across studies (Van den Noortgate et al., 2013). Although the participant level only contributed to one effect size (the sampling errors are conditionally independent), the non-independence could be introduced to the nested structure of the effect size. We calculated cluster-robust standard errors, significance statistical test, and confidence intervals for our estimates by using the CR2 method (Tipton & Pustejovsky, 2015). The method used to estimate the model parameters was restricted maximum-likelihood (REML). The inverse variance method was used to weigh each effect size. An average effect size and a 95% CI was finally calculated. The main strategy to obtain the average effect size was to use the means, standard deviations and sample size of each group. As a second strategy, when needed, we employed the raw difference in means, standard error of the means difference and sample size.

Lastly, as previously mentioned, some studies reported multiple conditions and comparisons (e.g., between different digital devices or different reading materials). In these cases, found in the between-participants dataset, the following strategies were applied: a) when the study consisted of multiple independent between-participants conditions, the effect size for each comparison was estimated (this is indicated in figures below by adding letters in alphabetical order after the publication year in the study reference); b) when there were multiple-treatment groups but they were dependent subgroups, the data from each sub-group was merged into one effect size (e.g., if the study provided data on tablets and e-readers from the same participants); and c) if two on-screen reading groups were compared with the same control group, the sample size for the control group was divided by two to minimize dependence (Higgins & Green, 2011).

Heterogeneity and sensitivity indexes

Cochran's Q statistic was employed to evaluate the heterogeneity among the effect sizes included in the datasets. The multilevel I^2 statistic was also considered to estimate the proportion of the variance in observed outcomes that reflects variation in true effect sizes rather than sampling error, with $I^2_{\text{Level 2}}$ and $I^2_{\text{Level 3}}$ representing within- and between-study heterogeneity, respectively. Moreover, a 95% prediction interval around the main effect size was calculated to know how the true effect varied across populations (Borenstein, 2019; Cheung, 2014). Furthermore, outlier cases were assessed with standardized deleted residuals larger than 2.5 (Viechtbauer, 2021; Viechtbauer & Cheung, 2010).

To explain the heterogeneity among the selected studies, analyses of moderating variables were performed by using multilevel meta-regression models with robust variance estimation methods for both continuous and categorical moderators.

Publication bias was assessed by using different approaches. Firstly, a funnel plot was examined for asymmetry. In addition, the Egger MLMA test was applied, incorporating multilevel meta-analysis with a modified measure of effect size precision (Pustejovsky & Rodgers, 2019), and testing significance using cluster-robust variance estimation methods. Lastly, another multilevel meta-regression model was applied to evaluate whether publication status moderated the effect size, that is, to compare the mean estimated effect sizes for samples obtained from published and unpublished research. Critically, all approaches were able to appropriately handle dependent effect sizes, as ignoring dependency can inflate the type I error rate (Rodgers & Pustejovsky, 2020).

Transparency and Openness

The two datasets used in this study are openly available at https://osf.io/sm3ba/?view_only=f9bc60a67b3041eeb1ef5b043a645d0b . This study design and its analyses were not pre-registered. All the statistical analyses above were performed using R 4.2.1 version with Metafor (Viechtbauer, 2010) and clubSandwich (Pustejovsky, 2020) packages.

Results

Descriptive characteristics of the studies

In the final sample ($n = 49$), 30 studies used a between-participants design. Of these 30 studies, nine had multiple independent comparisons, so that the final dataset consisted of 40 effect sizes with 161,469 participants that were initially included in the meta-analysis. In addition, 19 studies used a within-participants design, providing 23 independent comparisons (there were four studies that provided information for more than one independent sub-sample) with 1,379 participants.

Next, we will detail the variables that represent the four possible moderators for the screen-inferiority effect both for between and within-participants studies. Please note that all the information regarding these factors was not available in some cases. Thus, when describing the studies, the total number of effect sizes is not equal to the size of the datasets for some moderators. The characteristics of all the studies are described in detail in Tables 3 (between-participants) and 4 (within-participants) in the supplemental online materials.

Description of between-participants studies

First, regarding readers' characteristics we observed that over half of the effect sizes were from comparisons conducted on undergraduate students (k = 21). Following, there were thirteen cases in which primary (k = 6) and secondary school students (k = 7) participated; while for three effect sizes participants were categorized as non-students (e.g., adults not enrolled in academics).

Second, when looking at text factors we determined that the most frequent text genre was expository (k = 21), followed by narrative (k = 10), while a small proportion used both types of genres (k = 5). Next, the average word length was 1,549 words per text (range: 325-10,800). The shortest texts were typical for those studies in which primary and high-school students participated (e.g., Nishizaki, 2015a; Salmerón et al., 2021) while the longest texts were characteristic of undergraduate samples (e.g., Mangen et al., 2019) or non-students, such as senior participants (e.g., Hou et al., 2017). The most common way to navigate through the text pages on digital devices was horizontally (i.e., swiping; k = 19), closer to print reading, while only five reported vertical navigation (i.e., scrolling). With regard to the handheld digital device, most of the studies employed tablets as the digital reading device (k = 28), while eight used e-readers.

Third, concerning the activity factors we observe that in 23 effect sizes of the included studies, participants were allowed to read freely, contrary to the participants of the other 11 groups, who performed under time pressure conditions. Most of the studies assessed text comprehension using tests that mixed both textual and inferential questions (k = 20), while few used exclusively inferential (k = 2) or textual questions (k = 4). Regarding the reading session, almost half of the study settings were individual assessment sessions (k = 18), while in 15 of the studies the sessions were performed in groups (either in lab settings or regular classrooms). Lastly, only two studies indicated that participants were not allowed to hold the reading device, while more than half of the studies indicated free manipulation for participants (k = 21). The rest of the studies did not report this information. Due to the lack of variability in this variable, it was not possible to consider it for the moderator analysis.

Regarding the socio-cultural context, we found that the selected studies were published or presented between 2012 and 2022, but 20 effect sizes correspond to the last four years (2018 to 2022).

Regarding the methodological quality variables, most of the studies included convenience samples (k = 38). Only 2 of the effect sizes came from data collected following a probabilistic procedure (Jewsbury et al., 2020a, 2020b). Randomization was the most frequent strategy for group allocation (k = 36). Due to the lack of variability of these two variables, they were not considered in the moderator analysis. As for the reading comprehension assessment instrument, only 7 studies employed standardized tests. Around half of the studies used instruments with at least acceptable levels of internal consistency (k = 19). Finally, most of the studies (k = 31) used closed tests (different types of multiple-choice questions or true/false tests), and only a few used open tasks (short open-ended, essays) (k = 4). For additional sociodemographic information and details about the experimental materials see Table 3 in the supplemental online material.

Description of within-participants studies

Looking at readers' characteristics of the within-participants' dataset, most of the participants were undergraduate students (k = 14), followed by primary (k = 3), and secondary school (k = 3) students. Only in one study participants were categorized as non-students (e.g., adults not enrolled in academics), and two studies integrated students from several educational stages.

For the text factors, most of the within-participants studies used expository texts for the reading task (k = 13), while only few employed narrative texts (k = 3), or both types (k = 3). The average length of the texts was 867.96 words (range: 222 – 2,804). Regarding navigation, in this dataset, only in one study participants navigate through the texts vertically (i.e., scrolling), while 10 of them had to swipe horizontally (i.e., paging). Therefore, we could not use this variable as a moderator in the analyses. It should be noted that in four cases we could not report the specific type of handheld reading digital device. This was the case for studies in which a combination of both types of devices was required to calculate an independent effect size (Kretzschmar et al., 2013a, 2013b; Wang & Lin, 2016), or the information provided was not enough to differentiate between tablet or e-reader device (Liang & Huang, 2014). Nevertheless, for the rest of the studies, the most used reading digital device was the tablet (k = 13), compared to e-readers (k = 6).

Concerning the activity factors, eight studies allowed participants to read without a time limit, while another eight required participants to read within a specified time frame. In seven studies, information about time restrictions was not reported. Most of the studies assessed text comprehension by means of questionnaires that assessed both textual and inferential comprehension (k = 16) while only in few cases it was employed either textual (k = 2) or inferential questions (k = 2). Regarding the reading session, six studies (k = 7) reported having group sessions while participants in 7 studies (k = 10) read in individual settings. Regarding whether the participants were allowed to hold the reading device or not, out of 16 effect sizes, eight reported freedom to do so, while in the other eight participants were not allowed to.

Lastly, for the socio-cultural context factor, the reported studies for the within-participants' dataset were published or presented between 2010 and 2022, being the average year of publication 2017.

Regarding the methodological quality variables, all of the studies included convenience samples. Only 3 studies (k = 4) employed standardized reading comprehension tests, and less than half used a test with at least acceptable levels of internal consistency (k = 9). Finally, most of the studies (k = 15) used closed tests and a few open tasks (k = 5).

For additional sociodemographic information and details about the experimental materials see Table 4 in the supplemental online material.

The mean effect size, heterogeneity, and sensitivity analyses

We first examined possible outlier cases and tested normality assumption before analyzing the average effect sizes. Regarding the between-participants dataset, two effect sizes were identified as possible outliers based on standardized residuals (values > 2.5): Nishizaki (2015a) and Hsiao and Chen (2015), with standardized deleted residuals of 3.18 and 3.49, respectively. Once removed, the final sample for between-participants studies included 38 effect sizes (from 29 studies).

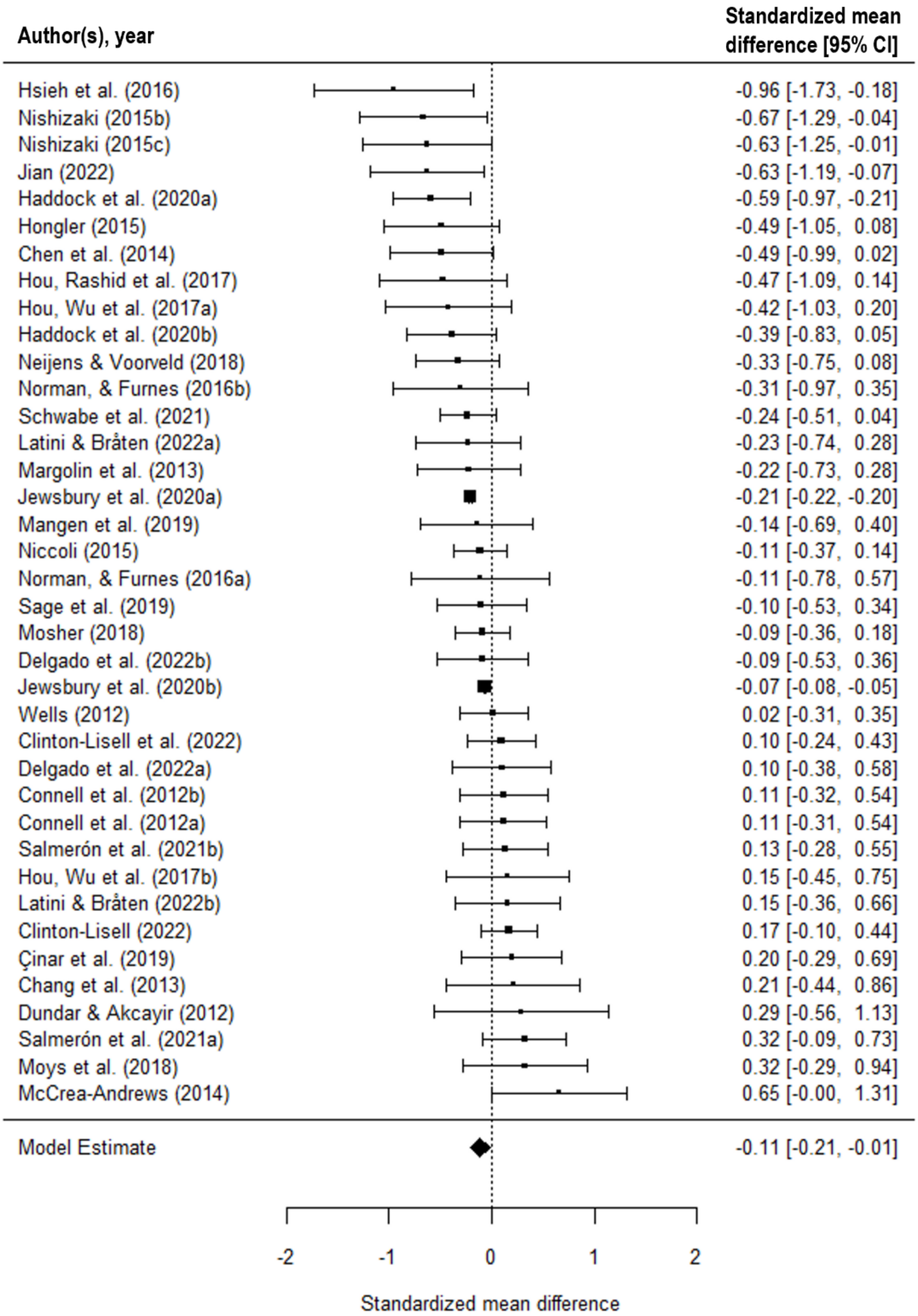
In the within-participants dataset, two studies showed standardized residuals above 2.5: Herman (2017) and Jeong and Gweon (2021b), with standardized deleted residuals of 2.88 and 3.20, respectively. They were removed from the within-participants dataset, which finally included 21 effect sizes (from 18 studies).

Media effect in between-participants designs

The meta-analytic effect was significant (g = -0.113, 95% CI [0.215, -0.012]; p = .030; k = 38, n = 29), indicating slightly poorer comprehension when reading on handheld digital devices than on paper. In interpreting this result, it should be considered that only a few of the primary studies resulted in significant differences. An overview of the effect sizes can be seen in Figure 3, which provides a graphical representation of the estimated result of each reading media comparison.

Figure 3

Forest plot of reading media effect sizes on reading comprehension from studies using between-participants designs



Note. The letters after the publication year differentiate several sub-samples from the same study. The diamond represents the average effect size. Please note that positive values favor reading comprehension on digital devices and negative values favor reading in print.

The heterogeneity of the effect sizes reached statistical significance, $Q(37) = 256.27$, $p < .001$. The I^2 statistic was 94.39, that is, 94.39% of observed variance reflected variance in true effects rather than in sampling error. The majority of true variance comes from between-study variance (73.54%), versus within-study variance (20.85%). Furthermore, the prediction interval was 95% PI [-0.486 to 0.259]. This range is usually wider than the confidence interval, particularly when there is great heterogeneity in the data (Al Amer & Lin, 2021), as is the case in our meta-analysis. We would expect that the true effect size of all comparable populations would fall in this range in 95% of all populations. In sum, this wide range of predicted effects and the significant heterogeneity above justified further analyses to examine potential moderating factors of the average effect sizes. The results are reported in a later section.

Sensitivity and bias analysis for between-participants studies

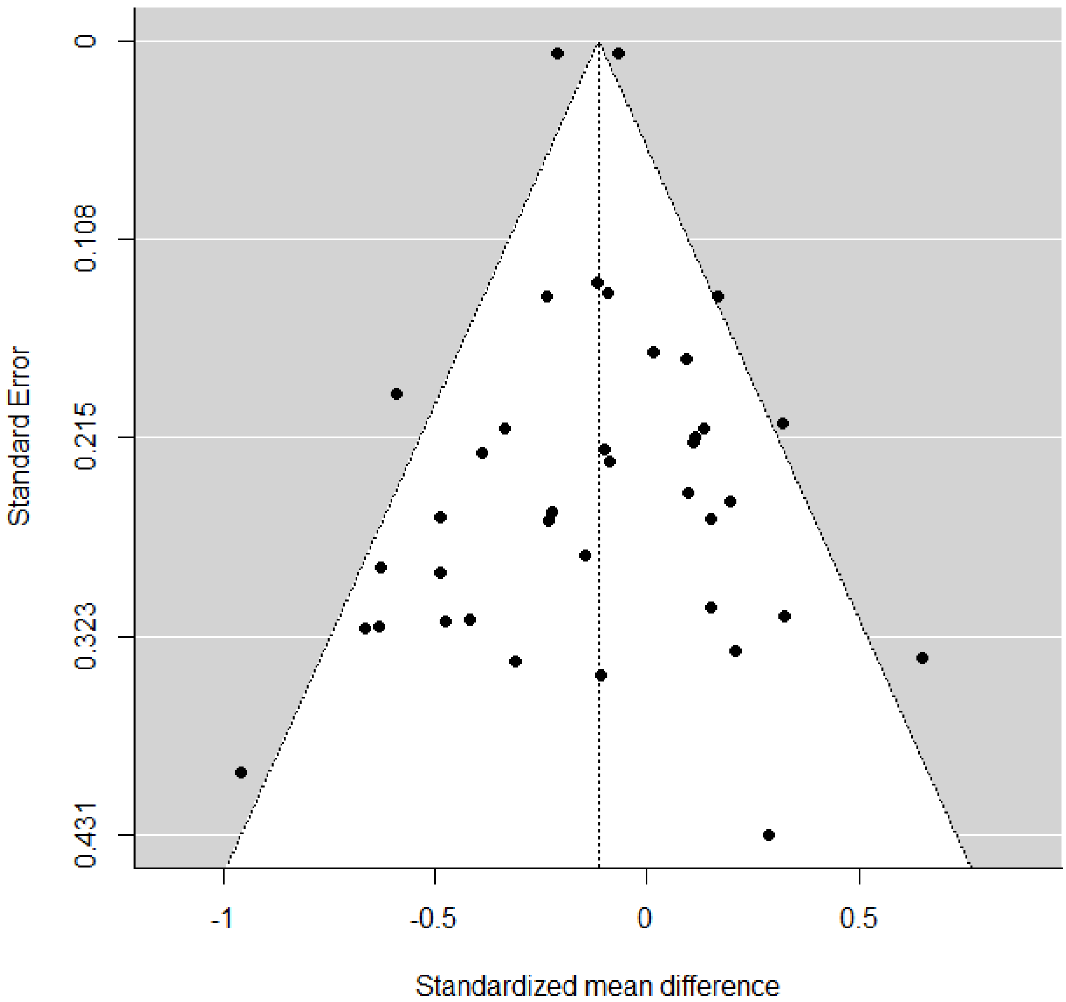
For the sensitivity analysis, we calculated the average effect size for the studies including the outliers (excluded from the results presented above). When comparing the calculations with the outliers we observed that the average value was slightly smaller and not significant (average effect size including outliers: $g = -0.069$, 95% CI [-.177, .038]; $p = .191$; $k = 40$, $n = 30$).

Secondly, we performed multilevel meta-regression analysis for moderator variables including the outliers. The significance of the moderators remained except for the percentage of male participants, which became significant ($b_{with\ outliers} = .012$; 95% CI [-.812, -.140]; $p = .012$).

In order to assess whether publication bias might be a threat against the meta-analytic results we applied three techniques. With respect to the funnel plot approach, a visual assessment revealed an apparent asymmetry (Figure 4). In addition, the MLMA Egger test for asymmetry on the intercept did not reach a statistically significant result (*b* = -0.364; 95% CI [-1.46, 0.73]; *p* =. 501). Likewise, multilevel meta-regression analyses with robust variance estimation revealed that the mean effect sizes from published versus unpublished studies were not statistically different (*p* = .678). These results suggested that there was no bias due to selective reporting or over-influence of studies with small samples.

Figure 4

Funnel plot of reading media effect sizes on reading comprehension from studies using between-participants designs



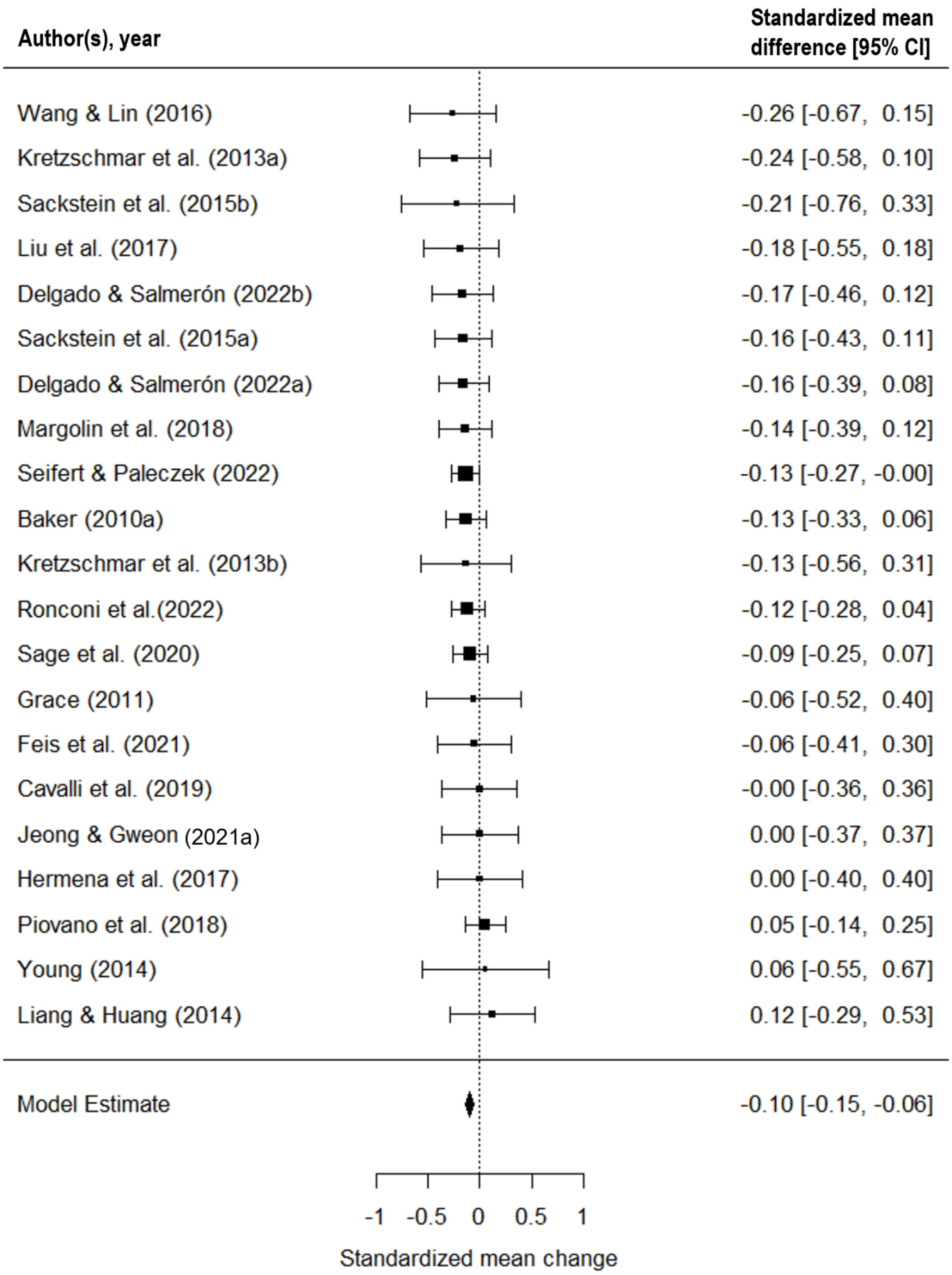
Media effect in within-participants designs

The meta-analytic effect was significant (d_c= -0.103, 95% CI [-0.145, -0.062], $p < .001$; $k = 21$, $n = 18$), indicating slightly poorer comprehension when reading on handheld digital devices versus paper. Of note is that none of the primary studies resulted in significant differences. An overview of the effect sizes can be seen in Figure 5.

Regarding the variability of the effect sizes, the heterogeneity *Q* statistic did not reach statistical significance, $Q(20) = 7.309$, $p = .995$. Furthermore, the prediction interval was small: 95% PI [-0.145, -0.062]. Additionally, the I^2 was equal to 0, indicating low heterogeneity (i.e., most of the variation in effect estimates is due to chance). Therefore, it is reasonable to assume that the effect sizes are estimating the same underlying effect. Due to the low heterogeneity found in the results we did not perform moderator analysis.

Figure 5

Forest plot of reading media effect sizes on reading comprehension from studies using within-participants designs



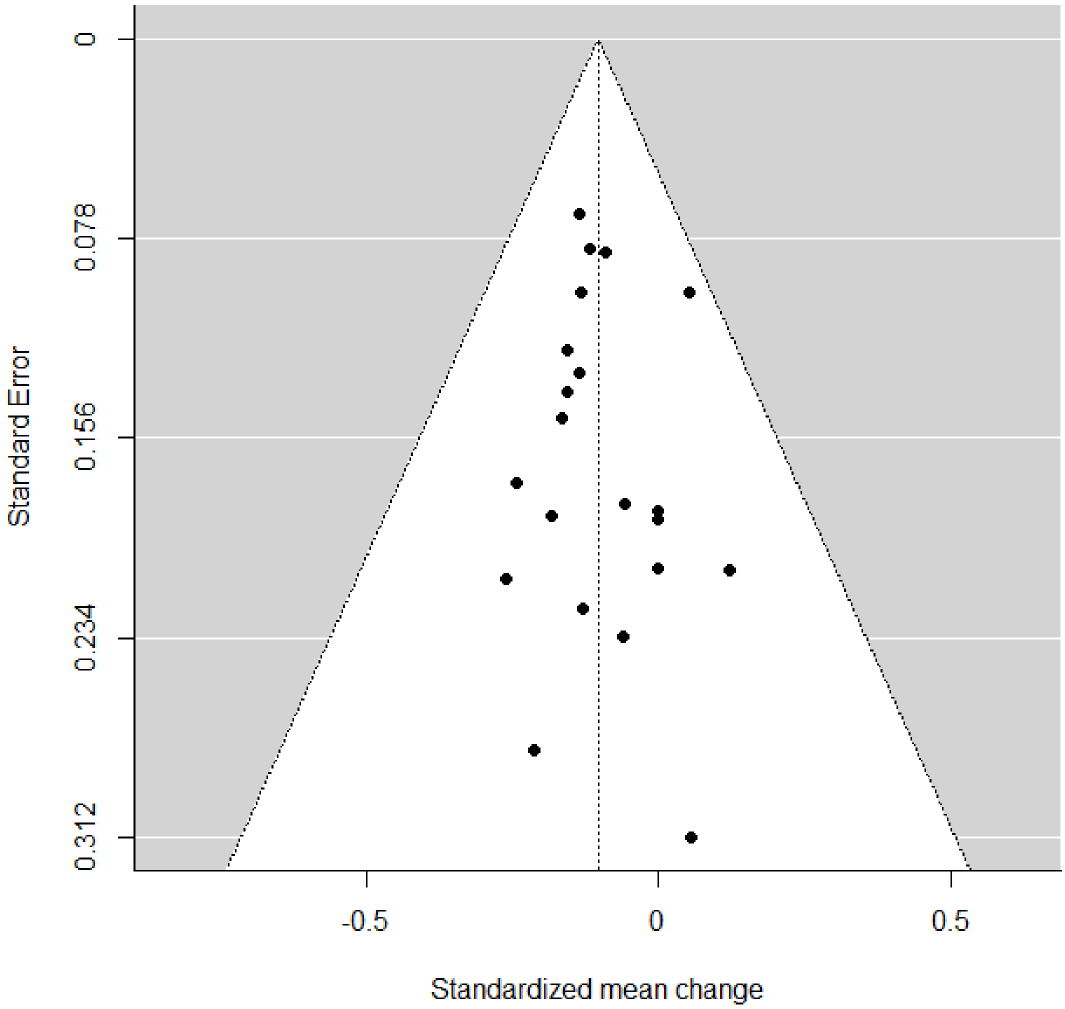
Note. The letters after the publication year differentiate several sub-samples from the same study. The diamond represents the average effect size. Please note that positive values favor reading comprehension on digital devices and negative values favor reading in print.

Sensitivity and bias analysis for within-participants studies

To compare the average effect size when the outliers were included we calculated the average effect size (d_c = -0.093, 95% CI [-.153, -0.34]; p = .006; k = 19, n = 23), which remained significant and exhibited similar values as the effect size presented above. As noted before, we did not carry out sensitivity analysis regarding moderator variables because of the lack of heterogeneity among the studies. Lastly, we performed bias analysis regarding publication status. The funnel plot does not appear to show an asymmetric distribution (Figure 6). Besides, the Egger MLMA test of the intercept did not reach a statistically significant result (b = 0.081; 95% CI [-0.042, 0.20]; p = .181). Even though we did not carry out moderator analysis, we did perform meta-regression analysis regarding publication status, which did not show significant differences between published and unpublished studies (p = .336). In sum, these analyses suggest that there was no publication bias.

Figure 6

Funnel plots of reading media effect sizes on reading comprehension from studies using within-participants designs



Moderator variables of the mean effect size in between-participants studies

Tables 1 and 2 show the results of examining the influence of qualitative and quantitative moderator variables, respectively, on the mean effect size. Firstly, regarding reader characteristics, educational stage was a significant moderator (p = .026). Due to the small quantity of effect sizes for non-students and the presence of primary studies that reported samples with mixed educational stages, we merged primary and secondary school students and focused on the differences between school students (altogether) and undergraduate students. The results showed that the mean effect size for undergraduates was significantly different from that of school students: whereas the mean effect size was higher for undergraduate students (*g* = -0.227) than for school students (*g* = 0.057), indicating a larger screen inferiority effect for undergraduates. Secondly, contrary to our expectations, none of the text characteristics analyzed (genre, length, navigation, handheld device) evidenced significant differences. Thirdly, regarding activity factors, only reading session modality was a significant moderator of the screen inferiority effect (*p* = .036). Results indicated that print reading favored text comprehension when the session was performed individually (*g* = -0.230) but not when it was performed in groups (*g* = 0.004). The other activity factors (time frame, comprehension questions, and embodied interactions) were not significant. Lastly, regarding socio-cultural context, year of publication (as indicator of generation) was not significant. Critically, none of the quality indicators regarding the reading comprehension instrument used (standardized vs. non standardized tests, and the degree of internal consistency) resulted in significant differences.

Table 1

Results of the mixed-effects meta-regressions for the qualitative moderator variables on the effect sizes obtained from the studies

						Statistical moderator test		**Heterogeneity test**
Variable	***k***	***n***	***b***	***SE***	**95% C.I. [LL, UL]**	***F(df)***	***p***	***Qe(df)***
Reader factors								
Academic level	32	24				6.321 (1, 12.67)*	.026	251.038 (30)***
Primary and Secondary school	11		.057	.077	[-.136, .250]			
Undergraduate	21		-.227	.083	[-.405, -.049]			
Text factors								
Text genre	36	27				2.161 (3, 5.43)	.203	250.657 (33)***
Narrative	10		-.206	.102	[-.447, .036]			
Expository	21		-.083	.067	[-.228, .062]			

Variable	*k*	*n*	*b*	*SE*	95% C.I. [LL, UL]	Statistical moderator test *F(df)*	*p*	Heterogeneity test *Qe(df)*
Page navigation	24	18				0.031 (1, 5.10)	.867	29.727 (22)
Paging	19		-.109	.069	[-.260, .042]			
Scroll	5		-.149	.212	[-.744, .447]			
Test medium	36	29				0.273 (1, 8.23)	.615	255.941 (36)***
Tablet	28		-.103	.051	[-.211, .005]			
E-reader	8		-.155	.097	[-.392, .083]			
Activity factors								
Time-frame	35	27				0.036 (1, 12.58)	.853	56.321 (33)*
Free	23		-.108	.069	[-.254, .039]			
Time pressure	11		-.126	.081	[-.313, .061]			
Comprehension type	25	20				0.9100 (2, 1.78)	.909	237.714 (22)***
Textual	4		-.015	.169	[-.593, .563]			
Inferential	2		-.033	.157	[-.2.03, 1.96]			
Mixed	19		-.100	.075	[-.265, .065]			
Session modality	33	24				5.127 (1, 17.74)*	.036	248.428 (31)***
In group	15		.004	.062	[-.142, .149]			
Individual	18		-.230	.083	[-.413, -.048]			
Standardized test	38	29				1.265 (1, 3.52)	.332	254.008 (36)***
Standardized	6		-.239	.125	[-.708, .230]			
Non-standardized	32		-.085	.055	[-.199, .028]			
Reliability of the instrument	38	29				0.700 (1, 20.84)	.412	256.220 (36)***
Questionable	20		-.152	.065	[-.292, -.013]			
Acceptable	18		-.070	.074	[-.239, .098]			
Extrinsic variables								
Published status	38	29				0.180 (1, 13.11)	.678	255.137 (36)***
Non-published	27		-.098	.063	[-.232, .035]			
Published	11		-.142	.082	[-.344, .058]			

Note. *k:* number of effect sizes. *n:* number of studies (clusters). *g:* mean effect size (*Hedge's g). F*: statistical test for testing the significance of the moderator variable. Q_E: statistic for testing the model misspecification. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 2

Results of the mixed-effects meta-regressions for the continuous moderator variables on the effect sizes obtained from the studies

Variable	*k*	*n*	*b*	*SE*	95% CI	$Q_{E(df)}$	*F(df)*	*p*
Text factors								
Total word count	22	16	-.000	.000	[-.000, .000]	28.713 (20)	.587 (1, 2.79)	.503
Average word count	22	16	-.000	.000	[-.000, .000]	28.977 (20)	.001 (1, 1.54)	.983
Context factors								
Publication year	38	29	.003	.016	[-.031, .036]	256.241 (36)***	.027 (1, 12.49)	.872
Additional methodological factors								
Sample size	38	29	-.00	.00	[-.000, .000]	255.155 (36)***	.023 (1, 1.17)	.902

Note. *k:* number of effect sizes. *n:* number of studies. *b:* unstandardized regression coefficient. *SE:* Standard error. *F*: statistical test for testing the significance of the moderator variable. Q_E: statistic for testing the model misspecification. *$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

The present study represents a unique contribution to the field of research on the influence of reading medium on text comprehension. Besides the synthesis of factors relevant for the screen inferiority effect based on the reading for understanding framework (Snow et al., 2002), this investigation represents the first attempt to meta-analyze empirical evidence on reading comprehension outcomes specifically from reading on handheld devices (mostly tablets), as compared to reading in print. Results from the two meta-analyses reported yielded a significant mean effect indicating slightly better text comprehension outcomes when reading in print than on handheld devices. The pattern concurs with recent correlational evidence showing a negative association between frequency of use of tablets in school and reading comprehension (Bryant et al., 2020). Both effect sizes are small according to the standards proposed by Cohen (1988). Evidence from the between-participants studies showed that the medium effect was significantly larger for adult students (vs. school-age children), and when the experimental task was completed in individual sessions (vs. group). Evidence from the within-participants studies showed low heterogeneity, meaning that such mean effect size was reasonably consistent across studies (i.e., the effect size varied over roughly 0.083 points in *dc* units), leaving little room for any effects of moderators. We next discuss our results in detail.

Handheld devices and the screen inferiority effect

Results from our two meta-analyses qualify previous claims suggesting that handheld devices provide a reading experience that is closer to books than computers (Mangen & Pirhonen, 2022). The overall effect sizes in the two reported meta-analyses indicated a screen inferiority effect for handheld devices, which nevertheless is approximately half the size of that found in previous synthesis that mostly analyzed paper-computer comparisons (g = -0.25, Clinton, 2019; g = -0.21, d_c = -0.21, Delgado et al., 2018; *RVE* = -0.21, Kong et al., 2018).

In interpreting the overall effect sizes from our meta-analyses a note of caution is needed. Almost all original studies identified used small convenience samples, a situation that raises concerns about lack of statistical power and limited generalizability (Stanley et al., 2018). Consequently, only a few cross-media comparisons included in our datasets provided significant effects. Notably, the only two comparisons that used large and representative samples and a standardized test yielded significant effects in favor of in-print comprehension (Jewsbury et al., 2020). In addition, an overview of our dataset yields that in the cases in which differences between reading media among the effect sizes in primary studies were significant, they consistently favored text comprehension outcomes when reading in print. All in all, this pattern of results suggests that differences favoring reading in print are not arbitrary.

Nevertheless, despite the statistical significance in our study, the small magnitude of the effect sizes, which is approximately half the size of that found for computers (e.g., Clinton, 2019; Delgado et al., 2018), suggests that handheld devices may still play a role in supporting students' text comprehension. As compared to computers, handheld devices are deemed to allow physical interaction that is closer to reading in print, especially because they can be held by the reader (Mangen & Pirhonen, 2022). This feature, for example, allows the reader to get physically closer to the text when processing the content more in-depth (Ballenghein et al., 2020). Moreover, the layering of text pages on a handheld device is also more similar to that of pages in printed texts, which supports strategic backtracking to a higher extent than scrolling (Haverkamp et al., 2023).

Furthermore, according to the results from our analyses of the moderators following the reading for understanding framework (Snow et al., 2002), the overall screen inferiority increased among undergraduate students and when reading in individual sessions. We next elaborate on these findings.

Reader factors

Results from the two meta-analyses reported depict a complex pattern for the potential moderator effect of academic level on the relation between reading media and comprehension. As discussed above, the low heterogeneity found in the studies that used within-participants designs left little room for moderating effects. Accordingly, from this evidence we can conclude that the educational stage does not moderate the screen inferiority effect on handheld devices. On the contrary, the analyses of moderators for the studies using between-participants designs revealed that the screen inferiority effect was higher for undergraduate students than for school-age students. Further caution should be taken in interpreting this pattern because, as discussed above, our between-participants analysis only included two studies that used large and representative samples, from fourth and eighth grade (Jewsbury et al., 2020). In both cases, the effect size was significant and favored groups that read in print, as opposed to those reading on tablet. In sum, considering the evidence from the studies using within-participants designs, as well as the results from the only two studies using between-participants designs that assess representative samples, we cannot rule out the possibility that the screen inferiority effect of handheld devices negatively affects primary and secondary school students.

What remains unquestionable from the results of our two meta-analyses is that undergraduate students are negatively affected by reading on handheld devices, which is in accordance with previous meta-analyses (Clinton, 2019; Delgado et al., 2018; Furenes et al., 2021; Kong et al., 2018; Öztop, & Nayci, 2021). Through the lens of the shallowing hypothesis (Annisette & Lafreniere, 2017), it is plausible to consider that the effect remains salient among those who have more extensive experience with superficial processing on screen during the course of their lives. In addition, we cannot discard a confounding effect between reader and text characteristics, namely academic level and text length or genre. The effect of shallow digital processing may only be visible on the comprehension outcomes when students are confronted with particularly challenging, expository texts. This issue remains open and could be explored in future studies.

Other relevant reader factors identified in our review, such as reading comprehension skills, topic interest, and learning difficulties, could not be analyzed in our metaanalysis due to the limited number of studies available. Thus, these factors could be explored in further research.

Text factors

Among the text factors that did not moderate the screen inferiority effect, the absence of influence of the text genre —expository vs. narrative— was the most unexpected finding. Previous meta-analyses have concluded that screen inferiority appears when reading expository texts, whereas reading narrative texts yields similar comprehension outcomes regardless of the medium (Clinton, 2019; Delgado et al., 2019; Schwabe et al., 2022). Similar to the explanation for the effect of the reading time-frame also found in previous research, expository texts often demand greater cognitive effort (Clinton et al., 2020), thus representing a particularly challenging reading scenario for shallower on-screen reading. However, this was not the case among the studies analyzed in the two meta-analyses, which suggests that readers' cognitive effort on handheld devices is closer to that when reading in print. This could also explain why the text length did not play a moderating role.

With respect to page navigation, no difference was found between horizontal paging and vertical scrolling. Although previous research suggested that scrolling could hinder readers' construction of text structure (Haverkamp et al., 2023), this seems not to be the case when reading on handheld devices. However, it should be noted that only five primary effect sizes in our dataset came from studies in which vertical scrolling was used to navigate the texts. Given that this group of effect sizes yielded a mean effect size slightly higher than those that used paging, more research is necessary to experimentally test the effect of different types of page navigation.

In addition, our results showed that tablets and e-readers are equivalent reading devices in terms of text comprehension, which corroborates prior studies that analyzed both types of devices in a single study (Baker, 2010; Kretzschmar et al, 2013; Norman, & Furnes, 2016, Exp. 1; Wang & Lin, 2016). Nevertheless, this pattern of results contradicts the expectation that uses other than reading, which are typical for tablets, may boost the association of this type of device with a superficial processing mindset. In sum, both types of handheld devices provide similar physical reader-text interaction that, due to their increased similarity to print reading as compared to computers, seems to mitigate the screen inferiority effect to some extent.

Activity factors

Our meta-analyses showed mixed evidence for the role of activity factors on the screen inferiority effect. Unexpectedly, our moderator analyses indicated that reading pace did not influence the screen inferiority effect. We predicted that the studies imposing reading time limits would yield higher reading medium differences than those providing self-paced reading time, which supports the idea that a shallow digital mindset would be particularly detrimental when the task conditions demand high cognitive engagement, as is the case of performing a task under time pressure. This moderating effect has been found in experimental studies using computers as digital devices (Ackerman & Lauterman, 2012; Delgado & Salmerón, 2021) and in the meta-analysis by Delgado et al. (2018). Although our analyses did not provide a clear explanation for this conflicting finding, we suggest that given that handheld devices seem to be more suitable than computers for in-depth reading comprehension, readers may invest sufficient cognitive effort for successful time management when reading on this type of device.

A second activity factor analyzed, text comprehension questions, did not moderate the effect of reading medium on comprehension, corroborating previous findings (Clinton, 2019; Delgado et al., 2018). It should be noted that only 6 of the 30 between-participants designs' studies used measures that targeted either textual or inferential information. Among those, a tendency arose towards a higher screen inferiority effect in inferential questions, as predicted by the shallowing hypothesis.

Of concern is that a majority of studies analyzed used multiple-choice questions to measure text comprehension. This approach offers several benefits, for example, it allows for constructing reliable and efficient measurements. Nevertheless, it also comes with limitations, as students may conceive such scenarios as problem-solving tasks rather than purposeful comprehension reading (Rupp et al., 2006). Accordingly, results from those studies may not generalize to more naturalistic non-testing scenarios. Future research should incorporate additional methods better aligned to capture higher order comprehension processes, such as integrative open questions or argumentative essays (e.g., Latini et al., 2019, 2020).

A third activity factor, group setting, significantly affected between-participants studies. The overall negative effect of on-screen reading on comprehension outcomes appeared in those studies in which the experimental reading task was performed in individual sessions, as compared to group sessions. This is against the expectation that, as readers may need further increased attention during on-screen reading to perform similarly to reading in print, a context that adds further distraction, such as group reading, may particularly potentiate the screen inferiority effect (Anguiano, 2020; Latini & Bråten, 2022). A potential explanation for this paradoxical effect may look at an interaction between activity and reader characteristics. As discussed above, among studies that used between-participants designs, the screen inferiority effect was significant for undergraduates, but not for students of earlier educational stages. Studies conducted with school-age students are often performed at students' schools, often in their own classrooms, whereas those conducted with undergraduate students are more often performed in research labs at universities. Thus, even if the undergraduates participate in group experimental sessions, it is easier to control for external distractions in a research lab than in school classrooms in which situations that divert participants' attention from the experimental task appear much more easily. It is thus plausible to argue that the superiority of print in terms of reader's engagement with the text appears in situations in which the reading context allows for concentration and engagement —which would explain the absence of this effect in group settings—, and that such experimental settings are typical of studies conducted with adults. A more detailed examination of this possibility could be performed by meta-analyzing the interaction between activity and text factors, namely the type of experimental setting (individual vs group) and the participants' educational stage. However, our dataset also did not allow us to perform such a comparison due to the reduced number of some of those combinations. Future research could further explore this issue by manipulating the level of distraction of an individual or group activity, as in studies analyzing students' multitasking (Clinton-Lisell, 2021).

Lastly, examining the influence of the possibility of holding the reading device would help to clarify whether handheld devices provide a reading experience closer to printed texts because of similarities in reader-text physical interaction (Mangen & Pirhonen, 2022; Mangen & Van der Weel, 2016). Unfortunately, only two studies in our dataset did not allow the participants to hold the reading device. This line of research can be explored in future experimental studies.

Context factors

With respect to the socio-cultural context, we analyzed the effects of generation, as indicated by publication year. Based on prior evidence, we expected larger differences between reading in print and on tablets in more recent studies (Delgado et al., 2018). Given that access to digital reading devices has become almost permanent and ubiquitous in recent years, the chances that people have established a strong digital shallow mindset have increased. Contrary to this expectation, we found null association between year of publication (2010-2022) and the screen inferiority effect. There are two possible, complementary explanations. On the one hand, this result can be due to the screen inferiority being less influential when reading on handheld devices. Thus, it is reasonable to argue that this effect has not increased when it comes to this type of device. Of note is that the studies analyzed in Delgado et al. (2018) were mostly conducted with computers. On the other hand, another major difference with Delgado et al.'s (2018) meta-analysis is the range of years, as that work included earlier studies and a longer timespan (2000-2017). Given that most of the Web 2.0 systems that are assumed to be associated with shallow processing (e.g., social media) emerged during the last part of the 2000s, it can be argued that on-screen reading during the first decade of 2000 resembled print reading practices to a larger extent than current on-screen reading. Accordingly, the absence of studies from the first decade of 2000 in the present meta-analysis, due to the fact that tablets were created in 2010, could explain the lack of generational changes.

Our review pointed to a surprising lack of research on other relevant context factors that could moderate the effects of reading medium on comprehension, particularly students' SES and degree of digitization in schools (OECD, 2015). Potentially, those factors could contribute positively to text comprehension on handheld devices via increased familiarity with digital technology, or more advanced teachers' practices to regulate digital distractions and shallow processing. Again, those are questions that could be analyzed in further research.

Limitations

These meta-analyses are not exempt from limitations. Importantly, in interpreting the results from our meta-analyses, caution should be exercised to avoid overgeneralizing the findings and drawing definitive conclusions without considering potential limitations of our dataset. The screen inferiority effect for handheld devices identified in our analyses characterizes students' performance in reading situations where they are taking multiple-choice text comprehension questionnaires. The effect was robust and arised regardless of the reliability of the reading comprehension test used. Nevertheless, our results cannot inform about the potential nuances to the screen inferiority effect that could come from factors (reader, text, activity, or context) that have not been systematically analyzed in the original studies. Accordingly, cautiousness is advised in using our results to inform educational policy. For example, the negative association between frequency of use of digital reading devices and text comprehension observed in NAEPS 2017 was attenuated when teachers had taken courses on how to integrate digital literacy in their classrooms, or when they used devices to support authentic reading tasks such as reading projects (Salmerón et al., 2023). In addition, it is important to highlight that only two primary studies from our dataset were conducted with large representative samples, in both cases reporting a significant screen inferiority effect in fourth and eighth grade students (Jewsbury et al., 2020). Future studies should be performed on larger and representative samples. Moreover, it should be noted that the analyses of potential moderators were also jeopardized by the lack of necessary information in the research reports of the primary studies. We therefore call for increasing effort in detailing research methods exhaustively. This will help not only explain the results in the field, but also will make it possible to perform replicating studies. Lastly, given that the screen inferiority is a small effect from a statistical point of view, the fact that most of the primary studies in our two datasets assessed small samples requires us to be cautious in interpreting our results. In fact, the only two studies conducted with large representative samples found a significant screen inferiority effect in fourth and eighth grade students (Jewsbury et al., 2020). Future studies should be performed on larger and more representative samples.

Conclusion and educational implications

As tablets become common devices in schools, a critical question for the research community is to assess their potential impact on educational outcomes (Slavin, 2020). Unfortunately, it has become common practice that research about tablets in education tends to overemphasize their positive aspects, while uncritically ignoring any negative effects (Mertala, 2021). An honest interpretation of our results identifies a more nuanced scenario: while handheld devices seem to provide a reading experience that is closer to printed books than that of computers, still they negatively affect comprehension. Accordingly, we recommend caution when using such devices in educational scenarios where the goal is to promote or to assess reading comprehension (Bryant et al., 2020). While software and digital pedagogical methods must be carefully designed to promote students' reading comprehension, we recommend using printed texts in schools for the time being, given their proven effectiveness and accessibility. Nevertheless, we recognize the potential benefits of digital media, and encourage educators to conscientiously explore the use of handheld devices, over computers, to work on reading comprehension tasks.

References

The complete reference list of the studies included in the meta-analysis datasets are located in the supplemental online material.

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior*, *28(5)*, 1816-1828. https://doi.org/10.1016/j.chb.2012.04.023

Al Amer, F. M., & Lin, L. (2021). Empirical assessment of prediction intervals in Cochrane meta-analyses. *European Journal of Clinical Investigation, 51*(7), e13524. https://doi.org/10.1111/eci.13524

Anguiano, C. J. (2020). *Effects of Multimedia-Enhanced Storybooks on Young Children'Vocabulary and Comprehension Knowledge: A Meta-Analysis*. [Doctoral dissertation, Washington State University]. https://www.proquest.com/dissertations-theses/effects-multimedia-enhanced-storybooks-on-young/docview/2451868239/se-

Annisette, L. E., & Lafreniere, K. D. (2017). Social media, texting, and personality: A test of the shallowing hypothesis. *Personality and Individual Differences*, *115,* 154-158. https://doi.org/10.1016/j.paid.2016.02.043

Baker, R. D. (2010). *Comparing the readability of text displays on paper, e-book readers, and small screen devices.* [Doctoral dissertation, University of North Texas]. https://www.proquest.com/dissertations-theses/comparing-readability-text-displays-on-paper-e/docview/746481524/se-2

Ballenghein, U., Kaakinen, J. K., Tissier, G., & Baccino, T. (2020). Cognitive engagement during reading on digital tablet: Evidence from concurrent recordings of postural and eye movements. *Quarterly Journal of Experimental Psychology, 73*(11), 1820-1829. https://doi.org/10.1177/1747021820931830

Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world.* Oxford University Press.

Baron, N. S. (2021). *How we read now: Strategic choices for print, screen, and audio*. Oxford University Press.

Baron, N. S., Calixte, R. M., & Havewala, M. (2017). The persistence of print among university students: An exploratory study. *Telematics and Informatics, 34(5),* 590-604 https://doi.org/10.1016/j.tele.2016.11.008

Ben-Yehudah, G., & Brann, A. (2019). Pay attention to digital text: The impact of the media on text comprehension and self-monitoring in higher-education students with ADHD. *Research in Developmental Disabilities, 89,* 120-129. https://doi.org/10.1016/j.ridd.2019.04.001

Bon, E. V., & Burke, M. (2022). Devices, settings and distractions: A study into how students read literature. In S. Zyngier and G. Watson (Eds.) *Pedagogical Stylistics in the 21st Century (pp. 183-206).* Springer.

Boon, H. J., Boon, L., & Bartle, T. (2021). Does iPad use support learning in students aged 9–14 years? A systematic review. *The Australian Educational Researcher, 48*(3), 525-541. https://doi.org/10.1007/s13384-020-00400-0

Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them.* Biostat, Inc.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Botella, J. & Sánchez-Meca, J. (2015). *Meta-análisis en Ciencias Sociales y de la Salud [Meta-analysis in Social and Health Sciences]*. Síntesis

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*(4), 927-931. https://doi.org/10.1016/j.compedu.2009.09.024

Bryant, J., Child, F., Dorn, E., & Hall, S. (2020). New global data reveal education technology's impact on learning. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/New%20global%20data%20reveal%20education%20technologys%20impact%20on%20learning/New-global-data-reveal-education-technologys-impact-on-learning.pdf

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

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

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, *42*(2), 288-325. https://doi.org/10.1111/1467-9817.12269

Clinton, V., Taylor, T., Bajpayee, S., Davison, M. L., Carlson, S. E., & Seipel, B. (2020). Inferential comprehension differences between narrative and expository texts: a systematic review and meta-analysis. *Reading and Writing*, *33*(9), 2223-2248. https://doi.org/10.1007/s11145-020-10044-2

Clinton-Lisell, V. (2021). Stop multitasking and just read: Meta-analyses of multitasking's effects on reading performance and reading time. *Journal of Research in Reading, 44*(4), 787-816. https://doi.org/10.1111/1467-9817.12372

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Routledge.

Delgado, P., & Salmerón, L. (2021). The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction, 71,* 101396. https://doi.org/10.1016/j.learninstruc.2020.101396

Delgado, P. & Salmerón, L. (2022). Cognitive effort in text processing and reading comprehension in print and on tablet: An eye-tracking study. *Discourse Processes, 59,* 237-274. https://doi.org/10.1080/0163853X.2022.2030157

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25,* 23-38. https://doi.org/10.1016/j.edurev.2018.09.003

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35,* 1297-1236. https://doi.org/10.1080/00140139208967394

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research, 91(4),* 483-517. https://doi.org/10.3102/0034654321998074

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32,* 139–156. https://doi.org/10.1111/jcal.12123

Haddock, G., Foad, C., Saul, V., Brown, W., & Thompson, R. (2020). The medium can influence the message: Print-based versus digital reading influences how people process different types of written information. *British Journal of Psychology, 111*(3), 443-459. https://doi.org/10.1111/bjop.12415

Haverkamp, Y. E., Bråten, I., Latini, N., & Salmerón, L. (2023). Is it the size, the movement, or both? Investigating effects of screen size and text movement on processing, understanding, and motivation when students read informational text. *Reading and Writing*, *50*(3), 113-176. https://doi.org/10.1007/s11145-022-10328-9

Hedges L. V., Olkin I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Herman, H. A. (2017). *The Effects of Literacy Support Tools on the Comprehension of Informational e-Books and Print-Based Text*. [Doctoral dissertation, Widener University]. https://www.proquest.com/docview/1955166316?parentSessionId=7jNqbwRMA0x5AEjnqyYL3RdACG2Pa3ogu9AaCYyKi%2BY%3D

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions (Version 5.1.0).* The Cochrane Collaboration.

Hou, J., Wu, Y., & Harrell, E. (2017). Reading on paper and screen among senior adults: Cognitive map and technophobia. *Frontiers in Psychology, 8,* 2225. https://doi.org/10.3389%2Ffpsyg.2017.02225

Hsiao, K. L., & Chen, C. C. (2015). How do we inspire children to learn with e-readers? *Library Hi Tech*, *33*(4), 584 - 596. https://doi.org/10.1108/LHT-04-2015-0038

Jensen, P. S., Mrazek, D., Knapp, P. K., Steinberg, L., Pfeffer, C., Schowalter, J., & Shapiro, T. (1997). Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 1672–1681. https://doi.org/10.1097/00004583-199712000-00015

Jeong, Y. J., & Gweon, G. (2021). Advantages of print reading over screen reading: A comparison of visual patterns, reading performance, and reading attitudes across paper, computers, and tablets. *International Journal of Human–Computer Interaction, 37*(17), 1674-1684. https://doi.org/10.1080/15391523.2020.1713264

Jewsbury, P., Finnegan, R., Xi, N., Jia, Y., Rust, K., & Burg, S. (2020). *2017 NAEP transition to digitally based assessments in mathematics and reading at Grades 4 and 8: Mode evaluation study.* https://nces.ed.gov/nationsreportcard/subject/publications/main2020/pdf/transitional_whitepaper.pdf

Kingston, N.M. (2008). Comparability of computer- and paper- administered multiple-choice tests for K–12 populations: A synthesis. *Applied Measurement in Education, 22*, 22-37. https://doi.org/10.1080/08957340802558326

Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education*, *123*, 138-149. https://doi.org/10.1016/j.compedu.2018.05.005

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: Concurrent EEG-eyetracking evidence from the reading of books and digital media. *PloS one, 8*(2), e56178. https://doi.org/10.1371/journal.pone.0056178

Latini, N., & Bråten, I. (2022). Strategic text processing across mediums: A verbal protocol study. *Reading Research Quarterly, 57*(2), 493-514. https://doi.org/10.1002/rrq.418

Latini, N., Bråten, I., & Salmerón, L. (2020). Does reading medium affect processing and integration of textual and pictorial information? A multimedia eye-tracking study. *Contemporary Educational Psychology*, *62*, 101870. https://doi.org/10.1016/j.cedpsych.2020.101870

Lenhard, W., Schroeders, U., & Lenhard, A. (2017). Equivalence of screen versus print reading comprehension depends on task complexity and proficiency. *Discourse Processes*, *54*(5-6), 427-445. https://doi.org/10.1080/0163853X.2017.1319653

Liang, T. H., & Huang, Y. M. (2014). An investigation of reading rate patterns and retrieval outcomes of elementary school students with E-books. *Educational Technology & Society*, *17* (1), 218–230. https://www.jstor.org/stable/jeductechsoci.17.1.218

Mangen, A. (2010). Point and click: Theoretical and phenomenological reflections on the digitization of early childhood education. *Contemporary Issues in Early Childhood*, *11(4)*, 415-431. https://doi.org/10.2304/ciec.2010.11.4.415

Mangen, A., Olivier, G., & Velay, J. L. (2019). Comparing comprehension of a long text read in print book and on Kindle: Where in the text and when in the story? *Frontiers in Psychology*, *10*, 38. https://doi.org/10.3389/fpsyg.2019.00038

Mangen, A. & Pirhonen, A. (2022). Reading, writing, technology, and embodiment. In S. L. Macrine and J. M.B. Fugate (Eds.) *Movement matters. How embodied cognition informs teaching and learning* (pp. 103-117). MIT Press.

Mangen, A., & Van der Weel, A. (2016). The evolution of reading in the age of digitisation: an integrative framework for reading research. *Literacy*, *50(3)*, 116-124. https://doi.org/10.1111/lit.12086

Mertala, P. (2021). 'It is important at this point to make clear that this study is not "anti-iPad"': Ed-Tech speak around iPads in educational technology research.

Learning, Media and Technology, 46(2), 230-242. https://doi.org/10.1080/17439884.2021.1868501

Nikkelen, S. W., Valkenburg, P. M., Huizinga, M., & Bushman, B. J. (2014). Media use and ADHD-related behaviors in children and adolescents: A meta-analysis. *Developmental Psychology*, *50*(9), 2228. https://doi.org/10.1037/a0037318

Nishizaki, D. M. (2015). The effects of tablets on learning: Does studying from a tablet computer affect student learning differently across educational levels. *CMC Senior Theses*, 1011. https://scholarship.claremont.edu/cmc_theses/1011

Norman, E., & Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior*, *54*, 301-309. https://doi.org/10.1016/j.chb.2015.07.043

Noyes, J. M. & Garland K. J. (2008) Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*, 1352-1375. https://doi.org/10.1080/00140130802170387

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91-121. https://doi.org/10.1080/10888438.2010.529219

Öztop, F., & Nayci, Ö. (2021). Does the digital generation comprehend better from the screen or from the paper?: A meta-analysis. *International Online Journal of Education and Teaching, 8(2),* 1206-1224. https://eric.ed.gov/?id=EJ1294459

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery, 88,* 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Pustejovsky, J. E. (2020). clubSandwich: Cluster-Robust (Sandwich) Variance estimators with small-sample corrections (0.4.2) [R package]. ttps://github.com/jepusto/clubSandwich

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, *10*, 57–71. http://dx.doi.org/10.1002/jrsm.1332

Rideout, V., & Robb, M. B. (2019). *The Common Sense census: Media use by tweens and teens*. Common Sense Media.

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2015). Matching learning style to instructional method: Effects on comprehension. *Journal of Educational Psychology*, *107*(1), 64–78. https://doi.org/10.1037/a0037478

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23(4),* 441-474. https://doi.org/10.1191/0265532206lt337oa

Rvachew, S., Rees, K., Carolan, E., & Nadig, A. (2017). Improving emergent literacy with school-based shared reading: Paper versus ebooks. *International Journal of Child-Computer Interaction*, *12*, 24-29. https://doi.org/10.1016/j.ijcci.2017.01.002

Salmerón, L., Delgado, P., Vargas, C., & Gil, L. (2021). Tablets for all? Testing the screen inferiority effect with upper primary school students. *Learning and Individual Differences, 86,* 101975. https://doi.org/10.1016/j.lindif.2021.101975

Salmerón, L., Vargas, C., Delgado, P., & Baron, N. (2023). Relation between digital tool practices in the language arts classroom and reading comprehension scores. *Reading and Writing, 36,* 175–194. https://doi.org/10.1007/s11145-022-10295-1

Schneps, M. H., Thomson, J. M., Chen, C., Sonnert, G., & Pomplun, M. (2013). E-readers are more effective than paper for some with dyslexia. *PloS one*, *8*(9), e75634. https://doi.org/10.1371/journal.pone.0075634

Schwabe, A., Lind, F., Kosch, L., & Boomgaarden, H. G. (2022). No negative effects of reading on screen on comprehension of narrative texts compared to print: A meta-analysis. *Media Psychology, 25(6),* 779-796. https://doi.org/10.1080/15213269.2022.2070216

Singer, L.M. & Alexander, P.A. (2017). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research, 87,* 1007-1041. https://doi.org/10.3102/0034654317722961

Slavin, R. E. (2020). How evidence-based reform will transform research and practice in education. *Educational Psychologist, 55*, 21-31. https://doi.org/10.1080/00461520.2019.1611432

Snow C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND Education.

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin, 144*(12), 1325. https://psycnet.apa.org/doi/10.1037/bul0000169

Stephens, M. (2007). Web 2.0, Library 2.0, and the hyperlinked library. *Serials Review, 33*(4), 253-256. https://doi.org/10.1016/j.serrev.2007.08.002

Stern, P., & Shalev, L. (2013). The role of sustained attention and display medium in reading comprehension among adolescents with ADHD and without it. *Research* *in Developmental Disabilities, 34*(1), 431-439. https://doi.org/10.1016/j.ridd.2012.08.021

Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: A mode-effect study. *Computers & Education, 151*, 103861. https://doi.org/10.1016/j.compedu.2020.103861

Tipton, E., and Pustejovsky, J. E. (2015), Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Valenzuela, Á., & Castillo, R. D. (2023). The effect of communicative purpose and reading medium on pauses during different phases of the textualization process. *Reading and Writing, 36,* 881–908. https://doi.org/10.1007/s11145-022-10309-y

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*, 576-594. https://doi.org/10.3758/s13428-012-0261-6

Vasilev, M. R., Kirkby, J. A., & Angele, B. (2018). Auditory distraction during reading: A Bayesian meta-analysis of a continuing controversy. *Perspectives on Psychological Science, 13*(5), 567-597. https://doi.org/10.1177/1745691617747398

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

Viechtbauer, W. (2021). Model checking in meta-analysis. In C. H. Schmid, T. Stijnen, & I. R. White (Eds.), *Handbook of meta-analysis* (pp. 219-254). CRC Press.

Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Wang, A. H., & Lin, K. C. (2016). Effects of display type and ambient illuminance on the comprehension performance of young and elderly readers. *Journal of Industrial and Production Engineering, 33*(7), 443-449. https://doi.org/10.1080/21681015.2016.1156033

Wang, S., Jiao, H., Young, M.J., Brooks, T. & Olson, O. (2007). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments. A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, *68*, 5-24. https://doi.org/10.1177/0013164407305592

Warrens M. J. (2010). A formal proof of a paradox associated with Cohen's kappa. *Journal of Classification*, *27*(3), 322–332. https://doi.org/10.1007/s00357-010-9060-x

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4,* 58. https://doi.org/10.3389/fpsyg.2013.00058

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