Review

Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension

Pablo Delgado[a], Cristina Vargas[b], Rakefet Ackerman[c], Ladislao Salmerón[a],[*]

[a] ERI Lectura and Department of Developmental and Educational Psychology - University of Valencia, Avd. Blasco Ibáñez, 21, 46010, Valencia, Spain
[b] Department of Developmental and Educational Psychology - University of Valencia, Avd. Blasco Ibáñez, 21, 46010, Valencia, Spain
[c] Faculty of Industrial Engineering & Management, Technion-Israel Institute of Technology, Haifa, 3200003, Israel

ABSTRACT

With the increasing dominance of digital reading over paper reading, gaining understanding of the effects of the medium on reading comprehension has become critical. However, results from research comparing learning outcomes across printed and digital media are mixed, making conclusions difficult to reach. In the current meta-analysis, we examined research in recent years (2000–2017), comparing the reading of comparable texts on paper and on digital devices. We included studies with between-participants ($n = 38$) and within-participants designs ($n = 16$) involving 171,055 participants. Both designs yielded the same advantage of paper over digital reading (Hedge's $g = -0.21$; $d_c = -0.21$). Analyses revealed three significant moderators: (1) time frame: the paper-based reading advantage increased in time-constrained reading compared to self-paced reading; (2) text genre: the paper-based reading advantage was consistent across studies using informational texts, or a mix of informational and narrative texts, but not on those using only narrative texts; (3) publication year: the advantage of paper-based reading increased over the years. Theoretical and educational implications are discussed.

1. Introduction

There has been a gradual shift from paper-based reading to reading on digital devices, such as computers, tablets, and cell-phones. Although there are clear advantages of digital-based assessment and learning, including reduced costs and increased in-dividualization, research indicates that there may be disadvantages as well, as described below. In addition, findings from previous reviews of studies on the effects of digital reading on comprehension have been inconclusive (Dillon, 1992; Kingston, 2008; Noyes & Garland, 2008; Singer & Alexander, 2017b; Wang, Jiao, Young, Brooks, & Olson, 2007). The current paper presents a meta-analysis of recent studies that investigated the effects of paper versus digital media on reading comprehension. In addition, we also explored the effects of several potential moderator variables whose influence may help to explain previous inconsistencies among study results.

1.1. Text comprehension and the role of media

Theoretical models of reading comprehension have extensively considered the interplay among reader characteristics, text content and design, and reading instructions (for a review see McNamara & Magliano, 2009). However, the factor of the medium has

been mostly ignored, despite empirical evidence suggesting that it influences reading outcomes (e.g., Lenhard, Schroeders, & Lenhard, 2017; Mangen, Walgermo, & Brønnick, 2013; Singer & Alexander, 2017a). In particular, Ackerman and Lauterman (2012) considered media-related differences in learning outcomes from a metacognitive perspective. In addition to learning outcomes, they compared learners' monitoring of their comprehension and allocation of their study time. On each medium, immediately after studying each text, participants predicted their success rates (in %) and were tested through multiple-choice questions. Moreover, to the best of our knowledge, these authors are the only ones who empirically considered the time frame as a potential moderating factor of media effects on learning outcomes. They examined the learners' adjustment to studying under time pressure, compared to free study time, on both media. Under time pressure, but not under free time, those who read from computers showed screen inferiority: they had more pronounced overconfidence than paper learners and achieved lower test scores. Moreover, only in paper-based reading, participants improved their efficiency under time pressure, compared to learning in a free time frame. Importantly, whereas theories of monitoring and allocation of study time assume close relationships between the two, Ackerman and Goldsmith (2011) found close relationships in paper-based reading, but more erratic time allocation decisions in digital-based reading. Before this study, conducted with young undergraduates, weak associations between monitoring and time allocation decisions were only found in elderly people and people with mental illnesses (Koren, Sneidman, Goldsmith, & Harvey, 2006; Pansky, Goldsmith, Koriat, & Pearlman-Avnion, 2009). Furthermore, several recent studies found that the preference for paper over digital-based reading persists despite technological advances (Baron, Calixte, & Havewala, 2017; Mizrachi, 2015; Kurata, Ishita, Miyata, & Minami, 2017; but see; Singer & Alexander, 2017a). Lauterman and Ackerman (2014) found that methods to overcome screen inferiority are effective only for people who prefer digital reading, but not for those who prefer paper reading. Together, the reviewed findings demonstrate several aspects of reading comprehension that have been overlooked so far in reading theories, highlighting the medium as an environment that affects reading outcomes, above and beyond reader and task characteristics.

In sum, the way the media affect reading comprehension outcomes is still unclear. Several researchers have explained screen inferiority under some conditions as being due to people's stronger inclination toward shallow work in digital-based environments than in paper-based ones (see Annisette & Lafreniere, 2017; Wolf & Barzillai, 2009), particularly when the task design indicates its legitimacy, as when working under a limited time frame (Lauterman & Ackerman, 2014; Sidi, Shpigelman, Zalmanov, & Ackerman, 2017).

A meta-analysis provides an opportunity to examine media effects on learning outcomes while considering overall task characteristics, such as time frames, participant characteristics, and the display technology, across theoretical frameworks, populations, and methodologies. Importantly, a meta-analysis makes it possible to consider potentially moderating factors, even across studies that did not include these factors in their designs, by comparing enough studies that used each level of the factor (e.g., only limited time frame vs. only free time allocation). Exposing moderating factors can guide future theoretical development and practical recommendations.

1.2. Previous reviews and meta-analyses

In the past ten years, only a few meta-analyses and literature reviews have been undertaken to determine the nature of the medium's influence on reading outcomes. Wang et al. (2007) focused on K-12 student population. Their meta-analysis examined media effects on performance on standardized tests, and it included 11 primary studies that yielded 42 comparisons. They found better reading outcomes in paper-based testing than in digital-based testing. The mean effect size (0.08) was significant, but small (see Cohen, 1988), and this difference between reading media was larger in studies that used fixed linear computerized tests ($n = 37$) than in those that used adaptive computerized tests ($n = 5$). Wang and colleagues concluded that differences between testing media are probably test specific, so that an analysis of potential media effects should be conducted for each type of test separately.

Kingston (2008) conducted a larger meta-analysis that included 81 effect sizes from 16 studies. This study focused on testing academic achievement across several academic topics in K-12 populations, and it showed a small advantage for digital administration in English Language Arts and Social Studies (effect sizes of 0.11 and 0.15, respectively), along with a small advantage for paper administration in Mathematics (effect size of −0.06). More relevant to our focus, eight of the studies included in Kingston's work assessed reading outcomes, five of which were included in Wang et al.'s (2007) meta-analysis, and found no effect of reading media. Regarding the digital disadvantage in Mathematics, Kingston alludes to possible difficulties when completing tests on a computer due to switching to sketch paper before answering. In sum, results from these meta-analyses are inconsistent. Some findings point to advantages of print text, whereas others favour digital text, and still other results indicate that media effects depend on the topic.

Recently, Kong, Seo, and Zhai (2018) performed a meta-analysis with 17 studies dating from 2000 to 2016. Results revealed better performance when reading from paper than when reading from digital devices (effect size of −0.21). This meta-analysis incorporated a relatively small number of studies which included great variability in terms of populations (e.g. second-language students), and tasks (e.g. perceived comprehension or proofreading). Interestingly, despite considering several potential moderating factors, this analysis did not reveal any significant effects. The authors acknowledged the need for considering additional moderating factors.

Two narrative literature reviews attempted to promote understanding of media effects on reading comprehension. Noyes and Garland (2008) reviewed media comparison studies that focused on reading outcomes but also on tasks such as examinations, writing, and filling in questionnaires (e.g., psychometric tests and surveys). They concluded that, although equivalence between the media was a challenge, differences, where found, appeared to be task specific. In particular, with respect to reading outcomes, the results were heterogeneous regarding comprehension and reading speed, with no clear conclusions about the influence of the media.

Recently, Singer and Alexander (2017b) described studies published from 1992 to 2017. They found it difficult to reach

conclusions and pointed to a lack of clarity in definitions of paper and digital reading, as well as a lack of important information in many studies, such as text features (genre and length), individual differences (e.g., reading rate and vocabulary), validity and reliability of the tasks used to measure reading outcomes, characteristics of the reading tasks, levels of comprehension evaluated, and scoring criteria. Singer and Alexander called on researchers to investigate how various factors interact with media and potentially explain the mixed results found in the literature.

The main conclusion drawn from the above review of previous meta-analyses and narrative research synthesis is that media effects are inconsistent. This may be partially explained by the difficulty of comparing paper texts to digital texts which include incomparable features such as hyperlinks, animations, or adaptive tests which may confound and hide media effects on learning processes. Another potential reason for the inconsistent results is the fact that most of the previous reviews did not consider or did not find moderating factors. Finding robust moderating factors can shed light on the reasons for the seemingly inconsistent media effects found. As mentioned above, Ackerman and Lauterman (2012) found inferior comprehension in digital-based reading compared to paper-based reading under time pressure, but media equivalence in free time conditions. This finding raises the option that the time frame allowed for reading is a factor that differentiates between studies that find an advantage of paper and those that find media equivalence. Considering the time frame as a moderating factor across a large collection of studies can inform us whether this specific study exposed a pattern which is robust across methodologies and populations.

In the present meta-analysis we aimed to facilitate comparisons between print and digital media by including only studies that used linear reading materials, where the digital texts closely resembled the printed versions. This focus allowed us to eliminate some of the aforementioned complexities. In addition, by performing a comprehensive meta-analysis we aimed to examine the influence of several potential moderating factors on media effects, in addition to the time frame just mentioned. We see high importance in identifying moderating factors for pointing to conditions that yield an advantage of print across methodologies and conditions, those that yield an advantage of digital devices, and those that result in equivalent outcomes.

1.3. Effects of experience with digital technologies

It could be argued that a potential straightforward moderator of digital text comprehension is experience using technology. In other words, potential comprehension difficulties in digital reading will disappear once students have enough experience with digital technologies. According to this view, as each new generation is surrounded by digital devices earlier and earlier in life (e.g. ASHA, 2015; Childwise, 2017), we should expect newer generations to achieve equivalent, or even better, comprehension levels in digital-based reading compared to paper-based reading (see illustration in Fig. 1, left panel). To explore this view, we investigated whether the publication date reveals a decreasing advantage of paper in recent years due to greater exposure to technology than in earlier years. If this was the case, with enough experience with digital technologies, readers would be able to overcome any potential detrimental effect on comprehension. In our schematic presentation (Fig. 1), we use paper comprehension as the reference level and illustrate potential changes in digital-based comprehension relative to it. Importantly, because we analyse effect sizes rather than objective measures of performance, we cannot know whether this paper-based reference level changes over time. In particular, one could also argue that because new generations may have less exposure to printed texts, paper comprehension will decrease rather than remaining constant. In any of those two cases, the prediction about the evolution of digital-based reading from this perspective is that reading ability on this medium will improve with further experience. Therefore, the advantage of print over digital-based reading will decrease over the years, regardless of the pattern of change in paper comprehension.

Several researchers have argued, however, that increasing exposure to technology, with its emphasis on speed and multitasking, may encourage a shallower kind of processing that leads to a decrease in deep comprehension in digital environments (e.g. Lauterman & Ackerman, 2014; Wolf & Barzillai, 2009). Indeed, current evidence supports the claim that mere experience with digital technology does not improve students' comprehension skills, but instead has a detrimental effect (Duncan, McGeown, Griffiths,

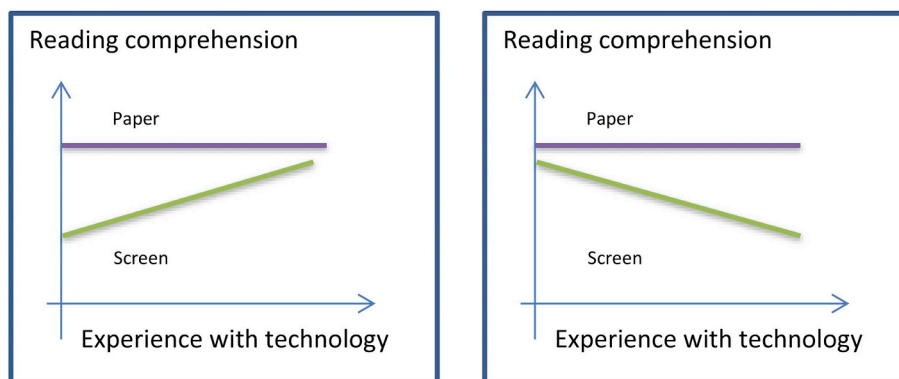**Fig. 1.** Schematic projection of trends for the effect of experience with technology on reading comprehension differences between print and digital devices. Left panel represents a situation in which more experience with technology reduces the difference between print and digital reading outcomes. Right panel represents a situation in which this potential difference increases over the years.

Stothard, & Dobai, 2015; Pfost, Dörfler, & Artelt, 2013). This view leads to the alternative hypothesis that the paper advantage over digital media increases with time (Fig. 1, right panel). If true, this would be a call for researchers, policy-makers, and education professionals to join forces to develop methods to support effective digital-based reading and learning.

1.4. Objectives

The aim of this meta-analysis was to gain a broad perspective of empirical studies comparing digital and print reading outcomes. Specifically, we had two objectives:

1) Examine whether the reading medium affects reading comprehension outcomes.
2) Identify moderating factors of the effects of the medium on reading comprehension outcomes.

2. Method

2.1. Selection criteria of the studies

Studies included in the meta-analysis met the following criteria:

1. The study compares comprehension in paper-based and digital-based reading, respectively defined as reading texts printed on paper and reading texts displayed on digital screens, including computers, tablets, mobiles phones, and e-readers.
2. Participants read individually and silently.
3. Reading materials are comparable across media in terms of text content, structure, and presence of images. Therefore, specific features of digital environments, such as hyperlinks or web navigation, are not present in the digital-based condition.
4. Participants study in their daily-used language.
5. Participants are a sample from a normative population (i.e., typical development, no reading difficulties, and no cognitive impairments or disorders).
6. The study makes an empirical contribution that includes the results of the comparison (i.e. the paper is not a review or an opinion).
7. The study was published or presented from the year 2000–2017. Formal publication was not required.
8. The report is written in English.
9. The report includes specification of the effect size or sufficient statistical information to calculate it (or this information was provided by the authors following a personal request).
10. The statistical data allow parametric analyses.

2.2. Search procedure

Several literature search procedures were used to locate relevant studies and previous reviews. Firstly, some electronic databases were consulted: PsycInfo, Eric, Proquest Psychology, Web of Science, Scopus (Physical Sciences and Social Sciences & Humanities), dissertation and theses (Proquest), and Google Scholar. The search included the following terms[1]: "("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 "media effects" OR "reading medium" OR "mode effect") AND (memory OR comprehension OR retention OR "test performance" OR learning)". These terms were searched as title, abstract, or keywords. As recommended by Card (2012), we complemented the search with additional strategies. Thus, secondly, references included in previous reviews were examined. Thirdly, we approached experts and societies in this area (The Society for Text and Discourse, Society for the Scientific Study of Reading, The European Association for Research on Learning and Instruction, and COST E-READ Action) asking for information about unpublished studies. Fourthly, a forward search was performed using Google Scholar to find studies that cited the works selected. Finally, references from the selected studies were also retrieved. The search ended in May 2017.

The search described above yielded 1,840 records. The selection process from this initial collection is described in Fig. 2. We ended up with 54 studies that satisfied all the inclusion criteria. Some studies reported more than one media comparison due to considering additional independent factors (e.g., educational level, text genre, digital devices). See the *effect size index* section below for details about the use of these subgroups. The final sample consisted of 76 media comparisons, each contributing an individual effect size. The meta-analysis is based on 171,055 participants. See Appendix (Table A1 and Table A2) for a detailed distribution of the participants among the studies.

---

[1] The study of media effects on reading comprehension has been the focus of several disciplines, including reading research, reading assessment, educational practice, media studies and learning technologies. Each discipline tends to use idiosyncratic words for similar, if not identical, scenarios. For example, the dependent variable in a situation where students read a text and answer comprehension questions is termed "test performance" in the assessment literature, but the term "comprehension scores" is used in the reading literature. Therefore, to avoid leaving out relevant studies from a particular field, we opted to include a broad range of search terms in our query.
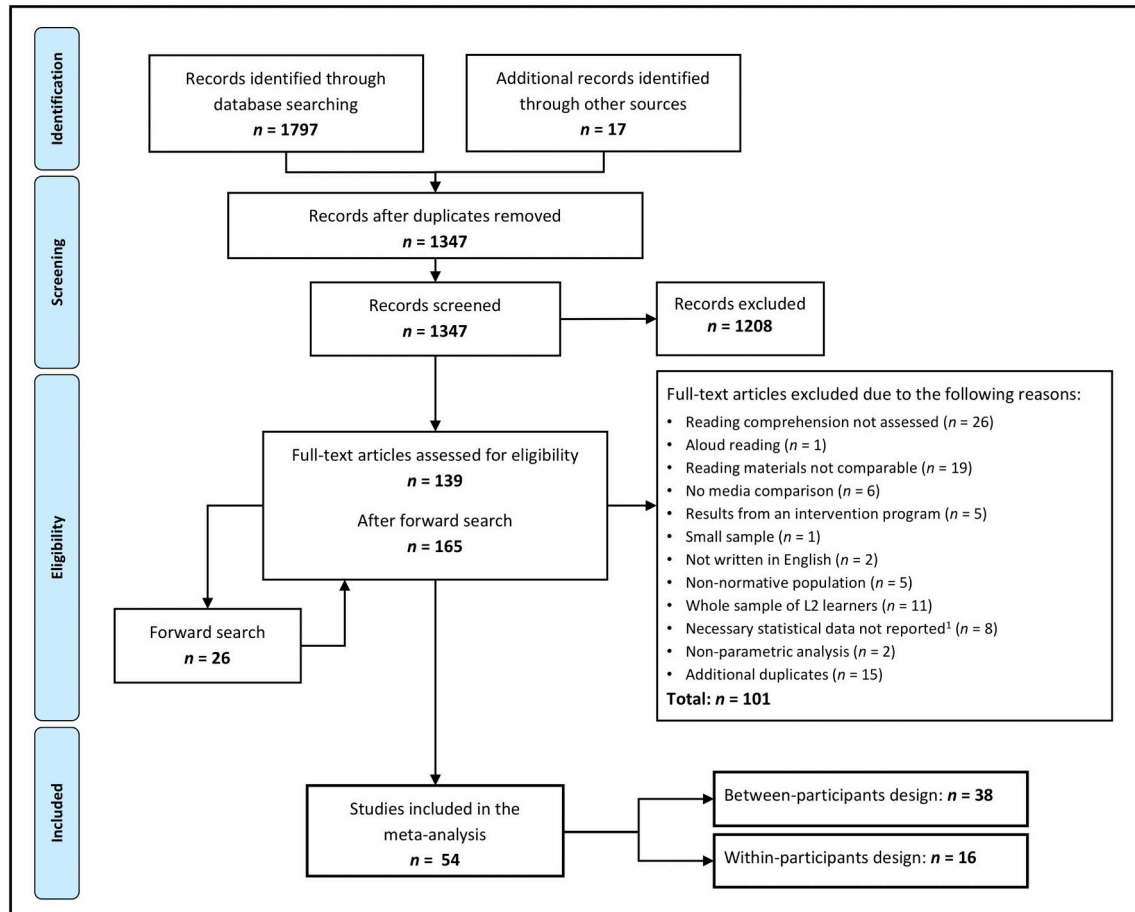
Note. [1]Not reported in the study report and not provided by authors following a personal request.

Fig. 2. Flowchart of the selection process.

2.3. Coding the studies

Several characteristics were coded for each comparison. This allowed for descriptive information and the consideration of moderating variables for the reported effect sizes. When necessary information was not included in the paper for a particular variable, it was coded as "Not reported" (N/r). When available, the following variables were coded:

Substantive variables:

1. *Participants' educational level:* elementary, middle or high school, undergraduates, or graduates and professionals.
2. *Text length:* number of words used in the reading task or other relevant information, such as the number of pages. Once coded, text length was categorized as (a) short (less than 1000 words) or (b) long
3. *Allowed reading time frame:* (a) free, when reading-time was self-paced by participants, or (b) limited, when time was restricted by experimental instructions.
4. *Type of digital device:* (a) computer (desktop or laptop) or (b) hand-held (tablet, e-reader, or smartphone).
5. *Text genre:* (a) informational, when texts were expository, descriptive or informative, (b) narrative, or (c) mixed, when both genre categories were used in the same task.
6. *Need for scrolling:* whether participants needed to scroll down the texts when reading in digital-based conditions. Coded as (a) yes or (b) no.
7. *Open testing:* whether participants could go back to texts when answering questions. Coded as (a) yes or (b) no.
8. *Type of comprehension:* (a) textual, when reading tasks asked for specific details or shallow level of comprehension; (b) inferential, high-level comprehension, when tasks required inferences based on parts of the texts, across parts, or involved previous knowledge; or (c) mixed, when tasks required both types of comprehension.
9. *Explicit strategy requirement:* whether participants were prompted or asked to implement a specific strategy in order to promote more in-depth reading, by means of selecting keywords, the use of highlighting or note-taking, or the use of reading strategies

promoted by the experimental instructions. Coded as (a) yes or (b) no.

Extrinsic variables:

10. *Publishing status:* (a) published paper, (b) official report, (c) master or PhD thesis, and (d) conference communication.
11. *Year of publication/presentation:* exact year.

Methodological variables:

12. *Sample size:* number of participants.
13. *Sampling method:* (a) probability (some process or procedure that ensures that the different units in the population have equal probabilities of being chosen) or (b) non-probability.
14. *Allocation of participants to media conditions:* (a) random, (b) quasi-random, (c) non-random but matched or controlled, (d) non-random and not controlled, and (e) within-participants design.
15. *Type of reading comprehension test:* (a) standardized/official test or (b) researcher-created task.
16. *Testing medium:* whether participants completed the comprehension test (a) on the same medium used for reading the texts, (b) always on paper, or (c) always on the digital device.

The coding process was conducted by two independent judges, based on a random sample (28%) of the studies included in the meta-analysis. Inter-rater reliability was adequate, showing a Cohen's kappa equal to .89 (minimum = .71, maximum = 1) for qualitative variables, and an intra-class correlation (95% CI) yielding absolute agreement for continuous variables (ICC = 1). Disagreements were discussed. For transparency and objectivity, a coding manual was developed and is available by request from the last author. A descriptive overview of the studies included is given in the *Results* section and in the Appendix (Table A1 and Table A2).

2.4. The effect size index

The effect size was calculated for each comparison, using means, standard deviations, and sample sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). 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 digital-based (treatment) and paper-based (control) groups' means on the post-test, divided by a pooled within-group standard deviation (Cohen, 1988). In addition, to estimate unbiased effect sizes, the correction factor for small sample sizes proposed by Hedges and Olkin (1985) was used. A positive Hedges' *g* indicates better comprehension results for the digital-based condition, whereas a negative Hedges' *g* indicates better outcomes for the paper-based condition.

For studies that used a within-participants design (each participant read on both paper and digital presentations), the standardized mean change index, d_c, was used to estimate the effect sizes. This effect size index is defined as subtracting the mean of the treatment group from the mean of the control group, and then dividing it by the standard deviation of the control group (Botella & Sánchez-Meca, 2015; Morris, 2000). In this case, in order to keep the interpretation of the direction of the mean effect size constant across both datasets (i.e., a positive value indicates better reading outcomes for the digital-based condition and vice versa), we used the digital-based condition as the control group. None of the studies reported the correlation coefficients, and thus, all values were imputed for a conservative estimate (*r* = .7), as recommended by Rosenthal (1991). As in the previous index, the correction factor for small sample sizes was applied to calculate this effect size index (Hedges & Olkin, 1985).

Finally, as indicated above, some studies reported multiple comparisons. In these cases, the following strategies were applied: a) when the study contained multiple between-participants treatments, the effect size for each subgroup was estimated; b) when there were multiple-treatment groups but they were dependent subgroups, effect sizes and their variances were combined into overall effect sizes and variances for these subgroups; c) if two digital-based 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); and d) when the study provided data on multiple outcome measures, effect sizes and variances were averaged to create a single effect size and allow statistical independence of the data (Lipsey & Wilson, 2001). In one case, a combination of strategies *b* and *c* had to be applied due to the existence of three digital-based reading groups.

2.5. Statistical analyses

Two separated meta-analyses were performed because it is not recommended to combine studies with between-participants and within-participants designs in one meta-analysis (Lipsey & Wilson, 2001). In each meta-analysis, a weighted mean effect size with its confidence interval (95%) was estimated, and a forest plot was made. Cochran's Q statistic was used to assess the presence of heterogeneity (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006), and I^2 index estimated the proportion of observed variance that is not due to sampling error. Furthermore, the prediction interval was calculated to provide additional context. A random-effects model was used to analyse effect sizes because it is generally regarded as more realistic (Borenstein et al., 2009; Borenstein, Higgins, Hedges, & Rothstein, 2017; Cooper, Hedges, & Valentine, 2009; Huedo-Medina et al., 2006).

Between-study heterogeneity was examined with ANOVAs for qualitative moderators and simple meta-regression for continuous moderators (Borenstein et al., 2009; Cooper et al., 2009), applying the adjustment proposed by Knapp and Hartung (2003). The

proportion of variance explained by moderators was estimated by the R^2 index (Raudenbush, 2009).

The normality assumption and outlier detection were assessed by examining the Q–Q normal plot, the Kolmogorov-Smirnov test with the Lilliefors correction, the Shapiro-Wilk test, and the standardized residuals (values greater than 3 in absolute magnitude were considered outliers). When potential outliers were identified, the robust model proposed by Beath (2014) was applied to confirm, removing effect sizes when a probability greater than .9 was found.

Sensitivity analysis was performed to evaluate the robustness of the results. The one-study-removal approach was used to evaluate the impact of each effect size on the mean estimate of the mean effect obtained (Borenstein et al., 2009). Moreover, when calculating the mean effect size for within-participants comparisons, due to the small number of effect sizes, additional methods were used to estimate τ^2 (in particular, the DerSimonian and Laird method with Knapp and Hartung adjustment, the maximum likelihood estimator, and the restricted maximum likelihood estimator). Finally, we also estimated the mean effect sizes, imputing different correlation coefficients (range of values from .10 to .90).

Publication bias was evaluated using Rosenthal's file drawer analysis (Rosenthal, 1979) and Egger's linear regression (Card, 2012), and applying ANOVA to compare the mean effect size of the published versus unpublished studies.

The statistical analyses were conducted using Comprehensive Meta-analysis software Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2014), R 3.1.1 software with Metafor (Viechtbauer, 2010) and Metaplus (Beath, 2015) packages, and a Microsoft Excel spreadsheet for computing prediction intervals.

3. Results

3.1. Descriptive characteristics of the studies

In the final sample ($n = 54$), 38 studies used a between-participants design. Of these 38 studies, 58 media comparisons (i.e., effect sizes) with 169,524 participants were initially included in the meta-analysis. Note that the majority of these participants (165,778) were from four large-scale studies (Eyre, Berg, Mazengarb, & Lawes, 2017; Lenhard et al., 2017; Pommerich, 2004; Puhan, Boughton, & Kim, 2005; see Appendix, Table A1). In addition, 16 studies used a within-participants design, providing 18 media comparisons with 1,531 participants. Within our dataset, two studies (Pommerich, 2004; Pomplun, Frey, & Becker, 2002) were included in both the Wang et al. (2007) and Kingston (2008) meta-analyses, mentioned above. Another study (Higgins, Russell, & Hoffman, 2005) was also included in Kingston's work. The remaining studies included in these two meta-analyses did not meet our inclusion criteria.

3.1.1. Between-participants studies

Focusing on the substantive variables described in the Appendix (Table A1), it is worth noting that the majority of the comparisons were conducted with undergraduate students (63.79%), used computers as digital devices (74.13%), included only informational texts (55.17%), and assessed comprehension by means of a mixture of textual and inferential questions (72.41%). In addition, in 44.83% of the comparisons, researchers imposed time constraints for reading the texts. Regarding extrinsic variables, 25 studies (39 effect sizes) were published papers, whereas the remaining 13 studies (17 effect sizes) included PhD dissertations ($n = 6$), a master thesis ($n = 1$), conference communications ($n = 4$), and an official report ($n = 1$). Moreover, an overview of the between-participants studies shows that 11 studies (16 effect sizes) were published or presented between 2000 and 2010, and 27 studies (42 effect sizes) between 2011 and 2017. Finally, regarding the methodological variables, 98.27% of the comparisons were from studies that recruited the sample through a non-probability sampling method, and 74.14% reported a randomized group allocation of participants. Researcher-created tasks were used in approximately 63.79% of the comparisons (see Appendix, Table A1, for additional information).

Finally, it is worth noting that several studies did not report information about some of the coded variables. However, they were included in the dataset whenever the information provided allowed us to calculate effect sizes because our purpose was to include a sample of studies in the meta-analysis that was as representative as possible.

3.1.2. Within-participants studies

The within-participants studies included are described in the Appendix (Table A2). Regarding substantive variables, a majority of the 18 comparisons reported that they were conducted with undergraduates (55.55%), used computers for digital-based reading (55.55%), used informational texts (61.11%), and assessed comprehension by means of a mixture of textual and inferential questions (55.55%). In relation to reading time, five comparisons imposed time constraints. Focusing on extrinsic variables, this dataset consisted of 11 published studies (13 effect sizes), a PhD dissertation, a bachelor thesis, and three conference communications (in all, 5 effect sizes from unpublished studies). Only four studies were reported before 2011. With regard to methodological variables, all the studies recruited the sample through a non-probability method, and eleven comparisons were conducted using researcher-created tasks.

3.2. The mean effect size, heterogeneity, and sensitivity analyses

Before calculating the mean effect size, preliminary analyses were conducted to identify outliers and verify normality of the sample. Two effect sizes were identified as possible outliers (Duran, 2013; Nishizaki, 2015; see Appendix, Table A1) by examining standardized residuals (values > 3), the Q–Q normal plot, and the Kolmogorov-Smirnov test with the Lilliefors correction ($p = .02$) in the between-participants dataset. The robust model was applied to further analyse these potential outliers, with both obtaining

	Hedges' g	Lower	Upper	p
Ackerman & Lauterman, 2012 (Exp. 1)a	-1,17	-1,82	-0,52	0,00
Ben-Yehudah & Eshet-Alkalai, 2014a	-0,80	-1,39	-0,21	0,01
Ackerman & Goldsmith, 2011 (Exp. 2)	-0,77	-1,23	-0,30	0,00
Nishizaki, 2015 (Exp. 1)b	-0,67	-1,29	-0,04	0,04
Beach, 2008b	-0,66	-1,26	-0,05	0,03
Nishizaki, 2015 (Exp 2)	-0,64	-1,08	-0,19	0,00
Chen et al., 2014a	-0,64	-1,26	-0,01	0,05
Ackerman & Lauterman, 2012 (Exp. 2)	-0,63	-1,09	-0,18	0,01
Jones et al., 2005	-0,58	-1,15	-0,00	0,05
Wästlund et al., 2005	-0,56	-1,03	-0,09	0,02
Norman & Furnes, 2016 (Exp. 2)	-0,52	-1,08	0,03	0,06
Mayes et al., 2001	-0,50	-1,12	0,12	0,11
Hongler, 2015b	-0,47	-1,15	0,22	0,18
Chen et al., 2014b	-0,44	-1,05	0,18	0,16
Mangen et al., 2013	-0,44	-0,92	0,05	0,08
Eyre et al., 2017a	-0,38	-0,44	-0,33	0,00
Eyre et al., 2017b	-0,38	-0,44	-0,32	0,00
Bartell et al., 2006	-0,37	-0,62	-0,11	0,00
Taylor, 2011b	-0,36	-1,01	0,30	0,28
Simian et al., 2016	-0,35	-0,76	0,07	0,10
Grimshaw et al., 2007a	-0,33	-0,87	0,22	0,24
Lenhard et al., 2017	-0,32	-0,40	-0,25	0,00
Green et al., 2010	-0,31	-0,84	0,22	0,25
Ben-Yehudah & Eshet-Alkalai, 2014b	-0,28	-0,84	0,29	0,33
Higgins et al., 2005b	-0,25	-0,64	0,14	0,21
Margolin et al., 2013b	-0,23	-0,73	0,28	0,38
Seehafer, 2014	-0,21	-0,69	0,26	0,38
Norman & Furnes, 2016 (Exp. 1)b	-0,21	-0,87	0,45	0,54
Lauterman & Ackerman, 2014 (Exp. 2)	-0,20	-0,64	0,25	0,39
Hongler, 2015a	-0,19	-0,87	0,49	0,59
Lauterman & Ackerman, 2014 (Exp. 1)	-0,16	-0,58	0,27	0,47
Pommerich, 2004 (Exp. 1)	-0,14	-0,23	-0,05	0,00
Norman & Furnes, 2016 (Exp. 1)a	-0,13	-0,80	0,54	0,70
Grimshaw et al., 2007b	-0,13	-0,65	0,39	0,63
Kaufman & Flanagan, 2016	-0,10	-0,55	0,34	0,64
Johnson, 2013	-0,06	-0,32	0,19	0,63
Higgins et al., 2005a	-0,05	-0,44	0,34	0,80
Daniel & Woody, 2013	-0,05	-0,55	0,46	0,86
Pommerich, 2004 (Exp. 2)a	-0,04	-0,14	0,07	0,50
Pommerich, 2004 (Exp. 2)b	-0,03	-0,13	0,08	0,58
Burkley, 2013	-0,02	-0,69	0,65	0,95
Beach, 2008a	-0,01	-0,77	0,75	0,98
Chen, 2015	0,00	-0,41	0,41	1,00
Niccoli, 2015	0,01	-0,24	0,27	0,91
Wells, 2012	0,02	-0,31	0,35	0,92
Ackerman & Goldsmith, 2011 (Exp. 1)	0,03	-0,44	0,49	0,91
Puhan et al., 2005	0,03	-0,06	0,11	0,51
Margolin et al., 2013a	0,04	-0,46	0,54	0,87
Porion et al., 2016	0,05	-0,41	0,51	0,83
Ackerman & Lauterman, 2012 (Exp. 1)b	0,06	-0,56	0,67	0,86
Morineau et al., 2005	0,08	-0,53	0,69	0,80
Connell et al., 2012b	0,11	-0,32	0,54	0,61
Connell et al., 2012a	0,11	-0,31	0,54	0,60
Taylor, 2011a	0,26	-0,40	0,92	0,13
Aydemir et al., 2013	0,39	-0,11	0,89	0,07
McCrea-Andrews, 2014	0,61	-0,04	1,27	0,07
Mean effect	**-0,21**	**-0,28**	**-0,14**	**0,00**

Note. Letters after the publication year differentiate several comparisons from the same study. Note that comparisons reported in the studies could have been recoded in the meta-analysis (see *Method* section). Please note that negative values indicate better outcomes for paper-based reading.

Fig. 3. Forest plot of reading media effect sizes on reading comprehension from studies using between-participants designs.

probabilities greater than 0.90. Therefore, they were removed from posterior analyses, and so the final sample of between-participants studies included 56 effect sizes. After removing outliers, effect sizes were normally distributed ($p = .40$).

When examining the within-participants dataset, no effect size was identified as an outlier, and so the initial 18 effect sizes were all included in the analysis. The Shapiro-Wilk normality test ($p = .52$) indicated that the dataset was normally distributed.
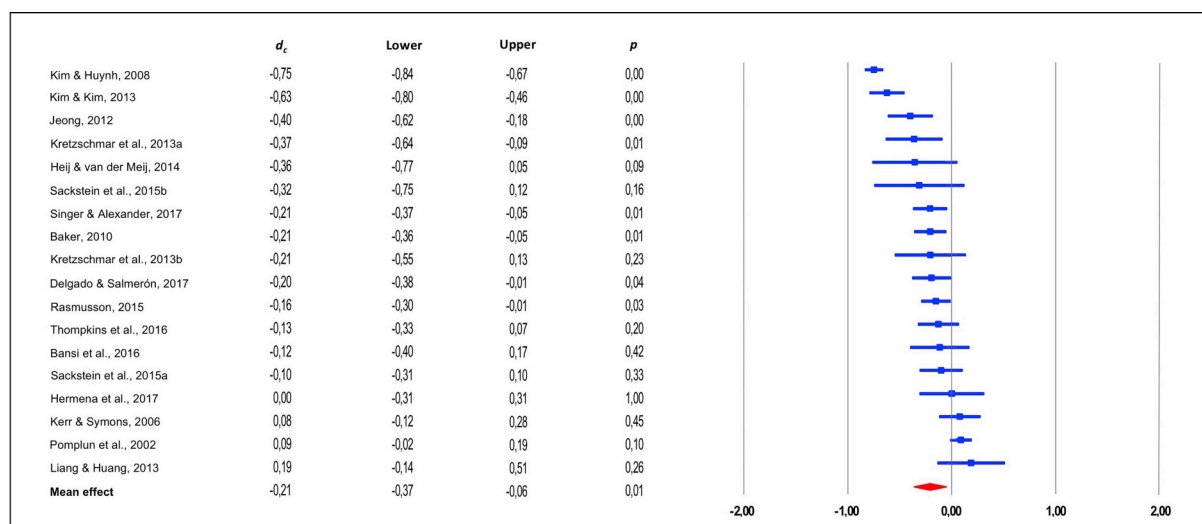
3.2.1. Media effect in between-participants designs

As explained above, comprehension in paper-based reading groups was used as the baseline. Therefore, negative values indicate that reading outcomes from digital-based devices were lower than their respective paper-based groups. The mean effect size of the sample was significant (Hedges' $g = -0.21$; 95% CI: -0.28, -0.14; $k = 56$), revealing an advantage of paper-based reading over digital-based reading. An overview of the effect sizes can be seen in Fig. 3, which provides a graphical representation of the estimated results of each reading media comparison. Each result is represented by a blue line with a dot in the centre. The dot indicates the value of the effect size (note the vertical lines marking values from -2 to 2), and the line that emerges from both sides of the dot represents the confidence interval. The longer the line, the larger the confidence interval. Lines that do not reach the zero value indicate significant effect sizes.

Regarding the variability of the effect sizes, the heterogeneity between individual effect sizes was medium-high ($I^2 = 72.24$) and statistically significant ($Q = 208.96, p < .001$). The prediction interval was -0.56 to 0.14, and so it was expected that the true effect size would fall in this range in 95% of all populations. Hence, the effects are large in some populations, but moderated and trivial in other populations. The wide range of effects calls for further analyses to examine potential moderating factors that would shed light on sources of differences among the studies. Thus, analyses were conducted to examine effects of substantive, extrinsic, and methodological variables. The results are reported below.

3.2.2. Sensitivity analyses for between-participants comparisons

The one-study-removal method (Borenstein et al., 2009) showed that effect sizes fell between Hedges' $g = -0.22$ and -0.20 ($p < .001$) and did not substantially affect the mean effect size, indicating a significant advantage of paper-based reading in all cases. Special attention should be paid to the four large-scale studies mentioned above (Eyre et al., 2017; Lenhard et al., 2017; Pommerich,

	d_c	Lower	Upper	p
Kim & Huynh, 2008	-0,75	-0,84	-0,67	0,00
Kim & Kim, 2013	-0,63	-0,80	-0,46	0,00
Jeong, 2012	-0,40	-0,62	-0,18	0,00
Kretzschmar et al., 2013a	-0,37	-0,64	-0,09	0,01
Heij & van der Meij, 2014	-0,36	-0,77	0,05	0,09
Sackstein et al., 2015b	-0,32	-0,75	0,12	0,16
Singer & Alexander, 2017	-0,21	-0,37	-0,05	0,01
Baker, 2010	-0,21	-0,36	-0,05	0,01
Kretzschmar et al., 2013b	-0,21	-0,55	0,13	0,23
Delgado & Salmerón, 2017	-0,20	-0,38	-0,01	0,04
Rasmusson, 2015	-0,16	-0,30	-0,01	0,03
Thompkins et al., 2016	-0,13	-0,33	0,07	0,20
Bansi et al., 2016	-0,12	-0,40	0,17	0,42
Sackstein et al., 2015a	-0,10	-0,31	0,10	0,33
Hermena et al., 2017	0,00	-0,31	0,31	1,00
Kerr & Symons, 2006	0,08	-0,12	0,28	0,45
Pomplun et al., 2002	0,09	-0,02	0,19	0,10
Liang & Huang, 2013	0,19	-0,14	0,51	0,26
Mean effect	-0,21	-0,37	-0,06	0,01

Note. Letters after the publication year differentiate several comparisons from the same study. Note that comparisons reported in the studies could have been recoded in the meta-analysis (see *Method* section). Please note that negative values indicate better outcomes for paper-based reading.

Fig. 4. Forest plot of reading media effect sizes on reading comprehension from studies using within-participants designs.

2004; Puhan et al., 2005). Given that their large samples yielded a small confidence interval for their effect sizes, their influence on the overall effect could skew the results. However, excluding these studies altogether (7 effect sizes), the mean effect size was Hedges' $g = -0.22$ ($p < .001$), which means they did not bias the overall effect of the reading media. Finally, given that we included "grey literature" (unpublished studies) in our meta-analysis, we repeated the meta-analysis without these studies in order to make sure that their inclusion does not compromise research quality. The mean effect size was Hedges' $g = -0.19$ (95% CI: -0.27, -0.11; $k = 38$) when excluding all the unpublished studies (i.e., official reports, conference communications, and dissertations) and Hedges' $g = -0.20$ (95% CI: -0.28, -0.13; $k = 51$) when only excluding the conference communications. Thus, "grey literature" did not substantially affect the overall mean effect size in this dataset.

3.2.3. Media effect in within-participants designs

The mean effect size of this sample of studies was also significant, and it replicated the advantage of paper-based reading over digital-based reading ($d_c = -0.21$; 95%; CI: -0.37, -0.06; $k = 18$). Fig. 4, similarly to Fig. 3, presents an overview of the effect sizes included in the dataset of studies that used a within-participants design.

As in between-participants studies, heterogeneity of the effect sizes was high ($I^2 = 89.88$; $Q = 167.94$, $p < .001$), with the prediction interval ranging from -0.90 to 0.47. Nevertheless, analyses of moderators were not performed in this dataset, due to the small number of effect sizes, and this should be taken into account when interpreting the results.

3.2.4. Sensitivity analyses for within-participants comparisons

One-study-removal analysis (Borenstein et al., 2009) indicated that effect sizes fell between $d_c = -0.18$ and -0.24 ($p < .001$), and were again significant, showing an advantage of paper-based reading in all cases. Additional results from Knapp and Hartung's adjustment of the DerSimonian and Laird estimator ($d_c = -0.21$; 95% CI: -0.33, -0.09; $k = 18$), the maximum likelihood approach ($d_c = -0.22$; 95% CI: -0.33, -0.10; $k = 18$), and the restricted maximum likelihood method ($d_c = -0.22$; 95% CI: -0.34, -0.10; $k = 18$) were also consistent. Moreover, a sensitivity analysis imputing different correlation coefficients (range of values from 0.10 to 0.90) was carried out. The findings were essentially identical (the largest difference between mean effect sizes was smaller than 3%) and revealed that the meta-analysis result was robust. Consequently, the result reported was based on a correlation of .70, as recommended by Rosenthal (1991). In addition, we also examined whether the inclusion of unpublished studies affected the overall effect of the reading media in this dataset. Thus, the mean effect size was $d_c = -0.22$ (95%; CI: -0.42, -0.13; $k = 13$) when excluding all the unpublished studies, and $d_c = -0.23$ (95%; CI: -0.41, -0.04; $k = 15$) when only excluding the conference communications. Therefore, "grey literature" did not affect the overall mean effect size in within-participants studies either.

3.3. Publication bias

3.3.1. Publication bias for between-participants comparisons

The risk of publication bias was examined with three different methods. First, results from Classic fail safe-N analysis indicated that 1,727 null effect sizes would be necessary to nullify the mean effect size of the medium. This value meets Rosenthal's criterion (5k + 10), which sets 290 as the minimum for this dataset. Second, Egger's linear regression indicated a non-significant publication bias ($p = .39$). Finally, an ANOVA revealed that the mean effect sizes from published versus unpublished studies were not statistically

Table 1

One-way analysis of variance of substantive variables on mean effect sizes for reading media from the studies using between-participants designs.

Variable[1]	Categories	k	Mean effect sizes		Q_B(df)	Q_W(df)	R^2
			Hedges' g	95% CI			
Participants' educational level[2]					2.33(2)	131.33(49)***	.00
	Grades 1 to 6	8	-0.19	[-0.35, -0.03]			
	Grades 7 to 12	8	-0.15	[-0.29, -0.02]			
	Undergraduates	36	-0.28	[-0.38, -0.18]			
Text length					0.14(1)	142.36(47)***	.00
	Short	22	-0.25	[-0.34, -0.16]			
	Long	26	-0.22	[-0.33, -0.11]			
Allowed reading time frame					4.12(1)*	185.17(45)***	.05
	Self-paced	20	-0.09	[-0.22, 0.05]			
	Limited	27	-0.26	[-0.35, -0.16]			
Digital device					1.55(1)	194.95(54)***	.02
	Computer	42	-0.23	[-0.31, -0.15]			
	Hand-held	14	-0.12	[-0.27, 0.03]			
Text genre					7.00(2)*	74.21(48)**	.31
	Informational	34	-0.27	[-0.36, -0.18]			
	Narrative	7	0.01	[-0.20, 0.20]			
	Mixed	10	-0.30	[-0.40, -0.21]			
Need for scrolling					1.99(1)	133.40(47)***	.00
	No	12	-0.13	[-0.27, 0.01]			
	Yes	37	-0.25	[-0.33, -0.16]			
Open testing					1.21(1)	183.46(47)***	.00
	No	33	-0.26	[-0.37, -0.16]			
	Yes	16	-0.18	[-0.29, -0.07]			
Type of comprehension[3]					0.14(1)	153.99(51)	.00
	Textual	9	-0.26	[-0.47, -0.04]			
	Mixed + Inferential	44	-0.21	[-0.29, -0.14]			

Note. k: number of effect sizes. Hedges' g: mean effect size. Q_B: between-categories Q statistic. Q_W: within-categories Q statistic. R^2: Proportion of total between-comparison variance explained. [1]Non-reported values for each variable were not included in these analyses. [2]Due to the small number of effect sizes, the category "Graduates or professionals" ($k = 3$) was not included in this analysis. [3]Due to the small number of effect sizes, comparisons that examined only inferential comprehension ($k = 3$) were included in the same group as those that examined both types of comprehension. *$p < .05$. **$p < .01$. ***$p < .001$.

different (Q_B (1, 54) = 0.14, $p = .71$). All these results suggested that there was no publication bias.

3.3.2. Publication bias for within-participants comparisons

In this dataset, Classic fail Safe-N analysis indicated that 475 null effect sizes would be necessary to nullify the mean effect size of the media, which again was a higher value than Rosenthal's criterion ($5k + 10 = 100$). Additionally, Egger's linear regression yielded a non-significant publication bias ($p = .20$), and an ANOVA between published and unpublished studies showed no significant differences (Q_B (1, 16) = 0.02, $p = .90$. Likewise, these three indicators suggested no risk of publication bias.

3.4. Moderating variables in between-participants comparisons

In the following analyses, we considered potential moderating variables, grouped by substantive, extrinsic, and methodological variables, for media effects on reading outcomes among the between-participants studies. As mentioned above, some studies lacked the necessary information about some of these variables, and so they were not included in the respective moderator analyses.

3.4.1. Substantive variables

We conducted an ANOVA for each substantive variable considered. These analyses indicated significant moderating effects of the allowed reading time frame (i.e., limited by task constraints vs. self-paced by participants) and text genre (i.e., informational texts vs. narrative texts vs. a combination of both genres). No moderating effects were found for educational level, text length, type of digital device, need for scrolling, open testing, or type of comprehension because Q_B values were not significant in all these cases (see Table 1). Examination of the reading time frame showed that comparisons in studies with time constraints yielded a significantly larger ($Q_B = 4.12$, $p = .04$) print advantage (Hedges' $g = -0.26$) than comparisons in studies in which participants were allowed to self-pace their reading (Hedges' $g = -0.09$). Thus, although there is an overall advantage of print over digital devices, the difference is larger with time constraints than with self-paced reading, which explains 5% of the mean effect size variance.

The moderator factor of text genre revealed a significant effect, explaining 31% of the mean effect size variance. Comparisons conducted with informational texts or a combination of informational and narrative texts showed significant mean effect sizes favouring paper-based reading over digital-based reading (Hedge's $g = -0.27$ and -0.30, respectively), whereas comparisons conducted only with narrative texts showed no effect of media (Hedge's $g = 0.01$) (see Table 1).

Table 2

One-way analysis of variance of moderating effect of extrinsic and methodological variables on mean effect sizes for reading media from the studies using between-participants designs.

Variable[1]	Categories	k	Mean effect sizes		Q_B(df)	Q_W(df)	R[2]
			Hedges' g	95% CI			
Publishing status					0.14(1)	186.47(54)***	.00
	Published	39	-0.22	[-0.31, -0.13]			
	Unpublished	17	-0.19	[-0.31, -0.07]			
Group allocation[2]					0.90(1)	167.33(49)***	.00
	Random	44	-0.20	[-0.28, -0.12]			
	Non-random	7	-0.28	[-0.46, -0.12]			
Type of reading comprehension test					0.01(1)	200.15(54)***	.00
	Standard./official	22	-0.21	[-0.31, -0.11]			
	Researcher-created	34	-0.21	[-0.32, -0.11]			
Testing medium					1.11(1)	180.06(45)***	.00
	Same for reading	27	-0.26	[-0.35, -0.17]			
	Always on paper	20	-0.17	[-0.31, -0.03]			

Note. k: number of effect sizes. Hedges' g: mean effect size. Q_B: between-categories Q statistic. Q_W: within-categories Q statistic. R[2]: Proportion of total between-comparison variance explained. [1]The variable sampling method was not included in the analyses due to lack of variability. [2]Due to the small number of effect sizes categories "Non-random but controlled" (k = 3) and "Non-random not controlled" (k = 4) were combined ("Non-random"). ***$p < .001$.

Two variables are worth mentioning, even though their moderating effects did not reach significance. The advantage of paper-based reading was significant when studies used computers (Hedges' $g = -0.23$, $p < .001$), but not when they used hand-held devices (Hedges' $g = -0.12$, $p = .11$). Similarly, the need for scrolling as a feature of digital-based reading resulted in a significant advantage of paper-based reading (Hedges' g = -0.25, $p < .001$), whereas the media effect was marginal and numerically smaller when scrolling was not necessary (Hedges' $g = -0.13$, $p = .06$) (see Table 1).

Finally, due to the small number of comparisons where in-depth reading was prompted by means of an explicit strategic requirement (k = 5), the moderating effect of this variable was not examined.

3.4.2. Extrinsic variables

As reported above, the ANOVA with publishing status was not a significant moderator, as indicated by the Q_B value (see Table 2). However, a meta-regression analysis revealed that the date of publication or presentation of the studies has a significant moderating effect on the mean effect size of the media. The advantage of paper-based reading over digital-based reading increased since 2000, as hypothesised in the right panel of Fig. 1. The beta coefficient of -0.01 ($Q_R = 4.95$, $p = .03$) indicates that the effect size favouring paper-based reading increased by 0.01 points a year, explaining 64% of the mean effect size variance (see Table 3).

3.4.3. Methodological variables

Four methodological variables were tested to examine their possible influence on the media effect. They were sample size, method of allocating participants to media conditions, the type of reading comprehension test, and the testing medium. Results revealed that none of these four methodological variables had a significant moderating effect, as indicated by the Q_B and Q_R values (see Tables 2 and 3). The sampling method variable was not analysed due to lack of variability (See Appendix, Table A1).

4. Discussion

This study sought to address an issue of great importance in education and work-related contexts, namely, whether and under what conditions media have an effect on reading comprehension. The strong appeal of digital-based assessment and learning environments has led many educational systems to adopt them. As findings from the current work reveal, however, digital environments may not always be best suited to fostering deep comprehension and learning. The straightforward conclusion is that providing students with printed texts despite the appeal of computerized study environments might be an effective direction for improving

Table 3

Meta-regression analysis of moderating effect of sample size and date of publication on mean effect sizes for reading media from the studies using between-participants designs.

Variable	k	b	Q_R	Q_E	R[2]
Sample size	56	-0.00	3.11	201.59***	.42
Date of publication	56	-0.01	4.95*	201.59***	.64

Note. k: number of effect sizes. b: unstandardized regression coefficient. Q_R: statistical test of between-comparison effects. Q_E: statistical test of between-comparison homogeneity of the effect sizes. R[2]: Proportion of total between-comparison variance explained. *$p < .05$. ***$p < .001$.

comprehension outcomes. However, given the unavoidable inclusion of digital devices in our contemporary educational systems, more work must be done to train pupils on dealing with performing reading tasks in digital media, as well as to understand how to develop effective digital learning environments.

The results of the two meta-analyses in the present study yield a clear picture of screen inferiority, with lower reading comprehension outcomes for digital texts compared to printed texts, which corroborates and extends previous research (Kong et al., 2018; Singer & Alexander, 2017b; Wang et al., 2007). These results were consistent across methodologies and theoretical frameworks.

Although the effect sizes found for media (-0.21) are small according to Cohen's guidelines (1988), it is important to interpret this effect size in the context of reading comprehension studies. During elementary school, it is estimated that yearly growth in reading comprehension is 0.32 (ranging from 0.55 in grade 1, to 0.08 in grade 6) (Luyten, Merrel, & Tymms, 2017). Intervention studies on reading comprehension yield a mean effect of 0.45 (Scammacca, Roberts, Vaughn, & Stuebing, 2015). Thus, the effects of media are relevant in the educational context because they represent approximately 2/3 of the yearly growth in comprehension in elementary school, and 1/2 of the effect of remedial interventions.

Our investigation of moderating factors indicated that the advantage of paper-based reading is significantly larger when a reading time limit is imposed, compared to self-paced reading. Such advantage is consistent across studies using informational texts (or a mix of informational and narrative), but no media effect is found when the studies used only narrative texts. In addition, the advantage of print reading significantly increased from 2000 to 2017. Furthermore, although they did not reach significance, the results suggest stronger media differences on computers than on hand-held devices, as well as disadvantages of digital texts that require scrolling. Finally, the results indicate that media differences do not vary according to the remaining substantive factors: age group (educational level), text length, type of comprehension assessed, or the option to revise the text to answer the questions; extrinsic factors: sample size and publishing status; or methodological factors: type of test, group allocation, and testing medium.

We discuss below the implications of the findings. In particular, how the screen inferiority effect is related to the reading practices of new generations, to theories of self-regulated learning, and to the genre of the reading materials. We then identify some of the limitations of the study and conclude by discussing several educational implications of our results.

4.1. Media effect and new generations

The adoption of new media practices often involves activating a set of cognitive processes appropriate for taking full advantage of the media. For children growing up surrounded by digital technologies, skills such as the ability to search and navigate, read critically, and multitask are essential (e.g. Salmerón, García, & Vidal-Abarca, 2018). Such skills place demands on attention and executive processes that may not be fully developed in children and adults reading digital texts. If simply being exposed to digital technologies were enough to gain these skills, then we would expect an increasing advantage of digital reading, or at least decreasing screen inferiority over the years. Contrary to this assumption, however, our results indicate that the screen inferiority effect has increased in the past 18 years, and that there were no differences in media effects between age groups. These surprising findings suggest that we cannot idly wait for screen inferiority to disappear as children are exposed to digital devices earlier and earlier in their lives, as adults gain more experience with the technology, or as technology improves. The data suggest that screen inferiority is a major challenge across age groups that becomes more severe as the presence of technology increases.

4.2. Media effect and time frames for learning

Our results do not address the cause of this persistent screen inferiority, but they provide evidence that people adopt a shallower processing style in digital environments (e.g. Lauterman & Ackerman, 2014; Wolf & Barzillai, 2009). The increase in media differences as technology becomes more integrated into our lives may be related to poorer quality of attention (Courage, 2017), where deep immersion in the text is challenged (e.g. Mangen & Kuiken, 2014). The *Shallowing Hypothesis* suggests that because the use of most digital media consists of quick interactions driven by immediate rewards (e.g. number of "likes" of a post), readers using digital devices may find it difficult to engage in challenging tasks, such as reading comprehension, requiring sustained attention (Annisette & Lafreniere, 2017). According to this perspective, the more people use digital media for these shallow interactions, the less they will be able to use them for challenging tasks. Such arguments are consistent with negative correlations reported between the frequency of digital media use and text comprehension in adolescents (Duncan et al., 2015; Pfost et al., 2013), and they suggest that we should be cautious about the introduction of digital reading in classrooms.

A relevant moderator found for the screen inferiority effect was time frame. This finding sheds new light on the mixed results in the existing literature. Consistent with the findings by Ackerman and Lauterman (2012) with lengthy texts, mentioned above, Sidi et al. (2017) found that even when performing tasks involving reading only brief texts and no scrolling (solving challenging logic problems presented in an average of 77 words), digital-based environments harm performance under time pressure conditions, but not under a loose time frame. In addition, they found a similar screen inferiority when solving problems under time pressure and under free time allocation, but framing the task as preliminary rather than central. Thus, the harmful effect of limited time on digital-based work is not limited to reading lengthy texts. Moreover, consistently across studies, Ackerman and colleagues found that people suffer from greater overconfidence in digital-based reading than in paper-based reading under these conditions that warrant shallow processing. Sidi et al. (2017) explained that time pressure and framing the task as preliminary both justify shallow processing, which has a stronger effect in digital environments where people are used to quick and shallow tasks (e.g., Facebook, chats; see also Lauterman & Ackerman, 2014). These empirical findings support Annisette and Lafreniere's (2017) Shallowing Hypothesis, which had previously been based on self-reports.

Our findings call to extend existing theories about self-regulated learning (see Boekaerts, 2017, for a review). Effects of time frames on self-regulated learning have been discussed from various theoretical approaches. First, a metacognitive explanation suggests that time pressure encourages compromise in reaching learning objectives (Thiede & Dunlosky, 1999). Second, time pressure has been associated with cognitive load. Some studies found that time pressure increased cognitive load and harmed performance (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). However, others suggested that it can generate a germane ("good") cognitive load by increasing task engagement (Gerjets & Scheiter, 2003). In these theoretical discussions, the potential effect of the medium in which the study is conducted has been overlooked. We see the robust finding in the present meta-analyses about the interaction between the time frame and the medium as a call to theorists to integrate the processing style adapted by learners in specific study environments into their theories.

The finding in this meta-analysis that most media effects come from tasks performed under limited time frames should be taken into account by designers of admission exams and educators. The disadvantage of digital-based reading would be especially critical if not all the examinees are tested in the same medium. Moreover, this could be also an influential factor even when they are all examined by means of digital tests, because of individual differences in adapting to the digital media. For instance, Lauterman and Ackerman (2014) found differences in media effects on learning outcomes based on people's media preference. Clearly, additional individual difference should be considered. Thus, digital exams outcomes probably reflect not only the knowledge or skill at hand, but also such digital-specific competencies.

An encouraging finding from Lauterman and Ackerman (2014) and Sidi et al. (2017) is that simple methodologies (e.g., writing keywords summarizing the text, framing the task as central) that engage people in in-depth processing make it possible to eliminate screen inferiority, in terms of both performance and overconfidence, even under a limited time frame. Together, these findings strongly suggest that pedagogy should play a significant role in identifying individual differences and guiding students to develop skills they miss that support a thoughtful approach to digital information, even when the task design seems to indicate the legitimacy of shallow processing.

4.3. Media effect and text genre

The text genre was another variable that moderated media effects. On the one hand, the paper-based reading advantage was consistent across studies using informational texts, or a mix of informational and narrative texts. On the other hand, studies using only narrative texts showed no effect of media on comprehension. Comprehending informational texts, compared to narratives, requires higher level processing, such as using complex academic vocabulary and structures, and these texts are less connected to real world knowledge, which makes them harder to comprehend (Graesser & McNamara, 2011). Thus, our finding may also point to the Shallowing Hypothesis as an explanation. Nevertheless, this result must be interpreted with caution due to the small number of comparisons that used only narrative texts. In addition, among the included studies that directly compared text genre and reading medium, only Simian et al. (2016) reported a significant interaction between these variables, revealing a positive effect of print-based reading only on informational texts, whereas two studies found no effect of text genre (Margolin, Driscoll, Toland, & Kegler, 2013; Rasmusson, 2015).

4.4. Additional potential moderators of media effects

Future research should aim to identify other variables that may interact with media effects. In particular, moderators with effects that approached significance deserve further consideration (see Table 1), such as the influence of the type of device. It is important to determine whether screen inferiority is limited to desktop computers and eliminated when using hand-held devices. If this proves to be the case, it would be important to understand what cognitive processes could allow media equivalence on hand-held devices. Of the three studies included in this meta-analysis that specifically examined differences among digital devices (Chen, Cheng, Chang, Zheng, & Huang, 2014; Hongler, 2015; Margolin et al., 2013), only Chen et al. (2014) found an interaction with media, reporting a negative impact of digital reading only on computers.

In addition, the need for scrolling was found to be a possible obstacle to comprehension during digital reading. Among the studies included in the meta-analysis, Pommerich (2004) and Higgins et al. (2005) found that participants who read non-scrolling digital texts outperformed those who read scrolling texts, although the differences were not significant. These studies, however, were performed more than a decade ago. Nonetheless, scrolling may add a cognitive load to the reading task by making spatial orientation to the text more difficult for readers than learning from printed text. One of the questions about the scrolling findings is whether the effect of scrolling is related to longer texts or some other artefact of mouse use while reading, although text length was not found to be a moderating factor in our meta-analyses.

4.5. Limitations

We would like to call attention to some limitations in our meta-analyses. First, ten studies that met the inclusion criteria could not be included due to lack of necessary statistical data ($n = 8$) or non-normal distributions ($n = 2$).

Moreover, the effect sizes included in the meta-analyses showed high heterogeneity. The moderators considered captured some of this variance, but there is clearly unexplained variance. Consequently, additional factors potentially influencing the results could be affecting the mean effect size. In particular, factors related to research methods (e.g., the reliability of the testing tools) or to sample characteristics (e.g., SES or degree of use of digital texts for learning purposes) could be considered. These factors were missing from

most of the reports we included in our meta-analyses. Therefore, we encourage researchers to investigate these possible moderators and describe their methods and samples in detail in future publications.

In addition, the interpretation of how the effect of reading media changes over generations was based on the studies' publication dates. Clearly, using the date as indicator of generation is simplistic and may affect several aspects (e.g., research methods may change throughout the years). In particular, we considered it relevant to examine how different age groups interact with the publication date. However, the distribution of age groups over the years was not broad enough to allow reliable analysis of this possible effect in our dataset. Thus, we recommend considering how different factors interact with the year of publication.

Finally, given that our purpose was to isolate the effect of media, per se, on reading outcomes, we excluded digital affordances (except for scrolling) such as hypertext reading or navigation through webpages. Their effect on reading comprehension is still an open question that warrants further research efforts.

5. Conclusions

In conclusion, it is clear that digital-based reading is an unavoidable part of our daily lives and an integral part of the educational realm. Although the current results suggest that paper-based reading should be favoured over digital-based reading, it is unrealistic to recommend avoiding digital devices. Nevertheless, ignoring the evidence of a robust screen inferiority effect may mislead political and educational decisions, and even worse, it could prevent readers from fully benefiting from their reading comprehension abilities and keep children from developing these skills in the first place. Thus, we call on researchers to consider how to guide students and exam takers in dealing with digital tasks such as admission tests (e.g., SAT and GMAT), tasks in work contexts, and school-related tasks that are very often performed with informational texts and under limited time frames. In particular, an important conclusion from our analysis is that there are predictable conditions that seem to allow media equivalence. It is important to appreciate these conditions, examine their validity for the task at hand, and use them whenever possible and relevant. We hope our meta-analysis will guide evidence-based decisions by policy makers and point designers and researchers toward conditions that support effective digital-based reading.

Acknowledgements

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2018.09.003.

References

The complete list of the references included in the meta-analysis can be found in the Appendix

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied, 17*, 18–32. https://doi.org/10.1037/a0022086.

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*, 1816–1828. https://doi.org/10.1016/j.chb.2012.04.023.

American Speech-Language-Hearing Association (2015). *Parent pooling: Better hearing and speech month.* http://www.asha.org/uploadedFiles/BHSM-Parent-Poll.pdf.

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.

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

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 570–585. https://doi.org/10.1037/0278-7393.33.3.570.

Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods, 5*, 285–293. https://doi.org/10.1002/jrsm.1114.

Beath, K. J. (2015). *Metaplus: Robust meta-analysis and meta-regression, 2016.* R package version 0.7-7.

Boekaerts, M. (2017). Cognitive load and self-regulation: Attempts to build a bridge. *Learning and Instruction, 51*, 90–97. https://doi.org/10.1016/j.learninstruc.2017.07.001.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2014). *Comprehensive meta-analysis version 3.* Englewood, NJ: Biostat.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*, 5–18. https://doi.org/10.1002/jrsm.1230.

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

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

Chen, G., Cheng, W., Chang, T., Zheng, X., & Huang, R. (2014). A comparison of reading comprehension across paper, computer screens, and tablets: Does tablet familiarity matter? *Journal of Computers in Education, 1*, 213–225. https://doi.org/10.1007/s40692-014-0012-z.

Childwise (2017). *The monitor trends report 2017* http://www.childwise.co.uk/reports.html#trendsreport.

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

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

Courage, M. L. (2017). Screen media and the youngest viewers: Implications for attention and learning. In F. C. Blumberg, & P. C. Brooks (Eds.). *Cognitive development in digital contexts* (pp. 3–28). London: Academic Press.

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

Duncan, L. G., McGeown, S. P., Griffiths, Y. M., Stothard, S. E., & Dobai, A. (2015). Adolescent reading skill and engagement with digital and traditional literacies as predictors of reading comprehension. *British Journal of Psychology, 107*, 209–238. https://doi.org/10.1111/bjop.12134.

Duran, E. (2013). Efficiency in reading comprehension: A comparison of students' competency in reading printed and digital texts. *Educational Research and Reviews, 8*(6), 258–269. https://doi.org/10.5897/ERR12.180.

Eyre, J., Berg, M., Mazengarb, J., & Lawes, E. (2017). *Mode equivalency in PAT: Reading comprehension.* Wellington: NZCER. http://www.nzcer.org.nz/system/files/ PAT%20Modes_report.pdf.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*, 33–41. https://doi.org/10.1207/S15326985EP3801_5.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3*, 371–398.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York, NY: Academic Press.

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions.* The Cochrane Collaboration: Version 5.1.0. [updated March 2011]. Available from http://www.cochrane-handbook.org.

Higgins, J., Russell, M., & Hoffmann, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning and Assessment, 3*(4)https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1657.

Hongler, K. I. (2015). *Superiority of paper as text presentation medium for effective and efficient learning – is it just an illusion?*Master thesishttp://www.mmi-basel.ch/wp-content/uploads/2016/04/hongler_masterarbeit.pdf.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*, 193–206. https://doi.org/10.1037/1082-989X.11.2.193.

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.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*, 2693–2710. https://doi.org/10.1002/sim.1482.

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.

Koren, D., Seidman, L. J., Goldsmith, M., & Harvey, P. D. (2006). Real-world cognitive—and metacognitive—dysfunction in schizophrenia: A new approach for measuring (and remediating) more "right stuff". *Schizophrenia Bulletin, 32*, 310–326. https://doi.org/10.1093/schbul/sbj035.

Kurata, K., Ishita, E., Miyata, Y., & Minami, Y. (2017). Print or digital? Reading behavior and preferences in Japan. *Journal of the Association for Information Science and Technology, 68*, 884–894. https://doi.org/10.1002/asi.23712.

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior, 35*, 455–463. https://doi.org/10.1016/j.chb.2014.02.046.

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Luyten, H., Merrell, C., & Tymms, P. (2017). The contribution of schooling to learning gains of pupils in Years 1 to 6. *School Effectiveness and School Improvement, Vol. 28*, 374–405. https://doi.org/10.1080/09243453.2017.1297312.

Mangen, A., & Kuiken, D. (2014). Lost in an iPad: Narrative engagement on paper and tablet. *Scientific Study of Literature, 4*, 150–177. https://doi.org/10.1075/ssol.4.2.02man.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research, 58*, 61–68. https://doi.org/10.1016/j.ijer.2012.12.002.

Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology, 27*, 512–519. https://doi.org/10.1002/acp.2930.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2.

Mizrachi, D. (2015). Undergraduates' academic reading format preferences and behaviors. *The Journal of Academic Librarianship, 41*, 301–311. https://doi.org/10.1016/j.acalib.2015.03.009.

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology, 53*, 17–29. https://doi.org/10.1348/000711000159150.

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

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.

Pansky, A., Goldsmith, M., Koriat, A., & Pearlman-Avnion, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology, 21*, 303–329. https://doi.org/10.1080/09541440802281183.

Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences, 26*, 89–102. https://doi.org/10.1016/j.lindif.2013.04.008.

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology, Learning, and Assessment, 2.* n6 https://eric.ed.gov/?id=EJ905028.

Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*, 337–354. https://doi.org/10.1177/0013164402062002009.

Puhan, G., Boughton, K. A., & Kim, S. (2005). *Evaluating the comparability of paper-and-pencil and computerized versions of a large-scale certification test*https://doi.org/10.1002/j.2333-8504.2005.tb01998.x ETS Research Report Series, 2005.

Rasmusson, M. (2015). Reading paper – reading screen. *Nordic Studies in Education, 35*, 3. http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-23417.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.). *The handbook of research synthesis and meta-analysis* (pp. 295–316). New York: Russell Sage.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641. https://doi.org/10.1037/0033-2909.86.3.638.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Salmerón, L., García, A., & Vidal-Abarca, E. (2018). The development of adolescents' comprehension-based Internet reading skills. *Learning and Individual Differences, 61*, 31–39. https://doi.org/10.1016/j.lindif.2017.11.006.

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

Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction, 51*, 61–73 doi: 1016/j.learninstruc.2017.01.002.

Simian, M., Malbrán, C., Fonseca, L., Grinberg, S., Gattas, M. J., Tascón, R., et al. (2016). Reading narrative, expository and discontinuous texts on paper versus screen: Impact on reading comprehension in Argentine high school kids. *Poster presented at 33rd annual meeting of the society for the scientific studies of reading (SSSR), porto, Portugal*.

Singer, L. M., & Alexander, P. A. (2017a). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*, 155–172. https://doi.org/10.1080/00220973.2016.1143794.

Singer, L. M., & Alexander, P. A. (2017b). 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.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024.

Viechtbauer, W. (2010). *"metafor: Meta-Analysis Package for R." R package version 1.4-0*.

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.

Wolf, M., & Barzillai, M. (2009). The importance of deep reading. *Educational Leadership, 6*, 32–37. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.7284&rep=rep1&type=pdf.