The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure

Pablo Delgado , Ladislao Salmerón [*]

ERI Lectura, Dept. of Developmental and Educational Psychology, University of Valencia, Av. Blasco Ibáñez, 21. 46010, Valencia, Spain

ARTICLE INFO

ABSTRACT

This study explored the influence of reading media and reading time-frame on readers' on-task attention, metacognitive calibration, and reading comprehension. One hundred and forty undergraduates were allocated to one of four experimental conditions varying on the reading medium (in print vs. on screen) and on the reading time-frame (free vs. pressured time). Readers' mindwandering while reading, prediction of performance on a comprehension test, and their text comprehension were measured. In-print readers, but not on-screen readers, mindwandered less on the pressured than in the free time condition, indicating higher task adaptation in print. Accordingly, on-screen readers in the pressured condition comprehended less than the other three groups. Mindwandering and text comprehension were similar under free reading time regardless of medium. Lastly, there were no differences in readers' metacognitive calibration. The results support the hypothesis of shallow information processing when reading on screen under time constraints.

1. Introduction

Major concerns about the utility of digital technologies in education have grown as their use becomes more and more pervasive. Scholars from different disciplines are bringing up their worries about their potential harming impact on human cognition, with especial emphasis on students' in-depth information processing and sustained attention capacity (e.g, Baron, 2015; Salmerón & Delgado, 2019; Wolf, 2018). Several empirical studies have reported that the use of digital technologies at school can lead to negative learning outcomes. In light of these considerations, it seems reasonable to deem digital technologies as not always suitable for academic reading and learning.

The conclusions of three recent meta-analyses on the medium effect on reading comprehension should be a matter of concern (Clinton, 2019; Delgado, Vargas, Ackerman, & Salmerón, 2018). Results demonstrated that people comprehend less the same texts on screen than on paper. While the overall effect sizes found by these studies were Hedges' g = -.21 (Delgado et al., 2018) and -.25 (Clinton, 2019), analyses of moderators identified three main qualifying factors. First, both Clinton (2019) and Delgado et al. (2018) found the on-screen inferiority to be clear in expository but not narrative texts, with $g = -0.32$ (vs. $g = -0.04$) and $g = -0.27$ (vs. $g = 0.01$), respectively. Second, the effect was significant only among studies in which participants read under time

constraints (g = -.26; Delgado et al., 2018). Finally, the effect of generation may also play a role, as the medium effect increased 0.01 points each year from 2001 to 2017 (i.e., the more recent the studies, the larger the on-screen inferiority; Delgado et al., 2018). Although from the classical approach of Cohen (1988) such effects are small, educational researchers have recently emphasized the need to interpret effects sizes in context (Funder & Ozer, 2019). Thus, as Delgado et al. (2018) argued, an effect size ranging from −0.21 to −0.32 is relevant in the reading comprehension field because it represents approximately 2/3 of the yearly growth in reading comprehension during elementary school (Luyten, Merrel, & Tymms, 2017).

The fact that the on-screen inferiority particularly emerges in expository texts and that it increases under time constraints suggests that such effect arises in cognitively demanding tasks. Although literary texts can be highly complex and difficult to fully understand, comprehension of expository texts (vs. narrative) is generally considered to demand increased cognitive efforts, as they present academic knowledge usually by means of a large number of ideas, infrequent vocabulary and complex text structures. For example, linguistic analyses of a random selection of 200 narrative and science seventh grade texts from the TASA corpus found that narrative texts use more frequent words, more concrete nouns and verbs, more connectives, or higher causal cohesion (Graesser & McNamara, 2011). Moreover, an increased

* Corresponding author.
 E-mail addresses: pablo.delgado@uv.es (P. Delgado), ladislao.salmeron@uv.es (L. Salmerón).

efficiency is required when performing tasks under limited time. In such cases, in-depth processing in combination with time management becomes critical (Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014). Thus, Delgado et al. (2018) pointed out to the *shallowing hypothesis* as an explanation for on-screen inferiority (Annisette & Lafreniere, 2017). This hypothesis considers that the daily, massive experience of reading on digital media promotes a superficial way of relating with textual information, which in turn is changing the way we process information. Although this hypothesis originally refers to the way we read on any type of medium, evidence suggests that such effect is more salient when reading on screen.

Building upon these empirical and theoretical backgrounds, our study seeks to disentangle the cognitive processes underlying shallow on-screen reading by analyzing undergraduate students' attention and meta-cognitive calibration while reading a lengthy print or digital text with or without time pressure.

1.1. On-task attention when reading on screen

A major concern regarding the impact of digitalization on information processing is a decreasing ability to focus on task (Baron, 2015; Wolf, 2018). From this perspective, reading on screen is inherently distracting as a result of frequent reading experiences based on skimming and multitasking. For example, Daniel & Woody (2013) found that engaging with competing activities when reading at home was more frequent among participants who read electronic versions of a textbook compared to those who read it in print. Nonetheless, to the best of our knowledge, no study has directly analyzed readers' attention while reading on screen, relative to reading in print. Our study is designed to fill in this gap.

On-task attention has been investigated by means of mindwandering measures. Mindwandering can be defined as unconstrained self-generated mental activity characterized by thoughts that arise independently of the task being performed, which have been called task-unrelated thoughts (TUTs; Smallwood, 2013). Mindwandering is part of a general process that implies attentional shifts from external to internal experiences. The most used method to capture the presence of TUTs is the *probe-caught technique*, where participants are periodically interrupted during the task and asked to report whether they were mindwandering. This method is considered valid and informative to assess the occurrence of TUTs (Smallwood & Schooler, 2006), and it has been used in reading research (e.g., Dixon & Bortolussi, 2013; Feng, D'Mello, & Graesser, 2013).

Reading tasks are unique to study mindwandering, as comprehending texts involves the construction of representations of the external environment (Smallwood & Schooler, 2006). In this type of task, the occurrence and maintenance of TUTs entails that top-down attention shifts from the text content to the individual's internal activity, causing a temporary mindless reading mode. Although recent approaches suggest that mindwandering could be beneficial to understand some passages of literary texts that, in fact, require the readers' mind to wander (see Fabry & Kukkonen, 2019), its detrimental consequences for reading comprehension have been widely reported (Feng et al., 2013; McVay & Kane, 2012; Soemer & Schiefele, 2019; Unsworth & McMillan, 2013). The present study tests the hypothesis that shallow processing on screen is related to inattentive reading. Examining how reading media affect readers' mindwandering would provide a direct explanation of the on-line processes responsible of shallow processing on screen. We expect that the effect of increased mindwandering may become more harmful when focused attention becomes more critical, such as when reading under time constraints (Ackerman & Lauterman, 2012; Delgado et al., 2018).

1.2. A metacognitive deficit when reading on screen and the inattentive reading

A different explanation for the shallower reading of digital texts is provided by the *metacognitive deficit* hypothesis. In one of the most relevant attempts to understand the underling mechanisms of on-screen inferiority, Ackerman and colleagues studied people's metacognitive calibration (Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014; Sidi, Shpigelman, Zalmanov, & Ackerman, 2017). Calibration is deemed a product of self-regulated learning processes which refers to a monitoring skill that reflects the accuracy of learners' perceptions of their own performance (Pieschl, 2009). Calibration tends to be poor, with learners often being overconfident (see Stone, 2002). In a series of studies, Ackerman and colleagues consistently found that the calibration accuracy of the participants in their studies was inferior when the experimental task was accomplished on a computer relative to printed materials, both when participants read texts to answer comprehension questions (Ackerman & Lauterman, 2012; Lauterman & Ackerman, 2014), and when they solved brief problems (Sidi et al., 2017). As a consequence of this heightening metacognitive inaccuracy, the authors argue, outcomes were poorer when performing the tasks on screen under time constraints. This is because the influence of self-monitoring processes becomes more relevant when time management is crucial. Authors also found that, under time pressure, participants' overconfidence in their own performance was larger on screen both when reading (Ackerman & Lauterman, 2012) and when solving short problems (Sidi et al., 2017).

Potentially, the relationship between a metacognitive monitoring deficit and inattentive reading when reading on screen may be bidirectional. On the one hand, lower on-task attention could hinder monitoring, as the occurrence of off-task periods will prevent readers from accurately judging their current level of understanding. Conversely, overconfidence in one's level of comprehension could liberate cognitive resources which could be dedicated to mindwandering (Smallwood & Schooler, 2006). Whatever the nature of this relationship, difficulties in time managing as a consequence of lessened on-task attention and the metacognitive deficit mentioned will lead to increased overconfidence on one's performance, especially under time constraints.

1.3. Theoretical and educational implications of the reading medium effect

The existence of a reading medium effect on text comprehension raises several theoretical and educational concerns. From a theoretical perspective, models of reading comprehension have accounted for a wide range of factors affecting comprehension processes, with especial attention to the interaction between individual's characteristics (e.g., decoding skills, attention capacity), task features (e.g., reading goals), and the text content and structure (e.g., text genre, complexity; see McNamara & Magliano, 2009). Nonetheless, in spite of the empirical evidence suggesting that medium affects reading comprehension (Clinton, 2019; Delgado et al., 2018), this factor has been consistently ignored in major theoretical models of comprehension.

According to the shallowing hypothesis (Annisette & Lafreniere, 2017) and the metacognitive deficit hypothesis (Ackerman & Lauterman, 2012; Sidi et al., 2017), we propose that screens themselves could activate an effortless cognitive style, characterized by lack of on-task attention, superficial processing, and lessened metacognitive monitoring. In this sense, a recent model of reading comprehension, the RESOLV model (Rouet, Britt, & Durik, 2017), assumes that contextual features play a decisive role in readers' engagement with the text. According to the RESOLV model readers construct a context schema based on their interpretation of the physical and social context and on previous typical experiences within similar contexts. Reading on screen is generally characterized by quick and superficial reader-text

interactions, (e.g., Liu, 2005; Pernice, Whitenton, & Nielson, 2014), which has been called the *zapping attitude* to text (van der Weel, 2011). Accordingly, a digital reading medium could activate a context schema which induces a particular shallowing processing style of the written information.

The influence of the reading medium on comprehension also has a substantial impact on education. If reading on screens prevents readers to fully engage with the text, either hindering their on-task attention or their metacognitive comprehension monitoring, they should not be recommended as a main source of information. Furthermore, as the negative effect of on-screen reading especially arises when reading under time constraints (Ackerman & Lauterman, 2012; Delgado et al., 2018), special attention should be taken when conducting timed exams on screens. Recent studies analysing data from the Program for International Student Assessment (PISA) have found that the change in presentation mode from print to computerized tests in PISA 2015 had a detrimental impact on students' PISA test scores in Germany, Sweden and Ireland (Jerrim, Micklewright, Heine, Salzer, & McKeown, 2018; Robitzsch, Lü;dtke, Goldhammer, Kroehne, & Köller, 2020). In a time when online education is becoming ubiquitous due to worldwide lockdowns associated to the COVID-19 pandemic, it is urgent to understand the underlying mechanisms of the medium effect as a first step to minimize the negative educational effects of on-screen reading.

1.3.1. The present study

The present study aimed to replicate the on-screen reading inferiority effect under time constraints, as well as to shed light on the explanation for such effect. Following the call for using more ecologically valid materials in reading research (Mangen, Olivier, & Velay, 2019), participants in our study read a text substantially longer than what is typical in this research field. Participants were randomly allocated to one of four experimental conditions, so that they read either in-print or on-screen, with or without time pressure. We measured text comprehension, mindwandering, and metacognitive calibration. Besides, we measured a comprehensive set of covariates to control for their potential influence on comprehension (Ackerman & Lauterman, 2012; Daneman & Merikle, 1996; Guthrie, Klauda, & Ho, 2013; Hidi, 2001; Naumann, 2015; Ozuru, Dempsey, & McNamara, 2009) and mindwandering (Feng et al., 2013; Fulmer, D'Mello, Strain, & Graesser, 2015; Kane & McVay, 2012; Randall, Oswald, & Beier, 2014; Unsworth & McMillan, 2013; Xu & Metcalfe, 2016). Our hypotheses were:

1. Participants reading on screen will mindwander more than those reading the printed text, regardless of time pressure.
2. Participants reading on screen under time pressure will show poorer calibration of comprehension (i.e., increased overconfidence in their performance on a comprehension test) than the other groups.
3. Participants reading on screen under time pressure will comprehend less than the other groups.

2. Method

2.1. Participants

One hundred and forty first-to-fourth year undergraduate students of pedagogy, teaching, and psychology of a large Spanish university volunteered for class credit. All participants had Spanish as their native language, and the mean age of the sample was 20.46 years ($SD = 1.57$). All participants provided informed consent, and they were debriefed after completing the study.

As indicated by a priori power analyses (G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007) with alpha and beta levels respectively set at 0.05 and 0.20, a 140-participant sample is appropriate to detect an interaction effect of medium and time-frame both on reading comprehension and on readers' calibration respectively equal to a partial eta-squared of 0.07 (minimum necessary sample size = 107) and 0.06

(minimum necessary sample size = 125; sizes of the interactive effects found by means of a similar experimental design by Ackerman & Lauterman, 2012).

2.2. Materials

Text. We used a lengthy expository text on human learning and artificial intelligence that comprised two figures and 3010 words (including figure captions), distributed across four pages. We used authentic versions of the text published in the science-dissemination magazine *Investigación y Ciencia*, the Spanish edition of the magazine Scientific American. As can be seen in Fig. 1, in the in-print reading condition we provided the article in the actual magazine, while in the on-screen condition we provided its pdf version on a desktop computer (screen size 17"). The pdf initially presented one page by screen, but participants were allowed to set the zoom at their own pace.[1] In case they did zoom in, scrolling down the text by using the mouse wheel was necessary.

The Inflesz Scale for Spanish (Barrio-Cantalejo et al., 2008) considers the number of syllables per word and the number of sentences to estimate texts' readability. According to this index, the readability of the text was 46.69, indicating that it was *Somewhat difficult*, a readability category equivalent to scientific-dissemination texts or specialized press. We chose such challenging text on purpose, as we expected it would allow for enough variation in terms of students' cognitive engagement, which is critical to test the shallowing hypothesis. Moreover, a sample of 20 participants in a pilot study took 21.93 min on average to complete the reading task. Thus, we considered the text long enough to capture variability in participants' mindwandering while reading.

2.3. Dependent measures

Multiple-choice comprehension test. We constructed 21 four-alternative questions, including seven questions for each of the following three comprehension processes: text-based (i.e., a single idea explicitly stated in a single sentence), local inference (i.e., a bridging inference linking two adjacent sentences), and global inference (i.e., a bridging inference linking information located more than two sentences apart; an example of each type of question can be found in the Appendix). The four response options for each question included the target and three different distractors: near-miss (an idea located in the text that conceptually taps the target answer), thematic (a plausible answer but containing common misconceptions), and unrelated distractor (an extremely improbable answer or inconsistent with the text content; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007).

We conducted exploratory factor analysis (EFA) to examine the structure of the 21 questions. We tested the existence of three factors addressing the three aforementioned comprehension processes, respectively, using the maximum likelihood method. Given the binary nature of the items (i.e., correct/incorrect responses), these analyses were based on a polychoric-transformed correlation matrix of the dataset. The EFA did not yield any acceptable model in which each factor was uniquely loaded by one type of question. Therefore, we assumed that our test assessed a single construct (i.e., reading comprehension). We then conducted EFA fixed to one factor and we excluded seven questions whose factor loading was non-significant and lower than $g = .30$ (Costello & Osborne, 2005). Thus, the test finally consisted of 14 questions (factor loading ranging from 0.31 to 0.72) Test reliability was measured using the omega coefficient (McDonald, 1999) also based on a polychoric-transformed correlation matrix, because this index is deemed more appropriate than Cronbach's alpha for dichotomous items

---

[1] If needed, participants set the zoom in advance using the last page of the previous article.

Fig. 1. First pages of the on-screen (pdf file) and the printed version of the reading material.

(Trizano-Hermosilla & Alvarado, 2016; Viladrich, Angulo-Brunet, & Doval, 2017). The results indicated good reliability for the 14 questions selected ($\omega = .81$). Maximum score on this test was 14.

Mindwandering probes. The frequency of TUTs was assessed by means of the probe-caught technique (Feng et al., 2013). While reading the article, either in print or on screen, the first author orally interrupted participants approximately each 99 s, and they immediately had to indicate whether they were paying attention to a TUT at that moment. They were previously instructed to identify on-task thoughts ("*Thoughts about the text content or about how well you are understanding it*") and TUTs ("*Thoughts about your daily stuff, a memory from the past, something in the future, your current state of being, or any other type of thought not related with the text content nor with the understanding of it*") (McVay & Kane, 2012). This measure was completed in a separated sheet of paper by ticking *yes* (*I was wandering*) or *no* (*I wasn't wandering*) for each probe. The TUT proportion on the probes was calculated for each participant, ranging from 0 to 1.

Metacognitive calibration. After reading, participants predicted their performance on the comprehension test by estimating the percentage of correct answers in a continuous 25–100% scale (Ackerman & Lauterman, 2012). Calibration for each participant was calculated by subtracting the percentage of correct answers in the comprehension test from their prediction of performance (POP), which allowed us to perform correlation analyses between this measure and the other measured variables. Besides, participants' POPs were statistically compared to the actual performance by means of repeated-measures analyses.

2.4. Covariates

2.4.1. Covariates related to participants features

Working memory. We used the Letter-Number Sequencing Test from the Spanish version of the Wechsler Adult Intelligence Scale-IV (Wechsler, 2008) to measure working memory capacity. In this test, the evaluator enunciates a series of alternating numbers and letters, and individuals report back the numbers from lowest to highest, and the letters in alphabetical order. Difficulty increases from a 3-item to an 8-item series. Its application procedure was adapted to a group application, and participants wrote down their responses (cf. Macedo-Rouet et al., 2019). Two experiment administrators ensured that participants did not annotate any digit/letter while the series were being read aloud. Moreover, when writing down the responses, participants were not allowed to start writing the final elements of the series and then complete it with other elements.

Prior knowledge. We constructed a self-reported 8-item questionnaire as an indicator of participants' prior topic knowledge. Participants rated their knowledge on four subtopics related to human learning (e.g., *brain processes involved in human learning*), and four related to artificial intelligence (e.g., *computer programming*), using a scale from 1 (*I know nothing*) to 10 (*I am an expert*). Cronbach's alpha was good for the items on human learning ($\alpha = .84$) and acceptable for the items on topics on artificial intelligence ($\alpha = .70$).

As can be seen in Table 3, participants' self-reported prior knowledge on human learning correlated with their scores on the comprehension test, whereas prior knowledge on artificial intelligence did not. These results support our decision to include participants' prior knowledge on each topic separately.

Topic interest. We constructed a self-reported 8-item questionnaire on participants' topic interest. They rated from 1 (*not interested at all*) to 10 (*very interested*) their interest in the same eight subtopics they rated in the prior knowledge questionnaire. The reliability was good for the items on human learning ($\alpha = .87$) and acceptable for the items on artificial intelligence ($\alpha = .75$).

Medium preference and use. We constructed a 5-item (two reversed) questionnaire to measure participants' medium preference for reading to learn. For each item, participants rated from 1 (*I totally disagree*) to 10 (*I totally agree*) a statement regarding the use of printed vs. digital texts

Table 1
Distribution of participants and participants' age, sex, grade year, and bachelor's degree across experimental condition.

Medium	In print		On screen	
Reading time-frame	Free	Pressured	Free	Pressured
n^a	33	33	33	33
Age	20.42 (1.27)	19.97 (1.36)	20.82 (1.90)	20.42 (1.64)
Percentage of females	81.82	75.75	84.85	81.82
Grade year[b]:				
1st year	1	3	2	3
2nd year	12	18	11	16
3rd year	14	7	12	10
4th year	6	4	8	4
Bachelor[a]:				
Pedagogy	6	4	8	4
Psychology	5	6	2	5
School Teaching	22	23	23	24

Note.
[a] After excluding outliers in comprehension.
[b] In number of participants.

Table 2

Means and standard deviations of the covariates in each experimental condition. Skewness and kurtosis of the covariates in the whole sample.

Medium	In print		On screen		Whole sample	
Reading time frame	Free	Pressured	Free	Pressured	Skewness[a]	Kurtosis[a]
Working memory	20.97 (2.26)	21.48 (2.37)	21.67 (2.55)	21.00 (2.75)	0.48	0.26
Prior knowledge on human learning	6.57 (0.94)	6.64 (0.96)	6.28 (1.16)	5.82(1.02)	−0.15	−0.75
Prior knowledge on artificial intelligence	3.17 (1.09)	3.93 (0.86)	3.07 (0.98)	2.95 (0.95)	0.55	−0.50
Topic interest (human learning)	8.72 (1.03)	8.52 (1.15)	8.70 (1.19)	8.78 (1.11)	−0.84	−0.08
Topic interest (artificial intelligence)	4.90 (1.69)	5.35 (1.54)	4.76 (1.72)	4.96 (1.71)	0.09	−0.78
Medium preference	8.24 (1.33)	7.58 (1.63)	8.32 (1.81)	8.33 (1.49)	−0.66	−0.54
Starting age of use of digital technologies	10.56 (2.49)	10.30 (2.43)	10.39 (2.67)	11.18 (2.54)	0.05	−0.30
Use of digital tech for leisure[b]	2.55 (1.52)	3.12 (1.68)	2.79 (1.22)	3.03 (1.58)	0.51	−0.36
Use of digital tech for study/professional purposes[b]	3.97 (1.90)	4.55 (2.55)	4.09 (2.13)	4.58 (2.47)	−0.37	−0.74
Perceived text difficulty	6.42 (1.94)	7.00 (1.16)	6.39 (1.64)	6.50 (1.54)	−0.84	0.37
Situational interest	7.18 (2.01)	6.76 (2.17)	6.52 (1.99)	7.05 (1.23)	−0.58	−0.52
Reading time (in min.)	24.18 (4.57)	16.30 (−)	22.49 (4.11)	16.30 (−)	0.32	0.42

Note.

[a] Once outlier values (±2*SD* from the sample mean) replaced using winsorization.

[b] In daily hours.

Table 3

Pearson correlations between all the measured variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Prior knowledge on artificial intelligence	–													
2. Prior knowledge on human learning	.30**	–												
3. Topic interest (artificial intelligence)	.56***	.11	–											
4. Topic interest (human learning)	-.05	.21*	.11	–										
5. Medium preference	-.10	.04	-.03	.14	–									
6. Starting age of use of digital technologies	.11	-.05	.12	.02	.15	–								
7. Daily use of digital tech for leisure	-.04	-.15	.11	-.02	-.00	-.09	–							
8. Daily use of digital tech for study/prof.	.06	-.03	-.06	-.07	-.08	-.10	.16	–						
9. Perceived text difficulty	-.07	-.20*	-.07	-.04	.14	.06	.09	-.02	–					
10. Situational interest	.07	.02	.18*	.09	-.07	.10	-.06	.06	-.29**	–				
11. Working memory	.18	.03	.20*	-.11	-.20*	.07	.03	.07	-.04	.16	–			
12. TUT proportion	-.02	-.02	-.18*	-.07	-.07	-.04	-.03	-.11	.22*	-.47***	-.05	–		
13. Calibration index	.08	.01	.04	-.10	.08	.09	.14	.08	-.12	-.03	-.19*	.10	–	
14. Comprehension questions	.01	.18*	.02	.03	-.14	-.13	-.07	-.10	-.25**	.28**	.22*	-.27**	-.71**	–
	1	2	3	4	5[1]	6	7	8	9	10	11	12	13	14

Note. *p < .05; **p < .01; ***p < .001.

for learning purposes (e.g., *I understand and memorize better when I study reading an electronic text than when I read on paper*). A mean score above 5 points indicated preference for paper. This questionnaire showed a good reliability level (α = .81). In addition, participants indicated at what age they started to use digital devices regularly, and how many hours a day they use them for leisure and for educational/professional purposes.

2.4.2. Covariates related to participants' perceptions of task features

Perceived text difficulty. Participants were asked to indicate, after reading, their perceived text difficulty from 1 (Very easy) to 10 (Very difficult) in a single Likert-scale item.

Situational interest. Participants rated, after reading, their interest in the text content from 1 (*Not interesting at all*) to 10 (*Very interesting*) in a single Likert-scale item.

2.5. Procedure

Tasks were completed in one small group session (six participants maximum). All the participants in each session performed the experimental task under the same condition (i.e., same reading time-frame and reading media). Sessions were conducted in a silent room and lasted approximately 75 min. Participants first completed the self-reported questionnaires on prior knowledge and topic interest, followed by the working memory test. Then, participants were introduced to the reading task: "*You are now going to read an article to learn as much as you can, because you will be asked to complete a test consisting of 21 four-alternative multiple-choice questions on the text content. Please note that you won't be allowed to go back to the text while answering the questions*". The

instructions for the free-time condition continued as follows: "*You can read the article at your own pace. When you consider you have read enough, raise your hand and wait to be given the comprehension test. It is important that you do not disturb the other participants, so please do everything very silently*". The pressured-time groups had only 16 min and 30 s to read the text, which represents the 75 percent of the mean time that the participants in the pilot study took to read the article at their own pace. The instructions for this condition continued as follows: "*You must keep in mind that you have little time to read the article. You only have 16 min and a half, which is the 75 percent of the time that a group of people spent on average when reading at their own pace. I will let you know when you have gone through half the time, also when there have 4 min left, and finally when there is only 1 min and a half left. You will have to stop reading when you are told that the time is up and you will then receive the questions*". Afterwards, participants were instructed in how to perform the mindwandering probe-caught task, and they were reminded to be honest when answering the probes. Participants in the pressured-time condition were probed 10 times, the last one just before reading time was over, while the number of probes for the participants in the free-time condition ranged from 10 to 19.

When the reading task was finished, participants predicted their performance on the reading comprehension test, and subsequently completed it. Then, they reported their perceived text difficulty and interest on the text, and answered the medium preference and use questionnaire. Finally, those who read under time pressure were asked whether they were able to finish reading the whole text and to indicate at what point they had to stop approximately.

3. Results

Two participants were excluded from the dataset because they did not perform the task properly.[2] Data distribution of the reading comprehension scores, TUT rate, and calibration was inspected before conducting the main analyses. Six participants were identified as outliers, as they scored in the reading comprehension test below 2 *SDs* from their time-frame group mean (sizeable differences in this measure existed between participants in the free and pressured-time groups). They were excluded from subsequent analyses, and therefore the final sample consisted of 132 participants, still appropriate according to the results of the power analyses reported above. Once outliers removed, sample's mean score on the reading comprehension test was 8.72 (SD = 2.63; range: 3–13) Table 1 shows the number of participants within each experimental group. ANOVA and chi-squared analyses indicated no significant differences between the experimental groups regarding participants' age, sex, grade year, and bachelor's degree (all *ps* > .12; see also Table 1).

Approximately half of the participants who read under time pressure mentioned that they could not read the whole text. Among those, they had to stop at some point of the second half of the last page of the 4-page text. Chi-squared analysis showed that they were equally distributed across reading medium groups, $X^2(1, N = 66) = .55, p = .46$ ($n = 14$ in the in-print group; $n = 17$ in the on-screen group).

3.1. Covariates

Table 2 includes descriptive information for all measured covariates for each group, and Table 3 shows Pearson correlations among all variables. Fig. 2 shows the correlation pattern among correlated variables. Outliers for each covariate were identified ($\pm 2SD$ from the sample mean) and replaced by the next highest or lowest score that was not an outlier (i.e., *winsorization*; Field, 2013). The number of outliers for each covariate ranged from 2 to 6 cases, representing less than 5% of the data. As can be seen in Table 2, all the covariates were then normally distributed, as values for Kurtosis and Skewness were within the ± 2 range (George & Mallery, 2010). Regarding correlations between covariates and dependent measures, scores in the comprehension test positively correlated with working memory and situational interest, and negatively with text difficulty. The TUT proportion positively correlated with perceived text difficulty, and negatively with participants' topic interest in artificial intelligence and situational interest. The remaining covariates didn't correlate with text comprehension scores or TUT proportion. Finally, participants' calibration index correlated with participants working memory.

To ensure that groups were comparable, we performed a series of ANOVAs with reading medium and reading time as independent variables, and each possible covariate as dependent variable. Only two of them differed between groups. Participants under pressured time reported higher perceived text difficulty regardless of medium, $F(1, 128) = 4.73, p = .03, \eta^2_p = .04$, and self-reported prior knowledge in human learning was higher in the in-print groups than in the on-screen groups, $F(1, 126) = 9.85, p < .01, \eta^2_p = .07$.

Based on the differences between groups and on the correlations between covariates and dependent measures, we included working memory, situational interest, and prior knowledge in human learning as covariates in the ANCOVA for text comprehension scores. The same covariates as well as participants' topic interest in artificial intelligence were included in the analysis of TUT proportion. In this case, controlling for working memory was a decision driven theoretically by previous evidence (i.e., working memory and mindwandering generally correlate negatively; Kane & McVay, 2012; Unsworth & McMillan, 2013; Randall

et al., 2014). Perceived text difficulty was not included as covariate in both cases due to its substantive dependence on the reading time-frame conditions (Miller & Chapman, 2001). Finally, participants' working memory was included as covariate in the ANCOVA for metacognitive calibration (see Fig. 2).

3.2. Effects of medium and reading time-frame on mindwandering

Means and standard deviations for the TUT proportion in each experimental condition can be seen in Table 4. ANCOVA assumptions with respect to normality, homogeneity of variances, and homogeneity of regression slopes between the three included covariates and the dependent variable were met. Thus, a two-way (medium x reading time) ANCOVA revealed no main effect of medium, $F(1, 124) = 1.21, p = .27$, and a and reading time-frame, $F(1, 124) = 2.09, p = .15$, on the TUT proportion. However, an interaction effect between these two factors qualified their effects on mindwandering, $F(1, 124) = 4.03, p = .047, \eta^2_p = .03$. Post hoc Bonferroni-corrected pairwise comparisons indicated that, across time-frames, participants reported a significant lower TUT proportion in the in-print group under time pressure than under free-time, $F(1, 124) = 5.56, p = .02, \eta^2_p = .04$, while it was not statistically different across reading time groups in the on-screen condition, $F < 1$. Moreover, across reading media, the TUT proportion was significantly lower in the in-print group than in the on-screen group when reading under time pressure, $F(1, 124) = 4.52, p = .03, \eta^2_p = .04$. There was no difference between media under free reading time, $F < 1$ (see Fig. 3).

3.3. Effects of medium and reading time-frame on calibration

Participants' POPs for the comprehension test were compared to their actual performance to examine whether their metacognitive calibration differed across experimental groups. We conducted a two-way repeated measures mixed ANOVA with medium and time-frame as between-participants factors, and calibration as a within-participants factor (i.e., participants' POPs vs. text comprehension scores). Given that calibration is our focus here, we report only results from the tests of within-participants effects. Results revealed no differences between POPs and actual performance in the whole sample, $F(1, 124) = 2,88, p = .09$, and no interaction effects between calibration and medium, $F < 1$, calibration and reading time-frame, $F(1, 124) = 3.50, p = .07$, and calibration, medium and time-frame, $F < 1$ (see also Table 4). Thus, participants showed to be well calibrated regardless of the reading medium and the reading time-frame.

The fact that some participants under time pressure could not finish reading the text might have affected their calibration. We therefore re-run the mixed ANCOVA described above only in the pressured-time groups, with medium and whether the participants had finished reading the text as between-participants factors. There was no effect of finishing reading the text and no interaction between this factor and medium, both $Fs < 1$.

3.4. Effects of medium and reading time-frame on reading comprehension

Finally, differences on text comprehension scores were examined by means of a two-way ANCOVA. Means and standard deviations for each group can be seen in Table 4. ANCOVA assumptions with respect to normality, homogeneity of variances, and homogeneity of regression slopes between the included covariates and the dependent variable were met in all cases except for participants' situational interest. Thus, we applied the blocking procedure (Tabachnick & Fidell, 2014) to include this covariate as an independent variable, as it allowed to focus on the effects of the independent variables of interest (i.e., medium and time-frame) once the variation of this covariate is removed from the estimated error. Following Tabachnick and Fidell (2014) procedure, a new independent variable was created by categorizing situational
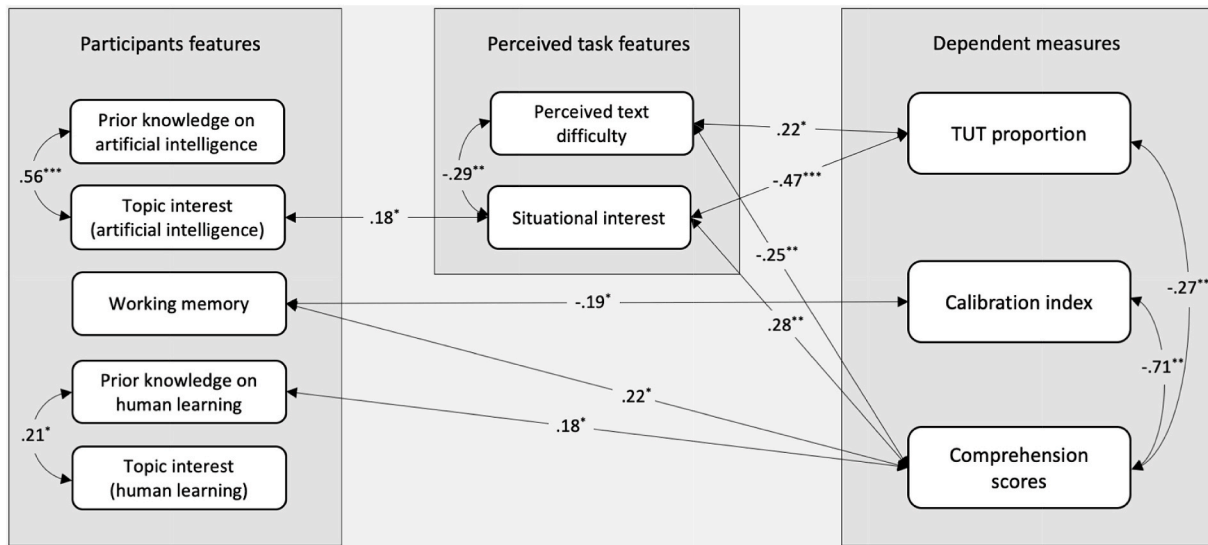
---

[2] They took approximately 11 min to read the text, which is below three *SD* from the mean of the free time condition.

Fig. 2. Correlational pattern between correlated variables in the study. *p < .05; **p < .01; ***p < .001.

Table 4
Means (SD) for TUT proportion, POPs, and reading comprehension scores. Skewness and Kurtosis for each dependent variable.

Medium	In print		On screen		Whole Sample	
Reading time-frame	Free	Pressured	Free	Pressured	Skewness[a]	Kurtosis[a]
TUT proportion	.22 (.15)	.15 (.14)	.22 (.15)	.22 (.14)	0.40	−0.67
POPs in comprehension test[b]	64.55 (12.28)	66.41 (13.66)	62.76 (11.50)	54.10 (15.16)	−0.25	−0.34
Text comprehension[c]	67.31 (16.27)	63.64 (20.62)	68.18(14.41)	50.00 (18.12)	−0.30	−0.72

Note.
[a] Once outliers in text comprehension removed (±2SD; n = 6).
[b] Estimated percentage of correct answers.
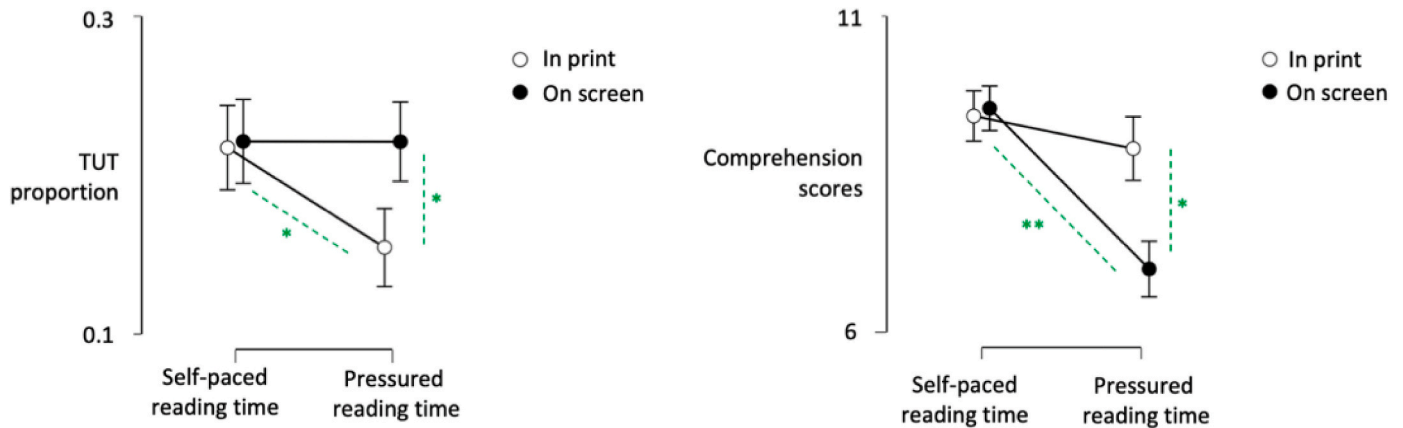[c] Percentage of correct answers.

Fig. 3. Descriptive plots of the mean TUT proportion (left panel) and the mean scores on the text comprehension test (right panel) in each experimental group. Vertical bars represent standard errors. Dashed green lines show significant differences *p < .05; **p < .01. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

interest values into three levels (low, medium, high) based on percentiles of 33 and 66. We then performed a 3-way ANCOVA, with medium, time-frame, and situational interest as independent variables, and working memory and prior knowledge on human learning as covariates. Results revealed no main effect of medium, $F (1, 118) = 1.00, p = .32$, a significant main effect of reading time-frame, $F(1, 118) = 13.36, p < .001, \eta^2_p = .10$, indicating higher scores in the free time condition, and a significant interaction effect of medium and time-frame on comprehension outcomes, $F(1, 118) = 4.82, p = .03, \eta^2_p = .04$. Across

time-frames, post-hoc Bonferroni-corrected pairwise comparisons showed that in the on-screen condition, participants who read under time pressure scored significantly lower than those who read at their own pace, $F(1, 118) = 17.50, p < .01, \eta^2_p = .13$. This was not the case for the in-print condition, where participants scored similarly regardless of the reading time-frame, $F < 1$. Furthermore, across reading media, participants who read under time pressure on screen scored significantly lower than those who read in print, $F(1, 118) = 4.89, p = .03, \eta^2_p = .04$, while no medium difference was found for participants reading with free

time, $F < 1$ (see Fig. 3).

Given the expected negative correlation between participant's TUT proportion and reading comprehension outcomes (see Table 3 and Fig. 2), we further explored the role of mindwandering on the effects of medium and time-frame on text comprehension scores. To that end, we re-run the ANCOVA described above introducing TUT proportion as covariate. The results showed that the interaction effect of medium and time-frame on comprehension scores lost significance, $F(1, 117) = 3.22$, $p = .07$.

4. Discussion

The present investigation examined for the first time how reading medium affects readers' on-task attention while reading an authentic, lengthy expository text. It also contributed to the research efforts on how medium affects reading comprehension and metacognitive calibration. Controlling for a comprehensive set of covariates, our findings revealed that reading on screen prevented readers from reducing their mindwandering when the task requirements called for efficient reading. Accordingly, participants reading on screen under time pressure scored significantly lower in the reading comprehension test, relative to those in the other three groups. Finally, contrary to our expectations, participants were equally well calibrated regardless of the experimental group. Next, we discuss the implications of these results with respect to the influence of mind-wandering on text comprehension, and how the lack of increased attention when reading on-screen can explain the screen-inferiority effect.

4.1. On-task attention and on-screen reading comprehension

Our experimental design provided a direct test of two potential underlying factors of the screen inferiority effect: inattentive reading and metacognitive calibration deficit. With respect to our first hypothesis, we expected to observe inattentive reading (i.e., higher TUT rate) when reading on screen, as compared to in-print reading, regardless of the time allocated to reading. As a consequence of this increased mindwandering, we hypothesized that disruption in reading comprehension due to mindwandering would be especially noticeable under time pressure. Yet, inattentive on-screen reading, although as such confirmed by our results, emerged in a different manner. When participants read at their own pace, they mindwandered to a similar extent, regardless of the medium. But when they read with time constraints, only in-print participants reduced the frequency of TUTs. Previous evidence has shown that learners can control the occurrence of mindwandering as the task demands call for it, especially those with greater working memory capacity (Rummel & Boywitt, 2014; Smallwood & Andrews-Hanna, 2013). Thus, given that in the present study participants' working memory did not differ across groups, we should have observed also reduced mindwandering when reading on screen under time pressure. Nevertheless, our results indicate that on-screen readers struggle to adjust to such high task demands. Although the size of this effect in our study was small according to Cohen's benchmark ($\eta^2_p = .04$, equivalent to $d = 0.41$), it was larger than the effect of increased mindwandering when reading difficult (vs. easy) texts found by previous studies ($OR = 1.24$, equivalent to $d = 0.12$, in Feng et al., 2013; $R^2 = .016$ equivalent to $d = 0.25$, in Mills, D'Mello, & Kopp, 2015). Similar findings were recently reported by Latini, Bråten, Anmarkrud, & Salmerón, 2019, who found that in-print readers were more able than on-screen readers to adapt to the learning demands of a multiple document comprehension task. Specifically, when instructed to prepare for an exam, as opposed to reading for pleasure, in-print readers wrote longer essays and indirectly integrated better the information from different sources. Such adaptation was not present among on-screen readers.

In line with previous findings (Feng et al., 2013; McVay & Kane, 2012; Soemer & Schiefele, 2019; Unsworth & McMillan, 2013), the TUT proportion in our study correlated negatively with text comprehension

scores. Results from the in-print group indicated that participants reduced their mindwandering while reading under time pressure, as compared to when reading with free time. This accommodation of readers' attention could have counteracted the detrimental effect of time pressure, resulting in similar comprehension scores across the two in-print groups regardless of the time-frame. However, that was not the case for participants who read on screen. They did not adaptively decrease the frequency of TUTs, when reading under time pressure, and their reading comprehension scores were significantly lower in this condition, which is in line with previous findings showing on-screen inferiority when reading under time constraints (Ackerman & Lauterman, 2012; Delgado et al., 2018; Lauterman & Ackerman, 2014; see). Moreover, the fact that the interaction effect of medium and time-frame on reading comprehension outcomes lost significance when controlling for participants' TUT proportion supports the idea that the on-screen reading inferiority in our study was driven, at least partially, by the fact that participants did not decrease their mindwandering when the task demanded higher on-task attention when reading on-screen.

In sum, our findings partially support the inattentive reading hypothesis by pointing to difficulties observed by on-screen readers to meet the task demands calling for increased on-task attention. Accordingly, the fact that on-screen participants in the pressured-time group didn't reduce their mindwandering provided direct evidence for a shallow processing during on-screen reading (Annisette & Lafreniere, 2017).

4.2. Metacognitive monitoring and on-screen reading

Unexpectedly, our results showed that undergraduates could accurately calibrate their reading comprehension regardless of reading medium and time-frame, contrary to previous findings yielding better calibration when reading in print (e.g., Ackerman & Lauterman, 2012). It was also unexpected that our participants were well calibrated in all the experimental conditions, as it has been widely reported that learners tend to be overconfident (Stone, 2002). The experimental procedure employed in the present study may have helped participants to make accurate predictions about their level of performance. Firstly, the available reading time in the pressured-time condition was certainly scarce, as indicated by the fact that approximately half of the participants in these groups could not reach the end of the text. Secondly, the caught-probe technique could make participants aware of their distraction while reading. Therefore, although our results indicated that not having finished reading the entire text did not provide additional support for participants' calibration, the time pressure as well as tracking their own attention might have been used by participants as cues to identify to what extent they could complete the reading assignment. The observed significant negative correlation between the TUT proportion and POPs supports this idea. Therefore, these two circumstances could have lead participants to be cautious rather than overconfident in their POPs. Thus, in spite of our results, the metacognitive deficit observed in prior studies for on-screen reading can't be discarded. So far, whether on-screen reading harms meta-cognitive calibration, and under what circumstances it occurs, if so, still remains an open question (Singer Trakhman, Alexander, & Silverman, 2018).

4.3. Inattentive on-screen reading and metacognitive monitoring

The deficit in meta-cognitive monitoring when performing tasks on screen, especially under time pressure, reported in previous research (Ackerman & Lauterman, 2012; Clinton, 2019; Lauterman & Ackerman, 2014; Sidi et al., 2017) could be related to difficulties in reducing mind-wandering. There is no reason to consider that such deficit is constrained to metacognitive judgments. On the contrary, mind-wandering could affect other monitoring processes related to task engaging, such as meta-consciousness. Several studies have examined the meta-cognitive status of mindwandering tracking participants'

awareness of their TUTs. According to the *meta-awareness hypothesis* (Schooler, 2002), results showed that readers are often unaware of their mindwandering (Smallwood, 2013; Smallwood & Schooler, 2006). Thus, a broader meta-cognitive deficit could lead not only to an increased calibration inaccuracy, but also to generate TUTs more often when reading on screen compared to printed texts. This possibility could be tested in future studies.

4.4. Limitations and future research

The present study is not exempt from limitations. Although using an authentic, lengthy text is a strength of our study, we can't generalize our results to shorter texts. Furthermore, we can only ensure that the reported differences appeared when reading this particular type of text. Therefore, future research should test if the inattentive reading hypothesis can explain screen inferiority in shorter learning tasks (cf. Sidi et al., 2017), as well as in a wide variety of text structures, difficulty, or genres. In addition, although our study included a comprehensive set of covariate measures, other individual factors, such as participants' sustained attention capacity, or a more exhaustive working memory measurement, could help to further explain mind-wandering during on-screen reading. Future studies could investigate whether the on-screen inferiority effect depends on those individual differences, to the extent that they can be related to the inattentive reading as an explanation (cf. Ben-Yehudah & Brann, 2019). Furthermore, this research line could be especially relevant among school-aged students, as they show a larger variability with respect to individual differences.

It is also noteworthy that we can't rule out the possibility that reading on screen is not inattentive in itself, but its disruptive effect may be caused by the fact that reading on desktop screens involves a body position that could hinder on-task attention. Reading is deemed a cognitive activity with an embodied nature, and the physical relationship between reader and text is different between these media (Mangen & van der Weel, 2016). In this vein, the regular body posture used while reading printed materials on a desk could facilitate the immersion in the text, increasing on-task attention, as compared to reading on screen. A recent study by Mangen et al. (2019) investigated reading a lengthy literary text on a Kindle vs. a printed book. They found that reading engagement was similar across both media, as well as most of the measures of participants' comprehension of the text. However, those who read in print constructed a better mental representation of the story chronology. Given that readers' engagement with the text was similar regardless of the medium, the authors concluded that the difference found in this mental representation was due to differences in readers' sensorimotor experience with the reading devices. If this circumstance somehow also impacts comprehension of complex and lengthy expository texts, the effect could have been even larger in our study, because our participants read on a stationary computer screen that did not allow any physical interaction with the reading device. To rule out this possibility, our study could be replicated by using tablets instead of desktop computers, given that handheld devices allow for a reader-text interaction more similar to that of reading on paper.

Another issue that was not addressed by our study was the interaction between the learners' age and the impact of the reading medium on text comprehension. It is still unclear whether the negative influence of reading on screens varies over individuals' development. Most of the studies comparing comprehension outcomes and metacognitive calibration across reading media have been conducted with undergraduates (Clinton, 2019; Delgado et al., 2018). Future investigations should examine the possible relationship between readers' age and the impact of the reading medium on attention and metacognition, given that executive functions are not fully developed until adulthood (see Best, Miller, & Jones, 2009).

Finally, we should also note that the probe-caught technique used to measure participants' mindwandering was undoubtedly disrupting for their reading performance. In this sense, although this circumstance is equally present in all the experimental groups, we cannot rule out the possibility that it could have affected each group differently. It is possible that the on-screen/time pressure readers in our study were disrupted to a greater extent than those who read in print. If so, the conclusion would be that on-screen readers are more susceptible to the negative impact of extraneous tasks on attention. Further research should address this issue by using less invasive methods to measure readers' on-task attention, such as electroencephalography (e.g., Broadway, Franklin, & Schooler, 2015) or post-task retrospective questionnaires (e.g., Sanchez & Naylor, 2018, Experiment 2).

5. Conclusion and implications

Our results show that reading on screen lead to inattentive reading particularly when the task demands an increase in on-task attention for efficient information processing. We argue that this inattentive reading is causing, at least in part, a shallow information processing and lower comprehension. The findings support current concerns indicating that digital technologies, under certain circumstances, hinder reading and learning (Baron, 2015; Salmerón & Delgado, 2019; Wolf, 2018). As argued above, although the size of the on-screen inferiority effect under time pressure found in our study was small according to Cohen's (1988) benchmarks ($\eta^2_p = .04$, equivalent to $d = 0.41$), analyzing it in context provides a richer picture. This effect size is located at the lower bound of Hattie's 'zone of desired effects' in educational contexts (i.e., a medium effect; Hattie, 2009), and it represents more than the yearly growth in reading comprehension during elementary school (0.32; Luyten, Merrell, & Tymms, 2017).

Our findings emphasize the need to incorporate the effect of reading medium on general models of text comprehension. Based on the RESOLV model (Rouet et al., 2017), we could argue that the medium seems to be one of those contextual factors that influence readers' engagement with texts. Still, many factors regarding reading medium effects remain unknown. Future research should shed further light on how reading medium interact with other factors, such us individual characteristics (e.g., attentional capacity; Ben-Yehudah & Brann, 2019), task demands (e.g., task goals; Latini, Bråten, Anmarkrud, & Salmerón, 2019), texts features (e.g., presence of illustrations; Latini, Bråten, & Salmerón, 2020), or additional contextual factors (e.g., classroom vs reading at home; Daniel & Woody, 2012). A major challenge for future research is to clarify under what circumstances the medium effect becomes more salient and what factors could mitigate its consequences.

Furthermore, our results should be a matter of concern to educational practitioners and policy makers. There are clear educational scenarios where on-screen reading, in light of our results, should be avoided. For example, taking exams on screen could prevent students from fully demonstrating their knowledge and skills, because they may struggle to adjust their attentional focus to their full potential. Recent studies analyzing the change of the mode (from print to computer) in the PISA tests support this conclusion, as indicated by lower scores from the computer-based tests (Jerrim et al., 2018; Robitzsch, Lüdtke, Goldhammer, Kroehne, & Köller, 2020). Moreover, mindwandering has demonstrated to exert a negative impact on academic performance (Hollis & Was, 2016; Seli, Wammes, Risko, & Smilek, 2016). Thus, if performing learning activities on screens hinders students' ability to control the generation of task-unrelated thoughts, this will negatively impact their learning processes and outcomes. That being said, suggesting a ban on digital technologies for educational purposes would be naïve in the XXI century. They are here to stay and they offer a wide range of educational possibilities. Nevertheless, we call educational practitioners and policymakers to consider the fact that printed texts are more appropriate when it comes to in-depth reading, especially with lengthy texts. In this regard, educational systems should be especially cautious with recent campaigns supporting a complete shift from printed to e-text books. Instead, on the one hand, we should find an appropriate balance between the use of printed materials and digital technologies by

means of evidence-based decisions. On the other hand, we find it necessary that academic curricula include training in the appropriate use of digital devices as learning tools in order to help students fully benefit from them. In this regard, further research efforts should address how to overcome the on-screen reading inferiority. These are major goals for education in the Digital Age.

CRediT authorship contribution statement

Pablo Delgado: Conceptualization, Investigation, Formal analysis, Writing - original draft. **Ladislao Salmerón:** Supervision, Writing - review & editing.

Declaration of competing interest

The authors do not have any interests that might be interpreted as influencing the research, and APA ethical standards were followed in the conduct of the study.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2020.101396.

Appendix

Multiple-choice comprehension test

The questions used in the text comprehension test are listed below (translated from Spanish). To construct the questions and the options we followed the guidelines of Ozuru, Best, Bell, Witherspoon, and McNamara (2017), and accordingly each question has 4 response options consisting of "one target and three type distractors: near-miss, thematic, and unrelated. Near-miss distractors have a large conceptual overlap with the target answer and the idea is located in the passage but in an inappropriate context. […] Thematic distractors are answers that are plausible but contain erroneous information based on common misconceptions (not located within the passage). […] Unrelated distractors are highly improbable, or inconsistent with the theme of the passage." (p. 408).

As reported in the manuscript, we constructed three types of questions, depending on whether they addressed the comprehension of text-based information, local inferences, or global inferences (also according to Ozuru et al., 2017). However, the performed exploratory factor analysis (EFA) did not support the assumption that our test measured three different constructs based on these three types of questions, respectively. Thus, after excluding seven questions based on the results from the EFA considering that the test measured only one factor, we assume that our test provided a general measure of participants' mental representation of the text, including both text-based and inferential comprehension, as well as their recall of what they had read in the text. Both the included and the excluded questions are listed below.

Questions included:

1. **Which of the following statements best represents the main idea of the text?** *(Global inference)*

   a. Understanding how the human brain learns will allow us to develop robots that are as intelligent as people. *(Thematic distractor)*
   b. Developing robots that learn autonomously will increase our understanding of how the human brain learns. *(Correct answer)*
   c. To understand reality, robots must create predictions that minimize their discrepancy from reality. *(Near-miss distractor)*
   d. Increasing our knowledge of artificial intelligence allows us to create robots that can learn autonomously. *(Unrelated distractor)*

2. **According to the learning model outlined in the text, what transports the signal that returns to the higher processing levels of the brain?** *(Text-based)*
   a. Predictions about reality. *(Near-miss distractor)*
   b. Prior knowledge about reality. *(Unrelated distractor)*
   c. The discrepancy between what is predicted and what is perceived. *(Correct answer)*
   d. Information perceived by the senses. *(Thematic distractor)*

3. **Regarding social behavior, artificial intelligence research suggests that humans perform some elementary behaviors to help others because:** *(Text-based)*
   a. We learn social norms that predict what is right. *(Thematic distractor)*
   b. We try to get our predictions confirmed. *(Correct answer)*
   c. We receive social rewards by predicting the behavior of others. *(Unrelated distractor)*
   d. We have learned to predict the norms of the social world. *(Near-miss distractor)*

4. **Artificial intelligence studies could be beneficial to our understanding of developmental disorders. In this regard, the text suggests a hypothesis that proposes that people with autism prefer repetitive routines because:** *(Text-based)*
   a. This is how they learn to minimize errors in their predictions. *(Near-miss distractor)*
   b. They are not able to detect errors in their predictions. *(Unrelated distractor)*
   c. They have no cognitive skills to make predictions. *(Thematic distractor)*
   d. They are too sensitive to the errors of their predictions. *(Correct answer)*

5. **If there is a mismatch between signals from the upper levels of the brain and those from the senses, the information processing system detects estimation errors that:** *(Local inference)*
   a. Hinder the acquisition of knowledge. *(Thematic distractor)*
   b. Don't allow to really know the reality. *(Unrelated distractor)*
   c. Improve our models about reality. *(Correct answer)*
   d. Process the basic features of reality. *(Near-miss distractor)*

6. **According to the learning model proposed in the text, we have a predictive brain, because almost all our actions are based on generating and updating predictions. This process aims to:** *(Global inference)*
   a. Minimize the number of predictions generated. *(Unrelated distractor)*
   b. Check all predictions generated. *(Near-miss distractor)*
   c. Improve the fit between previous beliefs and what is perceived. *(Correct answer)*
   d. Perceive reality to generate future predictions. *(Thematic distractor)*

7. **The text states that our brain constantly tries to predict the future. This is possible because the brain:** *(Global inference)*
   a. Generates stress in the face of the uncertainty. *(Unrelated distractor)*
   b. Sometimes is not able to determine what we perceive. *(Near-miss distractor)*

c. Refines its models of reality based on information from the senses. *(Correct answer)*

d. Objectively interpret information from reality. *(Thematic distractor)*

8. **In the discussion in the text about robots of the future, it is concluded that robots capable of developing human-like intelligence:** *(Text-based)*
    a. Will be built when we know how the brain works. *(Thematic distractor)*
    b. Are already a reality in the field of robotic research. *(Unrelated distractor)*
    c. Have already been created in games such as chess or go. *(Near-miss distractor)*
    d. Must be considered as a reality still quite distant. *(Correct answer)*

9. **Some researchers indicate that the body of a robot decisively modifies the learning process of the artificial intelligence system that controls it. This is because:** *(Local inference)*
    a. Its learning is determined by their bodily interactions. *(Correct answer)*
    b. Its algorithms must determine the movements of its body. *(Thematic distractor)*
    c. Its learning depends on a slower process. *(Unrelated distractor)*
    d. It can be programmed differently in each experiment. *(Near-miss distractor)*

10. **The text mentions that already existing artificial intelligence systems exceed the capabilities of humans to play chess or the Chinese game go. In this regard, the text concludes that this fact:** *(Global inference)*
    a. Demonstrates that machines can be smarter than humans. *(Thematic distractor)*
    b. Is not comparable to human capacity for spontaneous learning. *(Correct answer)*
    c. Indicates that humans make mistakes in trying to predict reality. *(Near-miss distractor)*
    d. Suggests that machines will be as smart as human with next 10–20 years. *(Unrelated distractor)*

11. **Roboticists try to create machines that mimic the spontaneous development of children. What does this contribute in their collaboration with psychologists and neuroscientists to study human development?** *(Local inference)*
    a. They create robots suitable to apply therapy to people with autism. *(Near-miss distractor)*
    b. They discover new factors that help us understand how we learn. *(Correct answer)*
    c. They identify the effects of gens on human development. *(Unrelated distractor)*
    d. They find ways to try to overcome human intelligence. *(Thematic distractor)*

12. **Studies find that children with low numeracy and computational skills struggle with:** *(Text-based)*
    a. Calculating with their eyes closed. *(Near-miss distractor)*
    b. Creating mental representations of their fingers. *(Correct answer)*
    c. Associating fingers with quantities. *(Unrelated distractor)*
    d. Calculating without using fingers. *(Thematic distractor)*

13. **In experiments in which researchers tried to teach robots vocabulary, they found that robots learned the name of an object more easily if, when naming it, it was placed:** *(Text-based)*
    a. Together with different objects. *(Thematic distractor)*
    b. Where the robot can move to look at it. *(Near-miss distractor)*
    c. Where its algorithm indicates. *(Unrelated distractor)*
    d. Every time in the same place. *(Correct answer)*

14. **According to the explanation proposed in the text for the attention deficit hyperactivity disorder, people with this disorder have trouble concentrating because:** *(Text-based)*
    a. They can't improve their predictions of reality. *(Near-miss distractor)*
    b. They perceive information from reality too quickly. *(Unrelated distractor)*
    c. They have an excessively activated nervous system. *(Thematic distractor)*
    d. They are constantly attracted to unpredictable stimulation. *(Correct answer)*

Questions excluded:

1. **"Intrinsic motivation" is the ability of a robotic system that makes it autonomously search for situations with the greatest learning potential. This determination enables it to learn:** *(Global inference)*
    a. Haphazardly when interacting with the environment. *(Correct answer)*
    b. What motivates it according to its programming. *(Thematic distractor)*
    c. Through external rewards or reinforcements. *(Unrelated distractor)*
    d. Not to make wrong predictions when interacting. *(Near-miss distractor)*

2. **To increase learning opportunities, children tend to choose learning situations in which:** *(Local inference)*
    a. They can make hypotheses that are easy to verify. *(Correct answer)*
    b. They run into something unknown. *(Near-miss distractor)*
    c. They find a wealth of stimuli. *(Thematic distractor)*
    d. They make few prediction errors. *(Unrelated distractor)*

3. **Artificial intelligence systems that incorporate a body can associate physical actions with the information they receive. This has been found to facilitate such systems:** *(Global inference)*
    a. To develop sensory capabilities. *(Unrelated distractor)*
    b. To create algorithms that determine those actions. *(Thematic distractor)*
    c. To recognize the characteristics of an image. *(Near-miss distractor)*
    d. To develop better numeracy skills. *(Correct answer)*

4. **Puppy robots from Sony are programmed in such a way that they search for different learning situations by themselves, which affects their level of exploration of the environment. This suggests that the type of skills people achieve depends on:** *(Local inference)*
    a. Both their gens and the environment. *(Thematic distractor)*
    b. The type of interactions they perform with the environment. *(Correct answer)*
    c. Their ability to reduce their prediction errors. *(Near-miss distractor)*
    d. The amount of correct predictions. *(Unrelated distractor)*

5. **It has been proposed to develop robots whose artificial intelligence system would allow a better understanding of attention deficit hyperactivity disorder. To this end, such system will be designed so that its predictive processing:** *(Global inference)*
    a. Imitates humans after learning by interacting with them. *(Correct answer)*
    b. Prefers unpredictable stimulation. *(Near-miss distractor)*
    c. Behaves according to what we know about said disorder. *(Unrelated distractor)*
    d. Is not able to pay attention when interacting. *(Thematic distractor)*

6. **In order to bring artificial intelligence systems closer to human intelligence, it is proposed that it would be necessary that their learning and development process take place as a complex "waterfall system". This is to represent that artificial intelligence needs to acquire complex knowledge:** *(Local inference)*

a. By adding simple knowledge. *(Thematic distractor)*
b. By minimizing the prediction errors. *(Near-miss distractor)*
c. Step by step, based on what is learned before. *(Correct answer)*
d. By making all predictions correctly. *(Unrelated distractor)*

7. **For both developmental psychology and robotics, a key area of learning research is knowing how humans learn to walk, speak, and recognize objects and people. This is very relevant because these are abilities:** *(Local inference)*.
   a. That are learned autonomously. *(Correct answer)*
   b. That are fundamental to our behavior. *(Thematic distractor)*
   c. That differentiate us from animals. *(Unrelated distractor)*
   d. That are essential to help others. *(Near-miss distractor)*

References

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.

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. (2015). *Words onscreen: The fate of reading in a digital world*. New York, NY: Oxford University Press.

Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M. I., & Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes [Validation of the INFLESZ Scale to assess the readability of texts addressed to patients]. *Anales del Sistema Sanitario de Navarra, 31*, 135–152. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-6627200 8000300004&lng=es&tlng=en.

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.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180–200. https://doi.org/10.1016/j.dr.2009.05.002.

Broadway, J. M., Franklin, M. S., & Schooler, J. W. (2015). Early event-related brain potentials and hemispheric asymmetries reveal mind-wandering while reading and predict comprehension. *Biological Psychology, 107*, 31–43. https://doi.org/10.1016/j.biopsycho.2015.02.009.

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

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

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*, 7. https://doi.org/10.7275/jyj1-4868.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433. https://doi.org/10.3758/BF03214546.

Daniel, D. B., & Woody, W. D. (2013). E-textbooks at what cost? Performance and use of electronic v. print texts. *Computers & Education, 62*, 18–23. https://doi.org/10.1016/j.compedu.2012.10.016.

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.

Dixon, P., & Bortolussi, M. (2013). Construction, integration, and mind wandering in reading. *Canadian Journal of Experimental Psychology, 67*, 1–10. https://doi.org/10.1037/a0031234.

Fabry, R. E., & Kukkonen, K. (2019). Reconsidering the mind-wandering reader: Predictive processing, probability designs, and enculturation. *Frontiers in Psychology, 9*, 2648. https://doi.org/10.3389/fpsyg.2018.02648.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review, 20*, 586–592. https://doi.org/10.3758/s13423-012-0367-y.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: Sage.

Fulmer, S. M., D'Mello, S. K., Strain, A., & Graesser, A. C. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary Educational Psychology, 41*, 98–110. https://doi.org/10.1016/j.cedpsych.2014.12.005.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*, 156–168. https://doi.org/10.1177/2515245919847202.

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference* (10th ed.). Boston: Pearson.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3*, 371–398. https://doi.org/10.1111/j.1756-8765.2010.01081.x.

Guthrie, J. T., Klauda, S. L., & Ho, A. N. (2013). Modelling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly, 48*, 9–26. https://doi.org/10.1002/rrq.035.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Oxon: Routledge.

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13*, 191–209. https://doi.org/10.1023/A:101666762.

Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. *Learning and Instruction, 42*, 104–112. https://doi.org/10.1016/j.learninstruc.2016.01.007.

Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015: How big is the 'mode effect' and what has been done about it? *Oxford Review of Education, 44*, 476–493. https://doi.org/10.1080/03054985.2018.1430025.

Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in Psychological Science, 21*, 348–354. https://doi.org/10.1177/0963721412454875.

Latini, N., Bråten, I., Anmarkrud, Ø., & Salmerón, L. (2019). Investigating effects of reading medium and reading purpose on behavioral engagement and textual integration in a multiple text context. *Contemporary Educational Psychology, 59*, 101797. https://doi.org/10.1016/j.cedpsych.2019.101797.

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.

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.

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation, 61*, 700–712. https://doi.org/10.1108/00220410510632040.

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, 28*, 374–405. https://doi.org/10.1080/09243453.2017.1297312.

Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, L., et al. (2019). How good is this page? Benefits and limits of prompting on adolescents' evaluation of Web information quality. *Reading Research Quarterly, 54*, 299–321. https://doi.org/10.1002/rrq.241.

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

McDonald, R. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates.

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.

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141*, 302–320. https://doi.org/10.1037/a0025250.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40–48. https://doi.org/10.1037/0021-843X.110.1.40.

Mills, C., D'Mello, S. K., & Kopp, K. (2015). The influence of consequence value and text difficulty on affect, attention, and learning while reading instructional texts. *Learning and Instruction, 40*, 9–20. https://doi.org/10.1016/j.learninstruc.2015.07.003.

Naumann, J. (2015). A model of online reading engagement: Linking engagement, navigation, and performance in digital reading. *Computers in Human Behavior, 53*, 263–277. https://doi.org/10.1016/j.chb.2015.06.051.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25*, 399–438. https://doi.org/10.1080/07370000701632371.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*, 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003.

Pernice, K., Whitenton, K., & Nielson, J. (2014). *How people read on the web: The eyetracking evidence.* Fremont, CA: Nielsen Norman Group.

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning, 4*, 3–31. https://doi.org/10.1007/s11409-008-9030-4.

Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mindwandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin, 140*, 1411–1431. https://doi.org/10.1037/a0037428.

Robitzsch, A., Lüdtke, O., Goldhammer, F., Kroehne, U., & Köller, O. (2020). Reanalysis of the German PISA data: A comparison of different approaches for trend estimation with a particular emphasis on mode effects. *Frontiers in Psychology, 11*, 884. https://doi.org/10.3389/fpsyg.2020.00884.

Rouet, J. F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist, 52*, 200–215. https://doi.org/10.1080/00461520.2017.1329015.

Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mindwandering to situational demands.

Psychonomic Bulletin & Review, 21, 1309–1315. https://doi.org/10.3758/s13423-013-0580-3.

Salmerón, L., & Delgado, P. (2019). Critical analysis of the effects of digital technologies on reading and learning. *Culture and Education, 31*, 465–480. https://doi.org/10.1080/11356405.2019.1630958.

Sanchez, C. A., & Naylor, J. S. (2018). Mindwandering while reading not only reduces science learning but also increases content misunderstandings. *Journal of Applied Research in Memory and Cognition, 7*, 332–341. https://doi.org/10.1016/j.jarmac.2018.05.001.

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6*, 339–344. https://doi.org/10.1016/S1364-6613(02)01949-6.

Seli, P., Wammes, J. D., Risko, E. F., & Smilek, D. (2016). On the relation between motivation and retention in educational contexts: The role of intentional and unintentional mind wandering. *Psychonomic Bulletin & Review, 23*, 1280–1287. https://doi.org/10.3758/s13423-015-0979-0.

Sidi, Y., Shpigelman, M., Zalmanov, H. y, & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction, 51*, 61–73. https://doi.org/10.1016/j.learninstruc.2017.01.002.

Singer Trakhman, L. M., Alexander, P. A., & Silverman, A. B. (2018). Profiling reading in print and digital mediums. *Learning and Instruction, 57*, 5–17. https://doi.org/10.1016/j.learninstruc.2018.04.001.

Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process–occurrence framework for self-generated mental activity. *Psychological Bulletin, 139*, 519–535. https://doi.org/10.1037/a0030010.

Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology, 4*, 441. https://doi.org/10.3389/fpsyg.2013.00441.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*, 946–958. https://doi.org/10.1037/0033-2909.132.6.946.

Soemer, A., & Schiefele, U. (2019). Text difficulty, topic interest, and mind wandering during reading. *Learning and Instruction, 61*, 12–21. https://doi.org/10.1016/j.learninstruc.2018.12.006.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*, 437–475. https://doi.org/10.1023/A:1009084430926.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. Harlow: Pearson Education.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*, 769. https://doi.org/10.3389/fpsyg.2016.00769.

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 832–842. https://doi.org/10.1037/a0029669.

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología, 33*, 755–782. https://doi.org/10.6018/analesps.33.3.268401.

Wechsler, D. (2008). *Wechsler Adult intelligence scale – IV*. San Antonio, TX: Pearson.

van der Weel, A. (2011). *Changing our textual minds. Towards a digital order of knowledge*. Manchester: Manchester University Press.

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

Xu, J., & Metcalfe, J. (2016). Studying in the region of proximal learning reduces mind wandering. *Memory & Cognition, 44*, 681–695. https://doi.org/10.3758/s13421-016-0589-8.