INTERNATIONAL LITERACY ASSOCIATION READING RESEARCH QUARTERLY

Strategic Text Processing Across Mediums: A Verbal Protocol Study

Natalia Latini

Ivar Bråten

University of Oslo, Norway

ABSTRACT

Using a sample of 116 Norwegian undergraduate readers in this experimental study, we investigated whether reading informational text on a tablet versus on paper would lead to differences with respect to strategic text processing and text comprehension. Strategic text processing was measured by means of verbal protocol analysis, and text comprehension was measured by means of postreading written products. Results were inconsistent with the shallowing hypothesis proposing that there are comprehension advantages for printed texts because digital texts are typically processed in a more shallow, superficial way. That is, we found no differences across the reading mediums with respect to strategic text processing or text comprehension. These results may suggest that there are boundaries to the shallowing hypothesis that are related to tasks, individual differences, and reading contexts. Such potential boundaries are discussed in light of the current findings, and suggestions for future research to clarify those boundaries are offered.

Contemporary researchers in learning and literacy largely agree that digital reading is a broad construct that includes using a plethora of information resources for a wide variety of purposes and tasks in contexts both in and out of school, with a range of individual differences coming into play and interacting with the constellations of information resources, purposes and tasks, and contexts that characterize digital reading (e.g., Alexander & the Disciplined Reading and Learning Research Laboratory [DRLRL], 2012; Bråten, Braasch, & Salmerón, 2020; Coiro, 2021; Kammerer, Brand-Gruwel, & Jarodzka, 2018; Salmerón, Strømsø, Kammerer, Stadtler, & van den Broek, 2018). Still, some of these researchers recently have been involved in efforts to clarify whether digital reading in its simplest or purest form, that is, as reading static text on a screen, may represent some additional challenges as compared with reading identical text in print (e.g., Delgado, Vargas, Ackerman, & Salmerón, 2018; Latini, Bråten, Anmarkrud, & Salmerón, 2019; Latini, Bråten, & Salmerón, 2020; Singer & Alexander, 2017; Singer Trakhman, Alexander, & Berkowitz, 2019). Of course, this is not because these researchers think that digital reading is a singular entity captured by reading static text on a screen, as some authors might suggest (Coiro, 2021), but rather because they think it is essential to understand whether there is some basic difference between reading digital and printed text when other variations across the reading mediums are controlled for. Arguably, this approach represents the opposite of a simplification of the challenges inherent in digital reading because it may reveal that even at the basic level of reading static text on a screen, digital reading may pose other challenges than does reading in print.

In the present study, we continued to pursue this line of research with determination, focusing on potential differences in strategic processing of

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 0(0)
pp. 1–22 | doi:10.1002/rrq.418

text when students read an identical informational text on a tablet and on paper. Specifically, the main purpose of this study was to investigate whether more shallow processing of digital text might lead to comprehension advantages for reading in print versus digitally, with strategic processing at different levels of depth assessed by means of verbal protocol analysis and comprehension performance assessed by means of written reports on the topic in question. To the best of our knowledge, this is the first study in this area using verbal protocol analysis to test assumptions regarding processing differences across the reading mediums. After a brief discussion of these assumptions and a review of relevant prior research, we specify the research questions and hypotheses that guided the current contribution. Because our study focused on strategic text processing as revealed by verbal protocol analysis, in particular, our background analysis also includes brief discussions of strategic text processing and verbal protocol analysis, respectively.

Theoretical Assumptions and Prior Research

In the last decade, the potentially detrimental consequences of individuals' digital media use for deeper processing and comprehension of text have received increased attention (Baron, 2015; Carr, 2011; Wolf, 2018). Among reading theorists, deeper processing and comprehension have long been regarded as a hallmark of skilled reading, required to construct a coherent mental representation of textual content that can be applied in new settings and problem-solving tasks (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; van den Broek, 2010). One reason this mode of reading has been considered endangered is that individuals may be likely to develop a more superficial, cursory way of dealing with text through their extensive use of digital media, such as through engaging with information on news, social networking, and entertainment websites. In such digital reading contexts, individuals more often than not seem to skim or skip over information in search for immediate rewards, with the danger being that this way of interacting with texts will bleed over (Wolf, 2018) and negatively affect their will and skill to engage deeply with more challenging reading tasks in which sustained attention is required for comprehension (Alexander & the DRLRL, 2012; Baron, 2015; Wolf, 2018). Specifically, Alexander and the DRLRL (2012) highlighted that students' superficial and passive interactions with digital media might carry over to reading task contexts involving higher motivational and cognitive demands "associated with the engagement in higher-order thinking, deep-level processing, and the attention to multiple intermediate goals required for the understanding and evaluation of text (Alexander et al., 2011; Fox, Maggioni, Dinsmore, & Alexander, 2008; Pressley & Afflerbach, 1995)" (p. 268).

The assumption that individuals to some extent will transfer a more superficial, cursory way of interacting with texts in many digital media contexts, such as those exemplified earlier, to the reading of all digital texts, including those that demand a deeper, more thorough processing, has been called the shallowing hypothesis (Annisette & Lafreniere, 2017). In essence, then, this hypothesis suggests that extensive use of digital media, including the internet, may strengthen a habit of mind that constrains individuals' processing and, in turn, their comprehension of digital texts more generally (Delgado et al., 2018; Latini et al., 2020). Consequently, individuals may actually be better off when trying to construct coherent mental representations from printed texts than when using digital texts with exactly the same content to complete the same reading task in the same context.

More generally, digital texts differ from printed texts in several important aspects. For example, hypertext environments represent unique affordances by connecting information across texts through embedded links, thus creating a navigable network of information that readers can access and reaccess at their own discretion (Landow, 2006). Moreover, such digital reading environments are often multimodal because they combine written language with other representations, such as spoken language, video, animations, and pictures (Mayer, 2014). Although these affordances certainly provide opportunities for learning and comprehension, they also pose new challenges for readers, such as with respect to navigation across linked information and integration across representations (e.g., Cho & Afflerbach, 2017; Scheiter & Gerjets, 2007). Still, much digital reading simply involves reading static text on a screen, such as when people read e-books for school or entertainment or when students, researchers, or professionals read downloaded text (e.g., articles in their field of study) without any hyperlinks or multimodal representations (Baron, 2020; Mizrachi, 2015; Walton, 2014). For example, students often read such texts in the form of articles, compendiums, and books that are part of the syllabus or used for assignments. This form of digital reading is also important because it is becoming increasingly common in standardized tests, including high-stakes tests (Goodwin, Cho, Reynolds, Brady, & Salas, 2020; Støle, Mangen, & Schwippert, 2020). As argued by Goodwin et al. (2020), comparing digital reading in its simplest form, that is, as reading static text on a screen, with reading on paper may also give insights into digital reading that can provide a basis for thinking about more complex digital reading environments (i.e., hypertext environments, multimodal representations). According to these authors, such comparison is therefore "part of the larger digital reading puzzle" (p. 1840) and, as such, may

contribute to building an overarching theory of new literacies (Leu, Kinzer, Coiro, Castek, & Henry, 2013).

Although the evidence is not consistent with respect to reading medium differences in comprehension, three meta-analyses (Clinton, 2019; Delgado et al., 2018; Kong, Seo, & Zhai, 2018) have bolstered the view that reading exactly the same text in print rather than digitally is advantageous in terms of comprehension performance. For example, based on a thorough review of 38 between-subjects and 16 within-subjects investigations, Delgado et al. (2018) reported an average effect size (Hedges's *g*) of 0.21 in favor of printed texts, with print advantage being greater with informational than narrative text and when reading time was restricted rather than unrestricted. Such effect sizes in the same direction were reported in the meta-analyses conducted by Clinton (2019) and Kong et al. (2018). More recent, individual studies have produced similar results. Using a long informational text, Delgado and Salmerón (2021) found that undergraduates reading in print obtained better comprehension scores than did those reading digitally, but only when reading time was restricted. Differences in comprehension performance across reading mediums were also observed in three within-subjects studies with younger students (Goodwin et al., 2020; Halamish & Elbaz, 2020; Støle et al., 2020). Interestingly, Støle et al.'s (2020) large-scale study of 10-year-old students documented that skilled comprehenders profited more from reading printed texts than did poorer or average comprehenders and that girls profited more from reading printed texts than did boys, indicating that high-performing girls had the most to lose from reading digital versions of the texts (Cohen's $d = 0.53$ for high-performing girls).

However, in two other recent studies, using between-subjects designs, Latini et al. (2019, 2020) did not find any main effect of reading medium on undergraduates' comprehension performance. Of note is that Latini and colleagues measured participants' integrated understanding by means of extended postreading writing tasks, whereas most other research on this issue has used multiple-choice and question-answering tasks (Clinton, 2019; Delgado et al., 2018).

At this point, the research base for the processing assumption of the shallowing hypothesis can be described as thin, at best. Still, there is some indirect evidence for the hypothesis in that several studies have found that participants read faster and were more likely to overestimate their comprehension when reading digital texts than when reading printed texts (Ackerman & Goldsmith, 2011; Dahan Golan, Barzillai, & Katzir, 2018; Halamish & Elbaz, 2020; Lauterman & Ackerman, 2014; Singer & Alexander, 2017; Singer Trakhman et al., 2019). It is an open question, however, whether faster reading leads to poorer calibration (i.e., overestimation), whether poorer calibration leads to faster reading, or whether the relation between reading time and calibration is bidirectional (Latini et al., 2020). Also consistent with the shallowing hypothesis, Støle et al. (2020) observed that outcomes requiring higher order reading processes were especially sensitive to reading medium differences. Moreover, Latini et al. (2019) found that when reading printed texts, but not when reading digital texts, participants took reading purpose into consideration and invested more effort in the task when reading in preparation for an exam than when reading for pleasure, with such adaptation to the reading purpose, in turn, leading to better comprehension performance.

A more direct test of the shallowing hypothesis was performed by Latini et al. (2020), who had participants read an identical illustrated text on a computer and on paper, while recording their eye movements as an indication of their integrative processing of textual and pictorial information. Although no direct or total effect of reading medium on integrated understanding was found in that study, participants reading print displayed more integrative processing during reading than did those reading on a computer. Integrative processing, in turn, was positively related to integrative understanding, indicating that reading print had an indirect, positive effect on performance via integrative processing. However, given the methodology used by Latini et al., their study could not clarify whether there were qualitative differences in processing across the reading mediums (i.e., in depth of processing), not just a quantitative one (i.e., more gaze transitions between textual and pictorial information). Likewise, Delgado and Salmerón's (2021) study, which showed that undergraduate readers reported less mind wandering (i.e., task-unrelated thoughts) when reading a printed text than when reading a digital text with restricted reading time, did not speak to the issue of qualitative differences in text processing across mediums. Finally, Goodwin et al. (2020), who found that students highlighted and annotated more when reading a printed text than when reading a digital text, actually suggested that the quality of the digital highlighting observed was superior to paper highlighting (and also contributed to comprehension performance). To test the shallowing hypothesis directly, we performed a qualitative verbal protocol analysis of participants' strategic text processing (Cho, Woodward, & Afflerbach, 2020) in the present study.

Strategic Text Processing

Strategic text processing involves effortful behavioral, cognitive, and metacognitive activities intentionally performed to improve some aspect of text-based learning or comprehension (Afflerbach, Pearson, & Paris, 2008; Bråten, Magliano, & Salmerón, 2020; Graesser, 2007). In several theories of text comprehension, such as the constructivist framework

of Graesser and colleagues (Graesser, 2007; Graesser et al., 1994) and the landscape model of van den Broek and colleagues (van den Broek, 2010; van den Broek, Young, Tzeng, & Linderholm, 1999), strategic text processing plays an important role in creating a coherent mental representation of textual content when readers are motivated to do so and automatic processes cannot do the job. Accordingly, there is ample evidence, both correlational and experimental, that strategic text processing is linked to better comprehension performance (Cho & Afflerbach, 2017; McNamara, 2007; Parris & Headley, 2015; Pressley & Harris, 2006). In particular, using deeper level strategies aimed at transforming textual information, such as by generating inferences that involve cross-text connections and connections between text and prior knowledge, has been considered to play a vital role in constructing coherence during reading (Beker, Jolles, & van den Broek, 2017; Graesser, 2007; Magliano, Millis, Ozuru, & McNamara, 2007). In comparison, surface-level strategies, such as rereading, rehearsing, and paraphrasing textual information without transforming what is already given in the text, have been considered less conducive in this regard, with such processing typically deemed more appropriate for retention and reproduction purposes than for constructing coherent understanding of textual information (Beker et al., 2017; Magliano & Millis, 2003). To what extent strategic text processing at different levels of depth is related to performance may depend on not only contextual (i.e., how performance is assessed) but also individual factors, however.

Recently, Dinsmore and Hattan (2020) performed a review of 15 previous review studies of levels of strategic processing and how different levels (i.e., deeper vs. surface level) relate to performance. In addition to demonstrating that levels of processing were not consistently defined in these reviews (if at all), Dinsmore and Hattan found it difficult to conclude regarding the relations between strategies at different levels of processing and performance, positing that "the degree to which these strategies are better or worse is…conditional" (p. 41). As noted earlier, one such condition is the way performance is assessed, that is, in terms of deeper, coherent understanding or in terms of retention or reproduction of text information. Moreover, Dinsmore and Hattan suggested that individual differences in competence, specifically in prior domain knowledge, may determine the effectiveness of different levels of processing, with efforts to use deeper level strategies when prior knowledge is low potentially hindering rather that helping with regard to comprehension performance (e.g., because it leads to incorrect inferences). Thus, in accordance with the model of domain learning (Alexander, 1997, 2005), readers could probably profit from the use of surface-level strategies to gain foundational understanding, or from a mix of surface-level and deeper strategies, unless they have high competence or expertise in a domain (which is characterized by high prior knowledge). In the same vein, reading researchers have suggested that paraphrasing may indicate efforts to comprehend and play an important role in text comprehension, especially when combined with other, deeper level strategies (Coté, Goldman, & Saul, 1998; van den Broek, Lorch, Linderholm, & Gustafson, 2001).

Another issue raised in Dinsmore and Hattan's (2020) review of reviews concerned the measurement of strategies at different levels of processing, with most previous research in this area relying on offline self-report inventories. In the present study, we took a different tack and tried to measure strategies at different levels of processing by means of verbal protocol analysis when students read the same text on a tablet and on paper.

Verbal Protocol Analysis

Verbal protocol analysis involves having readers think aloud as they read to create a verbal protocol that subsequently can be analyzed by the researchers (Cho et al., 2020; Ericsson & Simon, 1993). Typically, readers are instructed to verbalize all thoughts related to reading that are accessible in working memory, with such thoughts considered to represent effortful meaning-making activities that can be analyzed in terms of strategic text processing (Cho et al., 2020; Pressley & Afflerbach, 1995; Trabasso & Magliano, 1996). When readers are asked to verbalize their thoughts whenever they come to mind during reading, a concurrent verbal protocol becomes available; however, when readers are asked to think aloud after the reading of a particular segment of text (e.g., a sentence, a paragraph), researchers work with a retrospective verbal protocol in analyzing strategic text processing (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995; Trabasso & Magliano, 1996). In the latter case, it is important that processing has occurred so recently that the products of that processing remain in working memory. Accordingly, Ericsson and Simon (1993) concluded that the closer in time to reading individuals verbalized their thoughts, the more likely traces of that processing would remain in working memory and could be reported (see also Pressley & Afflerbach, 1995). Ericsson and Simon also reported that there were few differences between verbal protocols based on concurrent and retrospective verbalization when retrospective reporting was done immediately after a few sentences or a short paragraph was read. This was supported by Pressley and Afflerbach (1995), who claimed that

> it may simply be impossible to report what one is doing with respect to what is being read right at this instance. If that is so, there should be very little difference between what researchers claim are concurrent and what they view as briefly retrospective reports. (p. 130)

Coding systems used in analyzing verbal protocols have focused on different types of inferences that can support the construction of a coherent mental representation

and, as such, reflect deeper level text-processing strategies (e.g., Kendeou & van den Broek, 2007; Kopatich, Magliano, Millis, Parker, & Ray, 2019; Linderholm & van den Broek, 2002; Magliano et al., 2020; Trabasso & Magliano, 1996; van den Broek et al., 2001). Moreover, these systems have typically included categories reflecting metacognitive processing (e.g., monitoring of one's understanding) and evaluation of different aspects of the text (e.g., the believability of text content), as well as surface-level processing such as associations, text repetitions, and paraphrases (e.g., Coté et al., 1998; Kendeou & van den Broek, 2007; Linderholm & van den Broek, 2002).

Recently, Bråten, Magliano, and Salmerón (2020) reviewed three different approaches used to validate verbal protocol analysis, such as by correlating text-processing strategies derived from verbal protocol analysis with individual differences relevant to comprehension, showing that there are good reasons to consider verbal protocol analysis an effective tool for gaining insights into readers' strategic text processing. At the same time, however, it is difficult to exclude the possibility that oral expression of reading-related thoughts may compete for readers' limited cognitive resources, especially when readers lacking competence or expertise in the domain are asked to perform complex tasks (Schraw, 2010). When readers are asked to retrospectively verbalize their thoughts, it is also a possibility that verbal protocol analysis is prone to fabrication because readers are reporting other (e.g., deeper) processing than what actually occurred during reading (Bråten, Magliano, & Salmerón, 2020). Despite such possible confounds, it could be argued that verbal protocol analysis is particularly well suited to measure strategic processing at different levels of depth and relate such qualitatively different strategies to the construction of mental representations during text comprehension (Cho et al., 2020). Accordingly, verbal protocol analysis has been used effectively to distinguish between deeper level strategies (i.e., inferences) that promote coherence building and locally focused surface-level strategies (i.e., paraphrases) directed at understanding individual sentences (Magliano & Millis, 2003).

The Present Study

Based on the preceding theoretical and empirical background analysis, we set out to investigate whether undergraduates who read an identical informational text on a tablet or on paper would differ with respect to strategic text processing and text comprehension. To assess participants' strategic text processing, we asked them to think aloud immediately after having read each of 10 paragraphs, and to assess their text comprehension, we asked them to write a report on the topic in question based on their mental representation of the text content (i.e., without the text available). To ensure that participants' comprehension performance was not influenced by the think-aloud procedure, we included two additional groups of participants that read the same text on a tablet or on paper without thinking aloud. One possibility is, for example, that the think-aloud procedure might deflate any reading medium effects on comprehension performance because it would lead to more thorough text processing in both medium conditions. Conversely, any reading medium effects on comprehension might be inflated because participants in the print condition would use the prompt to retrospectively verbalize their thoughts to increase an already existing processing divide between the medium conditions. Thus, to exclude these possibilities, we used a 2 (digital, print) × 2 (think-aloud, no think-aloud) between-subjects design to analyze reading medium effects on comprehension performance in the present study.

In accordance with the shallowing hypothesis (Annisette & Lafreniere, 2017) and recent meta-analyses supporting this hypothesis (Clinton, 2019; Delgado et al., 2018; Kong et al., 2018), as well as with work within verbal protocol analysis indicating that significant reactivity (i.e., changes in task-related processes and performance) is unlikely when using this methodology (Ericsson & Fox, 2011; Fox, Ericsson, & Best, 2011), we expected that participants in the print condition would display better comprehension than those in the digital condition regardless of whether they thought aloud. In addition, we expected that among the participants who were asked to think aloud, participants reading in print would differ from those reading digitally with respect to strategic text processing, particularly with respect to deeper level strategies as revealed by inferences assumed to support mental model construction (e.g., backward, forward, and elaborative inferences; van den Broek, Fletcher, & Risden, 1993). We also grounded this hypothesis in the shallowing hypothesis (Annisette & Lafreniere, 2017) and prior empirical work consistent with this hypothesis (e.g., Ackerman & Goldsmith, 2011; Dahan Golan et al., 2018; Delgado & Salmerón, 2021; Halamish & Elbaz, 2020; Latini et al., 2020), which suggested that a lack of deep reading and, hence, more reliance on superficial text processing when reading on-screen is responsible for observed reading medium differences in comprehension performance. However, this is the first study to directly test the shallowing hypothesis by examining qualitatively different forms of text processing across the two reading mediums. In accordance with the shallowing hypothesis, we also expected that participants' text processing, as revealed by the verbal protocol analysis, would mediate the potential effect of reading in print versus digitally on participants' comprehension performance, as revealed by their postreading written products. Because not only deeper level strategies but also a mix of surface-level and deeper

(including metacognitive) strategies may be adaptive for nonexpert readers (Dinsmore & Hattan, 2020), we also explored the possibility that participants would differ across mediums in a combination of such strategies.

Finally, we included participants' working memory capacity, basic reading comprehension skills, prior knowledge, and task-based intrinsic motivation as potential covariates. These individual difference variables may influence text processing and comprehension performance (e.g., McNamara & Magliano, 2009; Schiefele, Schaffner, Möller, & Wigfield, 2012; Swanson & Alloway, 2012), so we wanted to rule out the possibility that any effects of our experimental manipulations depended on them.

Method

Participants

The sample consisted of 116 undergraduates in educational sciences and humanities at a large public university in southeast Norway. Most participants (78.4%) were enrolled in the first year of a bachelor program, and 15.5% and 6.1%, respectively, were enrolled in the second or third year. Participants' mean age was 21.95 years (*SD* = 2.88 years), and 77.6% were female. Most participants (89%) had Norwegian as their first language, and the rest were bilingual. In regard to reading medium, participants self-reported a preference for printed rather than digital reading materials in both the study context and their leisure time.[1]

We recruited participants in regular lectures. Those who volunteered received a gift card worth NOK 200 (approximately US$20) after the data collection. Collection and handling of all data met the requirements of the Personal Data Registers Act in Norway and were approved by the Norwegian Social Science Data Services.

Materials

Text and Experimental Manipulations

Each participant read one 10-paragraph informational text titled "Phobias" (in Norwegian), which was based on an encyclopedia on phobias (Milosevic & McCabe, 2015) in addition to diverse popular science articles dealing with this topic. Both the length and the language of these original texts were adapted to form a single text consisting of 1,000 words (approximately 100 words per paragraph). On a separate title page, source information (author's name and credentials plus venue and date of publication) was presented in addition to the title. The text was said to be authored by a female psychologist with a common Norwegian name and published in the *Journal of the Norwegian Psychological Association* in 2019.

The 10 paragraphs of the text were presented on 10 pages (one paragraph per page) and covered three main themes: what phobias are, why phobias occur, and how phobias can be treated. The first theme was covered in the first four paragraphs, with the first paragraph describing fear as a natural response to real threats that can result in fight-or-flight reactions, the second paragraph discussing phobias in terms of irrational fear that interferes with everyday life, the third paragraph describing the main categories of phobias (i.e., social and specific phobias) and their characteristics, and the fourth paragraph describing and explaining two less familiar phobias (i.e., phonophobia, trypophobia). The second theme was covered in the next four paragraphs, with the fifth paragraph discussing genetic causes of phobia and relating them to autonomous nervous system vulnerability and lower levels of an inhibitory neurotransmitter, the sixth paragraph discussing how phobias can be learned through the mechanisms of classical conditioning and observational learning, the seventh paragraph discussing how phobias may have an evolutionary origin and be derived from fears that increased the chances of survival in the distant past, and the eighth paragraph discussing gender difference in the prevalence of phobias and the potential contribution of stereotypical gender roles in this regard. Finally, the third theme was covered in the last two paragraphs, with the ninth paragraph discussing medical treatment by means of sedative drugs or cortisol and the 10th paragraph discussing psychological treatment in the form of traditional or virtual reality exposure therapy. Taken together, the 10 paragraphs of the text thus provided an elaborated overview of the characteristics, causes, and treatments discussed in the literature on phobias. An English version of the text is provided in Appendix A.

We used Björnsson's (1968) formula, based on word and sentence length, to compute the readability of the text. This resulted in an average readability estimate of 48.00 (*SD* = 9.99) for the 10 paragraphs, indicating that the difficulty level of the text was comparable to that of informational texts from the Norwegian government (Vinje, 1982).

We manipulated the reading medium between participants, with participants randomly assigned to read the 10 paragraphs of text either in print or digitally. In the print condition, the text was presented in a stapled booklet using A5 (5.8 × 8.3 in.) size, 14-point Times New Roman font, and one-sided printing with one paragraph printed on each page. The spacing was 1.15, and the text was fully justified. In the digital condition, the same text was presented as a PDF file on a Samsung Galaxy Tab A with a 10.1-inch screen equivalent to A5 size at a resolution of 1920 × 1200 pixels and a font size equivalent to 14 point. One entire paragraph was available on the screen at once. To turn the page, participants had to swipe horizontally from right to left across the screen. Apart from the reading medium, the appearance of the text was therefore identical across the two conditions. In addition to this reading medium manipulation, we randomly assigned participants either to a

think-aloud condition in which they were asked to verbalize their thoughts after each paragraph or to a condition in which no think-aloud prompts were given (see the Procedure section).

Dependent Measures

To compare participants' comprehension performance when reading in print versus digitally, we used their scores on written reports about the topic as a dependent measure. To compare their text processing across these two conditions, we used categories of think-aloud comments included in verbal protocols as dependent measures. In the following, we describe these dependent measures and how they were scored, as well as how inter-rater reliabilities were established.

Comprehension Performance

To measure comprehension of the text content, we asked participants to write a report in which they discussed what phobias are, why they occur, and how they are treated. In each of the 10 paragraphs, we identified the main idea and awarded participants 0 or 1 point depending on whether this main idea was correctly represented in their written products. In addition, in each paragraph, we determined how the main idea was further elaborated, and awarded participants 0 or 1 point depending on whether this elaboration was present in their written products. For example, in the paragraph defining phobias (i.e., paragraph 2), we identified the main idea as phobias involve irrational fear, and an elaboration of this idea as phobias interfere with adaptive functioning. The scores for each paragraph thus ranged from 0 to 2, and the scores on the entire measure could possibly range from 0 to 20. Only the total scores were used in subsequent statistical analyses, with these reflecting the extent to which participants had represented an elaborated overview of the characteristics, causes, and treatments of phobias, as presented across the 10 paragraphs. The scoring system is described in detail and further exemplified in Appendix B.

We both scored the written responses blind to experimental conditions. First, 24 participants' responses were scored in collaboration. Next, a random selection of 30 participants' responses (i.e., 25%) was scored independently, resulting in a high inter-rater reliability estimate (Cohen's $\kappa = .93$). Also, our total scores for these 30 participants were highly correlated (Pearson's $r = .95$, $p < .01$). All disagreements were resolved in discussion, and we scored the remaining participants' responses separately.

Text Processing

The audio-recorded verbal protocols of participants in the think-aloud condition were transcribed and segmented into units of analysis, with a unit of analysis defined as a comment or set of comments on the same phrase, sentence, or group of sentences within a paragraph (e.g., Chi, De Leeuw, Chiu, & Lavancher, 1994; Coté et al., 1998; Strømsø, Bråten, & Samuelstuen, 2003). These units were coded into one of eight categories of text processing, with this coding system based on prior work by van den Broek and colleagues (e.g., Kendeou & van den Broek, 2007; Linderholm & van den Broek, 2002; van den Broek et al., 2001) and Magliano and colleagues (e.g., Kopatich et al., 2019; Magliano et al., 2020; Trabasso & Magliano, 1996).

The category of associations included comments made in response to text content that referred to prior knowledge or experience without facilitating understanding or creating coherence. Paraphrases involved comments that repeated or reworded text content. Backward inferences included comments that connected content in the current paragraph with content in one or more preceding paragraphs. Elaborative inferences involved comments that connected text content with relevant prior knowledge and experience that facilitated understanding and increased coherence. Predictive inferences were comments that anticipated content in one or more forthcoming paragraphs based on content in the current paragraph. Monitoring included comments that involved reflections on one's own thinking about text content or one's own (lack of) understanding or knowledge of text content. Evaluation included comments that were opinions about or affective responses to text content. Finally, comments that did not fit into any of these categories and did not facilitate understanding or contribute to coherence were coded as other. The entire coding system for the verbal protocols is described and exemplified in Appendix C.

We both coded the verbal protocols blind to reading medium condition. First, 14 participants' protocols were coded in collaboration. Next, a random selection of 23 participants' protocols (i.e., 20%) was coded independently, resulting in a substantial inter-rater reliability (Cohen's κ) of .70. All disagreements were resolved in discussion, and we coded the remaining participants' verbal protocols separately.

Because we were interested in the distinction between deeper and surface-level processing strategies in the present study, we recoded the three types of inferences described earlier (i.e., backward, elaborative, predictive) into a broader category of inferences indicating deeper level strategies (e.g., Coté et al., 1998; Kendeou & van den Broek, 2007) while retaining the category of paraphrases as an indication of surface-level strategies (e.g., Beker et al., 2017; Magliano & Millis, 2003). As explained previously (see the Strategic Text Processing section), deeper level strategies aim at transforming textual information by generating inferences involving cross-text connections and connections between text and prior knowledge, whereas surface-level strategies involve a more superficial engagement with text that is suitable for reproduction of information in the same or similar form (Bråten, Magliano, & Salmerón, 2020). Moreover, we

recoded the comments categorized as monitoring and evaluation, respectively, into the broader category of monitoring, indicating evaluations of text content and one's own understanding and knowledge of that content (e.g., Coté et al., 1998; Strømsø et al., 2003). The categories associations and other were combined into a broader category of irrelevant processing. An estimation of inter-rater reliability based on independent coding of 20% of the verbal protocols yielded a high Cohen's kappa of .83 for these four categories of text processing. Finally, we created a broad category including a mix of surface-level and deeper strategies by combining the categories of paraphrases, inferences, and monitoring, which might be particularly adaptive for nonexpert readers (Dinsmore & Hattan, 2020). We only used the four categories described in this paragraph (i.e., inferences, paraphrases, monitoring, irrelevant processing), together with the mixed processing category, in subsequent statistical analyses.

Covariates

In the following, we describe the measures of working memory, reading comprehension skills, prior knowledge, and task-based intrinsic motivation, which we included as potential covariates in the present study.

Working Memory

We used a 12-item Norwegian adaptation of Swanson and Trahan's (1992) working memory span task to measure working memory. Each item consisted of a set of unrelated sentences that was read aloud to participants with an interval of two seconds between each sentence. Across the 12 items, the number of sentences in a set increased from two to five. Participants were asked to comprehend the sentences and answer a question about the content of an unknown sentence as soon as the final sentence in a set was read. Also, on the same response sheet, they were asked to write down the final word of each sentence. Thus, in accordance with current conceptualizations of working memory (e.g., Baddeley, 2001), this task required simultaneous processing and storage of information. For each item, scoring was done by awarding 1 point for a correct answer to the comprehension question and 1 additional point for each of the final words that was recalled correctly. If a question was not answered or was answered incorrectly, that item was scored 0 regardless of how many words were recalled correctly. The possible range of scores on this measure was 0–54. The internal consistency reliability (Cronbach's $\alpha = .56$) was lower than desirable, presumably due to the high difficulty level of some of the items.

Reading Comprehension

We used a Norwegian version of a cloze test developed in Danish by Gellert and Elbro (2013) to measure basic reading comprehension skills. This Norwegian version has been validated in several recent studies (e.g., Bråten, Brante, & Strømsø, 2019; Latini et al., 2020). Participants read five narrative and five expository texts that ranged in length from 40 to 330 words and had a total length of 1,340 words. The 10 texts had 41 word gaps altogether, with four alternatives provided for each gap, and participants were asked to read the texts and fill as many of these gaps as possible during a period of 10 minutes. One point was awarded for each correctly filled gap (i.e., the possible range of scores was 0–41). Of note is that correct filling of all gaps required an understanding of the ideas in the text and some form of inferencing (sample item: Your skin may also become dry during the flight. Therefore bring [**water—ear plugs—medicine—cream**] on longer flights). The internal consistency reliability (Cronbach's α) for participants' scores was .77.

Prior Knowledge

We measured prior knowledge about phobias by asking participants to respond in writing to four open-ended questions: (1) What is a phobia? (2) Which types of phobias do you know? (3) Do you know why some people have phobias? (4) Do you know how phobias can be treated? On the first question, scores were based on the definition of phobias by Milosevic and McCabe (2015), which included four aspects: A phobia is a fear of something, it is irrational, it is intense and enduring, and it is maladaptive. Participants were awarded 1 point if they included one of these aspects, 2 points if they included two of these aspects, and 3 points if they included three or four of these aspects. To receive a score of 3 on this question, participants had to include the aspect of irrationality, though. On the second question, 1 point was awarded if participants included one or more phobias from one of the following six categories, which commonly figure in the literature on phobias (Milosevic & McCabe, 2015): animals/insects; nature (e.g., heights, water, darkness); blood, injection, or injury (e.g., needles, illness); situations (e.g., flying, elevators, driving, small rooms); social phobias; and others.[2] To receive 2 points, participants had to include phobias from two or three of these categories, and to receive 3 points, participants had to include phobias from four or more categories. On the third question, 1 point was awarded for including one of the following potential causes of phobias: genetics, learning (including traumatic experiences), evolutionary mechanisms, and gender role stereotypes. To receive 2 points, participants had to include two of these potential causes, and to receive 3 points, participants had to include three or four of them. Finally, on the fourth question, 1 point was awarded if participants included one of the three most common ways of treating phobias (i.e., exposure therapy, medication, conversation-based therapy). If participants included two or three of these treatments, 2 or 3

points were awarded. The possible range of scores was thus 0–3 on each question and 0–12 on the entire prior knowledge measure.

We both scored participants' responses to the four questions blind to experimental conditions. First, 12 participants' responses were scored in collaboration. Then, we scored a random selection of 30 participants' responses (25%) independently, resulting in an inter-rater reliability coefficient (Pearson's *r*) of .92. Disagreements were resolved in discussion. We scored the remaining participants' responses separately.

Task-Based Intrinsic Motivation

A five-item inventory completed immediately after reading the text (and before writing the report) was used to measure participants' task-based intrinsic motivation. This measure has been validated in prior research with both sixth-grade students (Bråten, Johansen, & Strømsø, 2017) and undergraduates (Latini et al., 2020). Participants used a 5-point scale (1 = *does not fit at all*; 5 = *fits very well*) to rate to what extent they had experienced reading the text as exciting, interesting, fun, attractive, and enjoyable, respectively. The internal consistency reliability (Cronbach's α) for participants' scores was .79.

Procedure

The first author collected the data in individual 60-minute sessions in a quiet room at the university. On arrival, participants were sequentially assigned to one of the four experimental conditions: print/think-aloud (*n* = 29), print/no think-aloud (*n* = 29), digital/think-aloud (*n* = 29), and digital/no think-aloud (*n* = 29). After participants completed the working memory measure, which was administered orally, they received a folder containing a demographic survey, the reading comprehension measure, and the prior knowledge measure and completed these materials on paper in this order. Participants assigned to the think-aloud conditions (i.e., print/think aloud, digital/think aloud) were then given a task in which they practiced thinking aloud as they read a three-paragraph text about schizophrenia. This text was similar to the experimental text in terms of layout, paragraph length, and writing style. This also means that participants in the print/think-aloud condition read the practice text printed in a booklet, whereas participants in the digital/think-aloud condition read the practice text on a tablet (see the Text and Experimental Manipulations section). Before reading the practice text, participants were given the following oral instruction:

> In this investigation, you will read 10 paragraphs of text, and after each paragraph, you are going to say aloud everything you are thinking about what you are reading. But first, you will get a practice task in which you read three paragraphs on another topic. After reading a paragraph silently, you turn the page, and on that page, you will be asked to verbalize everything you are thinking about what you are reading. Nothing is right or wrong to say; just say everything you are thinking about what you are reading. If you do not say anything, I will ask you to talk. When you have said everything you are thinking about what you are reading, you turn the page and continue reading silently. Do you understand what you are going to do?

Participants were reminded to talk if they remained silent for more than three seconds ("Don´t forget to think aloud"). After the practice session, participants in the print/think-aloud and digital/think-aloud conditions read the following instruction on paper:

> You are now going to read a text about phobias in order to write a brief report in which you discuss what phobias are, why they occur, and how they are treated. The text consists of 10 paragraphs, with one paragraph on each page. When you have read a page, you will be asked to say aloud everything you are thinking about what you are reading. You cannot look back to the text while thinking aloud.

Participants in the print/no-think-aloud and digital/no-think-aloud conditions read the same instruction except for the last two sentences. Moreover, participants in the no-think-aloud conditions were instructed that they could spend up to eight minutes reading the entire text, whereas participants in the think-aloud conditions were instructed that they could spend up to 15 minutes. This difference was based on piloting of the materials, which indicated that a time limit of eight minutes would allow all students to read the entire text and that the average additional time used to think aloud was approximately seven minutes. These time limits were intended to make participants in all conditions experience a certain time pressure because reading medium effects have been found to be more pronounced when reading times are restricted (Delgado et al., 2018; see also Delgado & Salmerón, 2021).

Participants reading in print were informed that the text was available in a folder labeled "Text" within the folder containing the other materials, and participants reading digitally were informed that the text was available as a clickable file labeled "Text" on the home screen of the tablet. Participants were informed that when they had finished reading the entire text, they should put the text back in the folder (when reading in print) or return to the home screen (when reading digitally). When participants in the think-aloud conditions had read a paragraph and turned the page in the booklet (when reading in print) or swiped the screen (when reading digitally), only the think-aloud prompt "What are you thinking about what you are reading?" appeared on the next page. After finishing thinking aloud, participants again turned the page or swiped the screen to access the next paragraph, and so forth. For participants in the no-think-aloud conditions, the 10 interspersed pages containing only the think-aloud prompts were not included in the booklet/digital file, of course.

Immediately after participants had finished reading the text or reached the time limit of eight or 15 minutes, they completed the inventory on task-based intrinsic motivation on paper. Then, all participants were provided with a laptop computer on which they accessed a web-based questionnaire by clicking on a Google Chrome window located on the taskbar. This questionnaire contained the following written task instruction:

> Based on the text you just read, you are going to write a brief report in which you discuss what phobias are, why they occur, and how they are treated. You can spend as much time as you want on this writing task. It is important that you express yourself as completely and elaborately as you can.

Below this instruction, participants wrote their report in a separate text entry box with no word limit. The task instruction was visible during writing, but participants could not reaccess the text. When finished, they submitted their report to a server by clicking a "Send" button.

Results

In this section, we first present the results for comprehension performance based on the entire sample of 116 participants. Then, we present the results for text processing based on the subsample of 58 participants for which verbal protocol data were collected and analyzed.

Effects on Comprehension Performance

Table 1 presents descriptive information and zero-order correlations for all measured variables for the entire sample. As can be seen, participants' prior knowledge of phobia was rather low, and on average, they obtained moderate scores with regard to comprehension performance. However, all the individual difference measures except task-based intrinsic motivation were positively and statistically significantly correlated with comprehension performance. Interestingly, participants reported a high level of intrinsic motivation for the reading task despite their limited prior knowledge of the text's topic and their modest performance on the writing measure.

TABLE 1
Descriptive Statistics and Zero-Order Correlations for All Measured Variables for the Entire Sample (*n* = 116)

Variable	1	2	3	4	5
1. Working memory	—				
2. Reading comprehension	.20*	—			
3. Prior knowledge	.25**	.32**	—		
4. Task-based intrinsic motivation	.00	.10	.04	—	
5. Comprehension performance	.21*	.26**	.35**	.12	—
M	27.98	30.13	5.50	4.13	7.48
SD	7.94	5.30	1.54	0.65	3.02
Skewness	0.01	−0.30	0.13	−1.13	0.39

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

Table 2 displays descriptive information about the individual difference variables and comprehension performance for the four subgroups differing with regard to reading medium (print or digital) and think-aloud prompt (think-aloud or no think-aloud). We performed one-way analyses of variance (ANOVAs) with the four subgroups as the independent variable and working memory, reading comprehension, prior knowledge, and task-based intrinsic motivation, respectively, as the dependent variables. Results showed no statistically significant differences among the subgroups: for working memory, $F(3, 111) = 1.67$, $p = .177$, $\eta^2 = .043$; for reading comprehension, $F(3, 112) = 0.41$, $p = .750$, $\eta^2 = .011$; for prior knowledge, $F(3, 112) = 0.84$, $p = .475$, $\eta^2 = .022$; and for task-based motivation, $F(3, 112) = 0.22$, $p = .879$, $\eta^2 = .006$.

To test our hypothesis regarding the effect of reading medium on comprehension performance, we performed a 2 × 2 between-subjects analysis of covariance (ANCOVA) with reading medium (print or digital) and think-aloud prompt (think-aloud or no think-aloud) as the independent variables and comprehension performance as the dependent variable. Covariates in this analysis were working memory, reading comprehension, and prior knowledge (task-based motivation was not included because it did not correlate with the dependent measure; Tabachnick & Fidell, 2014). Results of the evaluation of the assumptions for performing the ANCOVA were satisfactory.

Results of the ANCOVA showed that there was no statistically significant main effect of reading medium on comprehension performance (print: $M = 7.78$, standard error [*SE*] = 0.38; digital: $M = 7.23$, $SE = 0.37$; $F(1, 108) = 1.02$, $p = .316$, $\eta_p^2 = .009$). Also, there was no statistically significant main effect of thinking aloud on performance (think-aloud: $M = 7.50$, $SE = 0.37$; no think-aloud: $M = 7.51$, $SE = 0.38$), $F(1, 108) = 0.01$, $p = .981$, $\eta_p^2 = .000$; nor was there any statistically significant interaction between reading medium and think-aloud prompt, $F(1, 108) = 0.63$, $p = .428$, $\eta_p^2 = .006$. The effect of the covariate of prior knowledge, $F(1, 108) = 9.72$, $p = .002$, $\eta_p^2 = .083$, but not the effects of working memory, $F(1, 108) = 1.98$, $p = .162$, $\eta_p^2 = .018$, or reading comprehension, $F(1, 108) = 2.38$, $p = .126$, $\eta_p^2 = .022$, were statistically significant.

Thus, contrary to what we expected, there was no effect of reading medium on comprehension performance. Also, our analysis determined that this lack of effect was independent of whether participants were thinking aloud about what they were reading. The results

TABLE 2
Descriptive Information for Subgroups Differing With Respect to Reading Medium and Think-Aloud Prompt

Variable	Print		Digital	
	Think-aloud *M* (*SD*) (*n* = 29)	No think-aloud *M* (*SD*) (*n* = 29)	Think-aloud *M* (*SD*) (*n* = 29)	No think-aloud *M* (*SD*) (*n* = 29)
Working memory	25.66 (7.37)	27.25 (8.30)	29.21 (7.90)	29.79 (7.88)
Reading comprehension	29.62 (5.69)	30.76 (5.22)	29.55 (5.82)	30.59 (4.57)
Prior knowledge	5.48 (1.41)	5.14 (1.48)	5.66 (1.74)	5.72 (1.51)
Task-based intrinsic motivation	4.18 (0.44)	4.05 (0.81)	4.16 (0.77)	4.14 (0.54)
Comprehension performance	7.83 (2.88)	7.24 (3.48)	7.14 (3.04)	7.72 (2.72)

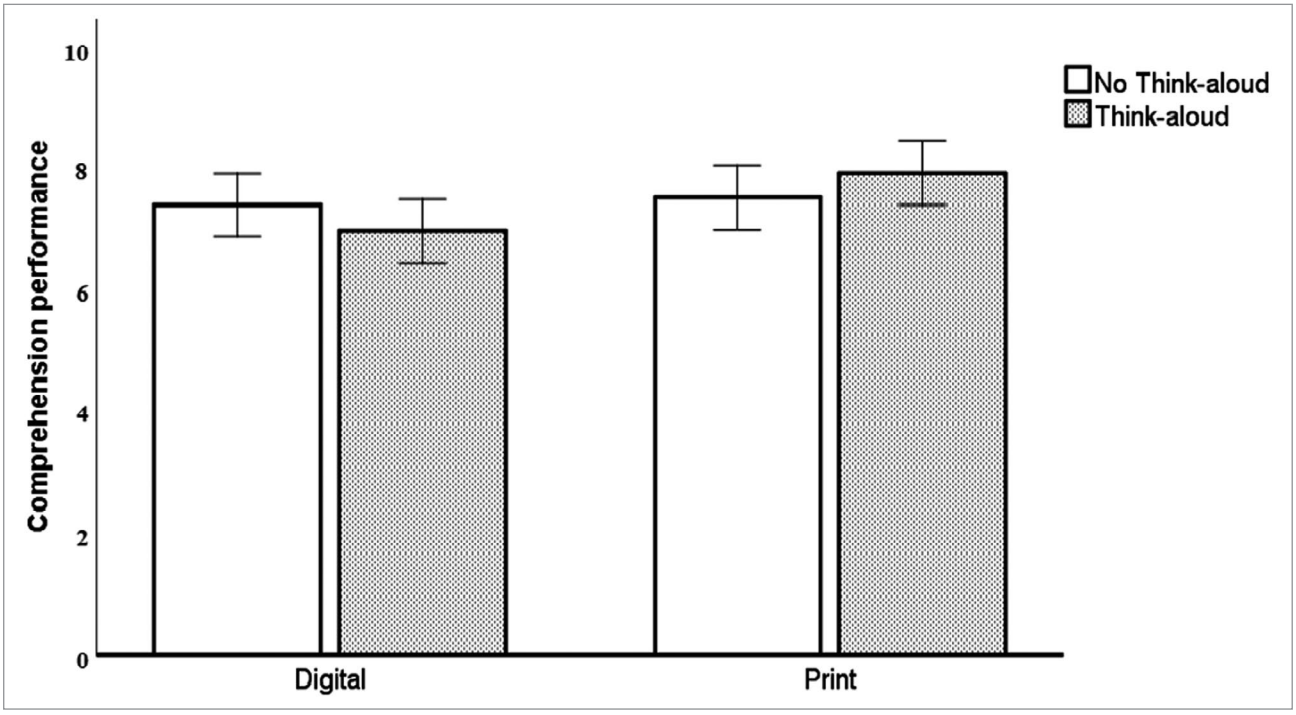
of our analysis regarding comprehension performance are displayed in Figure 1.

Effects on Text Processing

Table 3 includes descriptive information and zero-order correlations for the individual difference measures, the categories of strategic text processing, and comprehension performance for participants who were instructed to think aloud (i.e., the print/think-aloud and digital/think-aloud conditions). As can be seen, the category of irrelevant processing was negatively and statistically significantly ($r = -.37$, $p = .005$) correlated with comprehension performance, and the category of mixed processing was positively and statistically significantly ($r = .37$, $p = .004$) correlated with comprehension performance, whereas the positive correlation between paraphrases and comprehension performance ($r = .25$, $p = .055$) did not quite reach a conventional level of statistical significance with this sample size. The other categories of strategic processing (i.e., inferences, monitoring) had low positive and statistically nonsignificant correlations with comprehension performance.

Table 4 includes descriptive information about the text-processing variables based on verbal protocols for the two think-aloud groups. Regarding inferences, monitoring, and irrelevant processing, we performed a set of one-way ANOVAs with reading medium as the independent variable to examine whether participants reading in print

FIGURE 1
Estimated Marginal Means for Comprehension Performance for Each Reading Medium by Think-Aloud Condition



Note. Error bars represent standard errors.

TABLE 3
Descriptive Statistics and Zero-Order Correlations for All Measured Variables for Participants in the Think-Aloud Condition (*n* = 58)

Variable	1	2	3	4	5	6	7	8	9	10
1. Working memory	—									
2. Reading comprehension	.04	—								
3. Prior knowledge	.22	.33*	—							
4. Task-based intrinsic motivation	−.13	.01	−.07	—						
5. Irrelevant	.11	−.13	−.09	.07	—					
6. Paraphrases	.09	.10	.32*	−.15	−.45**	—				
7. Inferences	−.04	.17	.16	.20	−.02	−.26†	—			
8. Monitoring	.06	.06	−.10	.23	.17	−.69**	.44**	—		
9. Mixed processing	.14	.23	.40**	.06	−.45**	.63**	.43**	.05	—	
10. Comprehension performance	.13	.31*	.44**	.12	−.37**	.25†	.12	.01	.37**	—
M	27.43	29.59	5.57	4.17	3.09	8.62	3.00	5.62	17.24	7.48
SD	7.78	5.70	1.57	0.62	3.03	8.77	3.14	5.01	7.17	2.96
Skewness	0.21	−0.13	0.39	−1.19	1.74	0.67	1.53	0.67	0.27	0.35

†$p = .055$. *$p < .05$. **$p < .01$.

TABLE 4
Descriptive Information About the Text-Processing Variables for the Think-Aloud Groups

	Print *M* (*SD*) (*n* = 29)	Digital *M* (*SD*) (*n* = 29)
Irrelevant	3.00 (3.01)	3.17 (3.11)
Paraphrases	8.48 (9.04)	8.76 (8.64)
Inferences	2.69 (2.94)	3.31 (3.35)
Monitoring	4.83 (3.91)	6.41 (5.87)
Mixed processing	16.00 (6.54)	18.48 (7.65)

differed from those reading digitally with respect to these types of text processing. Regarding paraphrases and mixed processing, we performed one-way ANCOVAs with reading medium as the independent variable, using prior knowledge as a covariate because this individual difference variable was positively correlated with the dependent variables (i.e., paraphrases, mixed processing). Results of evaluation of the assumptions for performing these analyses were satisfactory.

Results of the ANOVAs showed no statistically significant effects of reading medium on inferences (print: $M = 2.69$, $SD = 2.94$; digital: $M = 3.31$, $SD = 3.35$), $F(1, 56) = 0.56$, $p = .457$, $\eta^2 = .010$; monitoring (print: $M = 4.83$, $SD = 3.91$; digital: $M = 6.41$, $SD = 5.87$), $F(1, 56) = 1.47$, $p = .231$, $\eta^2 = .026$; or irrelevant processing (print: $M = 3.00$, $SD = 3.01$; digital: $M = 3.17$, $SD = 3.11$), $F(1, 56) = 0.05$, $p = .831$, $\eta^2 = .001$. Moreover, the ANCOVAs showed no statistically significant effects of reading medium on paraphrases (print: $M = 8.64$, $SE = 1.57$; digital: $M = 8.60$, $SE = 1.57$), $F(1, 55) = 0.00$, $p = .987$, $\eta_p^2 = .000$, or mixed processing (print: $M = 16.15$, $SE = 1.23$; digital: $M = 18.33$, $SE = 1.23$), $F(1, 55) = 1.57$, $p = .215$, $\eta_p^2 = .028$. In both ANCOVAs, the covariate of prior knowledge had a statistically significant effect: for paraphrases, $F(1, 55) = 6.41$, $p = .014$, $\eta_p^2 = .104$, and for mixed processing, $F(1, 55) = 10.17$, $p = .002$, $\eta_p^2 = .156$.

Thus, contrary to what we expected, there was no effect of reading medium on text processing, neither on deeper level strategies indicated by participants' inferences nor on a mix of surface-level and deeper level strategies as indicated by participants' paraphrases, inferences, and monitoring in combination. Because there was also no effect of reading medium on comprehension performance among participants who were instructed to think aloud,[3] our pattern of results was not at all consistent with the notion that readers' text processing may mediate the effects of reading medium on comprehension performance.

Discussion

Whereas the shallowing hypothesis used to explain observed comprehension advantages for reading in print versus digitally maintains that more shallow processing when reading digitally is responsible for this difference (Annisette & Lafreniere, 2017; Delgado et al., 2018), this hypothesis has hardly been tested directly in prior research. In the present study, we therefore set out to investigate whether this hypothesis would stand up to scrutiny when verbal protocol analysis was used to identify strategic text processing at different levels of depth that could be compared across the reading mediums. Although we expected that this approach would produce results consistent with the indirect evidence available for this hypothesis (e.g., Ackerman & Goldsmith, 2011; Dahan Golan et al., 2018; Delgado & Salmerón, 2021; Halamish & Elbaz, 2020; Latini et al., 2020; Singer & Alexander, 2017), this was not the case. Thus, not only was comprehension performance quite similar across the reading mediums, but there were also no indications in our data that participants reading digitally processed the text more shallowly than did participants reading in print. As such, there was no evidence for a mediated effect of reading medium on comprehension performance via strategic text processing, which is at the heart of the shallowing hypothesis.[4]

Doubtless, most researchers find it more convenient to discuss findings in accordance with their hypotheses than findings posing challenges to those hypotheses. At the same time, however, such findings are essential to communicate because of the avenues they may suggest for future research within an area. Although we certainly do not consider this single study a reason to abandon the shallowing hypothesis altogether, it has given us some ideas about the possible boundaries of this hypothesis, that is, about the circumstances under which it may not hold water. As we discuss these circumstances in the following, we also suggest how future researchers may seek further clarification regarding these important issues.

First, like in the previous studies by Latini et al. (2019, 2020), who also did not find any direct medium effects on comprehension, performance was measured by a longer writing task in the present study. Such a task can be considered ecologically valid because students more often than not are required to demonstrate their comprehension by means of written products.[5] In addition, the validity of this measure is supported by the correlations between participants' scores and relevant individual difference (i.e., working memory, prior knowledge, basic reading comprehension skills) and text-processing measures (i.e., irrelevant, mixed processing) in this study. Of note is also that the lack of any reading medium effect on comprehension was paralleled by a lack of any effect on strategic text processing. However, given that this way of measuring comprehension performance is in contrast to most other studies of reading medium effects on comprehension, which have used multiple-choice and short question-answering tasks (Clinton, 2019; Delgado et al., 2018), it may suggest that the boundaries of the shallowing hypothesis do not extend to tasks in which readers are asked to produce longer, coherent texts to demonstrate their comprehension. Possible reasons for this may be that such tasks increase self-regulation of processing during reading in both medium conditions, or alternatively or additionally, as suggested by Latini et al. (2020), that such tasks neutralize any reading medium effects by increasing reflection during performance regardless of the medium. Of course, further clarification of this issue requires that future researchers in this area systematically vary the way of assessing comprehension performance in one and the same study and also collect processing data not only during reading but also during writing in both medium conditions. To eliminate the potential influence of writing competence on comprehension performance, it seems pertinent to include tasks that do not require any writing skills among the assessment tools.

A related point is that in using multiple-choice and short question-answering tasks, prior research has mainly tested a more basic form of reading comprehension than what we did in the present study, such as when comparing performance on the same standardized reading tests across mediums (e.g., Eyre, Berg, Mazengarb, & Lawes, 2017; Lenhard, Schroeders, & Lenhard, 2017). In contrast, we investigated potential processing and performance differences across mediums on a more complex task after such basic comprehension skills that others have used as dependent measures were controlled for. To control for the potential influence of writing on comprehension performance, future researchers could also measure engagement and competence in writing independently and then partial out their contributions to comprehension as measured with a writing task.

Second, the boundaries of the shallowing hypothesis may not extend to motivated adult readers, such as the participants in the present study. Of note is that in the meta-analysis by Delgado et al. (2018), educational level was not found to be a statistically significant moderator. Still, the possibility exists that adult readers in higher education may more easily adapt their processing and comprehension to different reading mediums in accordance with task requirements, given that they are motivated to do so. Relevant in this regard is the fact that our participants were recruited in noncompulsory lectures (which were not attended by all students enrolled in the program) and actively chose to participate in a research project on reading, and that they also reported a high level of intrinsic motivation for the reading task. Thus, it is conceivable that these participants were more motivated for the reading task than students typically participating in this kind of research, particularly as compared with students performing reading tasks and

completing tests across mediums as part of regular class activities. As such, our participants may have been more willing to invest effort in self-regulated processing and performance regardless of medium than what is typically the case. Comparing participants across not only age and educational level but also across different levels of task-based motivation therefore seems to be an important avenue for future research in this area, with use of more heterogeneous samples also needed to probe the generalizability of the current findings.

Third, the boundaries of the shallowing hypothesis may not extend to reading task contexts in which distractions are minimized, such as when data are collected individually rather than in groups or intact classes. Although the cited meta-analyses (Clinton, 2019; Delgado et al., 2018) did not register or study the potential effects of such variation, most of the cross-medium studies they included seem to have been conducted in groups of participants or classes. However, if performing similarly across mediums requires generally increased attention and self-regulation on the part of the readers, such adaptive resource allocation is likely to be facilitated by working individually while being monitored and guided through the tasks by a researcher present in the same room. Of note also is that two studies included in Delgado et al.'s (2018) meta-analysis that collected data individually from undergraduate students did not find any reading medium differences on comprehension performance (Chen, 2015; Margolin, Driscoll, Toland, & Kegler, 2013). Thus, this reading task context may elicit a maximum (rather than a typical) performance that tends to reduce or remove reading medium differences that might otherwise occur. Needless to say, studies that systematically compare maximum and typical performance within one and the same study, such as by varying the procedure for data collection, should be conducted to test the boundaries of the shallowing hypothesis.

Arguably, there are also aspects of the data collection procedure that may have limited the ecological validity of our study. These include that we presented the text as one 100-word paragraph per page, that participants could not go back to previously read paragraphs, and that the reading time was restricted. Although such limitations, given our design, were deemed necessary to obtain experimental control in comparing processing and comprehension across reading mediums, efforts to further increase ecological validity are recommended in future research on this issue.

In addition to the individual factors of task, individual motivation, and contextual distraction that we, based on our study, suggest may demarcate the boundaries of the shallowing hypothesis, future research should also study the potential interaction between these factors. For example, more complex tasks may not remove reading medium effects on processing and performance regardless of task motivation and engagement, and reduction of distractions in the reading and performance context may be less needed for readers at higher levels of motivation. In brief, because reading digitally does not seem to be universally detrimental for processing and comprehension, future research should attempt to determine for which individuals performing which tasks in which contexts this might not be the case. Importantly, such clarification would have not only theoretical but also educational implications because it would make it possible to target the constellation of tasks, individual differences, and contexts for which digital reading per se may pose particular challenges. For example, some readers may be particularly prone to skipping or skimming over information in apparently easy digital reading tasks, and some readers may be more dependent on a less distracting environment when working with digital texts than are others. By adapting interventions, such as in terms of reading strategy instruction, to such potential interactions, more students would presumably be able to reap the benefits of the digital revolution in reading technologies. In this regard, it is also important to note that deeper may not always be better with respect to text processing. Thus, in accordance with Dinsmore and Hattan (2020), surface-level strategies or a mix of surface- and deeper level strategies may actually be more adaptive for readers who lack competence or expertise in a domain. In the present study, this view was supported by the correlational pattern that we observed between the different categories of text processing and comprehension performance.

The lack of any reading medium differences with respect to text processing or comprehension performance in the present study seems to represent a challenge to the shallowing hypothesis. Apparently, skilled, motivated readers allowed to reflect on the content in writing, with a minimum of distraction present during both reading and writing, can process and, in turn, comprehend a relatively long digital and printed informational text similarly. This might be because such readers under such circumstances realize the need to strategically counteract the tendency to shallowly read digital texts in a myriad of other digital reading contexts (Goodwin et al., 2020; Latini et al., 2019), consistent with the notion that it is the mind-set readers bring to each reading medium rather than the reading medium itself that determines their text processing and resulting comprehension (Baron, 2020). If this mind-set can be changed as needed and lead readers to rise to the occasion when reading digital text for comprehension, there should be little reason to worry that the transition to a digital reading medium per se represents a threat to deep reading. Although our findings, therefore, might be considered good news, the caveat is, of course, that much digital reading, including academic reading, is not performed by highly motivated students reading with a minimum of distraction to perform extended writing tasks. Thus, although our findings may indicate important boundaries of the shallowing hypothesis, much reading of digital texts

both in and out of school can be assumed to take place within those boundaries.

In conclusion, the inconsistency in research on reading medium differences in processing and comprehension may be considered an impetus to further investigate the conditions under which such differences occur. As noted by Wolf (2018), the discourse on reading medium easily becomes polarized, with advocates for technological advancement standing against traditionalists applauding conventional print. Such entrenched beliefs about the better reading medium may be prevalent among teachers, parents, educational publishers, and policymakers, as well as among researchers having a stake in this field of inquiry. As we highlighted in this discussion, however, this is not a question of being generally right or wrong but of understanding the complex interplay of factors that likely determines the effects of reading in print versus digitally.

NOTES

We thank Ladislao Salmerón for help in designing the experiment.

[1] Participants rated their preference for reading medium when reading in the study context and in their leisure time, respectively, using a scale ranging from 1 (*clear preference for printed materials*) to 5 (*clear preference for digital materials*). Their mean scores were 1.71 (*SD* = 0.90) for the study context and 2.11 (*SD* = 1.19) for their leisure time.

[2] The first four of these categories fall in the broader category of specific phobias.

[3] An ANCOVA with reading medium as the independent variable, comprehension performance as the dependent variable, and reading comprehension and prior knowledge as covariates did not show a statistically significant effect of reading medium for the 58 participants who were instructed to think aloud, $F(1, 54) = 1.34$, $p = .252$, $\eta_p^2 = .024$. Only the effect of the covariate of prior knowledge was statistically significant in this analysis, $F(1, 54) = 9.25$, $p = .004$, $\eta_p^2 = .146$.

[4] Such a mediated effect of reading medium on comprehension performance is possible even when a direct effect of medium on comprehension is not significant (Hayes, 2009).

[5] In regard to ecological validity, participants in all conditions were also free to underline/highlight and annotate the text while reading, using a pencil when reading on paper and the PDF reader when reading on the tablet, and they could also take notes with a pencil on paper in all conditions. Participants were not explicitly encouraged to use any of these opportunities while reading, however, and none of them spontaneously did.

REFERENCES

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Afflerbach, P., Pearson, P.D., & Paris, S.G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, *61*(5), 364–373. https://doi.org/10.1598/RT.61.5.1

Alexander, P.A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M.L. Maehr & P.R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213–250). Greenwich, CT: JAI.

Alexander, P.A. (2005). The path to competence: A lifespan development perspective on reading. *Journal of Literacy Research*, *37*(4), 413–436. https://doi.org/10.1207/s15548430jlr3704_1

Alexander, P.A., & the Disciplined Reading and Learning Research Laboratory. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist*, *47*(4), 259–280. https://doi.org/10.1080/00461520.2012.722511

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

Baddeley, A.D. (2001). Is working memory still working? *American Psychologist*, *56*(11), 851–864. https://doi.org/10.1037/0003-066X.56.11.851

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

Baron, N.S. (2020). Digital reading: A research assessment. In E.B. Moje, P.P. Afflerbach, P. Enciso, & N.K. Lesaux (Eds.), *Handbook of reading research* (Vol. 5, pp. 116–136). New York, NY: Routledge.

Beker, K., Jolles, D., & van den Broek, P. (2017). Meaningful learning from texts: The construction of knowledge representations. In J.A. León & I. Escudero (Eds.), *Reading comprehension in educational settings* (pp. 29–62). Amsterdam, Netherlands: John Benjamins.

Björnsson, C.-H. (1968). *Läsbarhet [Readability]*. Stockholm, Sweden: Liber.

Bråten, I., Braasch, J.L.G., & Salmerón, L. (2020). Reading multiple and non-traditional texts. In E.B. Moje, P.P. Afflerbach, P. Enciso, & N.K. Lesaux (Eds.), *Handbook of reading research* (Vol. 5, pp. 79–98). New York, NY: Routledge.

Bråten, I., Brante, E.W., & Strømsø, H.I. (2019). Teaching sourcing in upper secondary school: A comprehensive sourcing intervention with follow-up data. *Reading Research Quarterly*, *54*(4), 481–505. https://doi.org/10.1002/rrq.253

Bråten, I., Johansen, R.-P., & Strømsø, H.I. (2017). Effects of different ways of introducing a reading task on intrinsic motivation and comprehension. *Journal of Research in Reading*, *40*(1), 17–36. https://doi.org/10.1111/1467-9817.12053

Bråten, I., Magliano, J.P., & Salmerón, L. (2020). Concurrent and task-specific self-reports. In D.L. Dinsmore, L.K. Fryer, & M.M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 275–295). New York, NY: Routledge.

Carr, N. (2011). *The shallows: What the internet is doing to our brains*. New York, NY: W.W. Norton.

Chen, D. (2015). *Metacognitive prompts and the paper vs. screen debate: How both factors influence reading behavior* (Master's thesis). Retrieved from https://smartech.gatech.edu/handle/1853/53840

Chi, M.T.H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

Cho, B.-Y., & Afflerbach, P. (2017). An evolving perspective of constructively responsive reading comprehension strategies in multilayered digital text environments. In S.E. Israel (Ed.), *Handbook of research on reading comprehension* (2nd ed., pp. 109–134). New York, NY: Guilford.

Cho, B.-Y., Woodward, L., & Afflerbach, P. (2020). Qualitative approaches to the verbal protocol analysis of strategic processing. In D.L. Dinsmore, L.K. Fryer, & M.M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 373–392). New York, NY: Routledge.

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

Coiro, J. (2021). Toward a multifaceted heuristic of digital reading to inform assessment, research, practice, and policy. *Reading Research Quarterly*, *56*(1), 9–31. https://doi.org/10.1002/rrq.302

Coté, N., Goldman, S.R., & Saul, E.U. (1998). Students making sense of informational texts: Relations between processing and representation. *Discourse Processes*, *25*(1), 1–53. https://doi.org/10.1080/01638539809545019

Dahan Golan, D., Barzillai, M., & Katzir, T. (2018). The effect of presentation mode on children's reading preferences, performance, and self-evaluations. *Computers & Education*, *126*, 346–358. https://doi.org/10.1016/j.compedu.2018.08.001

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

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

Dinsmore, D.L., & Hattan, C. (2020). Level of strategies and strategic processing. In D.L. Dinsmore, L.K. Fryer, & M.M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 29–46). New York, NY: Routledge.

Ericsson, K.A., & Fox, M.C. (2011). Thinking aloud is not a form of introspection but a qualitatively different methodology: Reply to Schooler. *Psychological Bulletin*, *137*(2), 351–354. https://doi.org/10.1037/a0022388

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Eyre, J., Berg, M., Mazengarb, J., & Lawes, E. (2017). *Mode equivalency in PAT: Reading comprehension*. Wellington: New Zealand Council for Educational Research.

Fox, M.C., Ericsson, K.A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Gellert, A.S., & Elbro, C. (2013). Cloze tests may be quick but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, *31*(1), 16–28. https://doi.org/10.1177/0734282912451971

Goodwin, A.P., Cho, S.-J., Reynolds, D., Brady, K., & Salas, J. (2020). Digital versus paper reading processes and links to comprehension for middle school students. *American Educational Research Journal*, *57*(4), 1837–1867. https://doi.org/10.3102/0002831219890300

Graesser, A.C. (2007). An introduction to strategic reading comprehension. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3–26). Mahwah, NJ: Erlbaum.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Halamish, V., & Elbaz, E. (2020). Children's reading comprehension and metacomprehension on screen versus on paper. *Computers & Education*, *145*, Article 103737. https://doi.org/10.1016/j.compedu.2019.103737

Hayes, A.F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Kammerer, Y., Brand-Gruwel, S., & Jarodzka, H. (2018). The future of learning by searching the web: Mobile, social, and multimodal. *Frontline Learning Research*, *6*(2), 81–91. https://doi.org/10.14786/flr.v6i2.343

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577. https://doi.org/10.3758/BF03193491

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

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

Kopatich, R.D., Magliano, J.P., Millis, K.K., Parker, C.P., & Ray, M. (2019). Understanding how language-specific and domain-general resources support comprehension. *Discourse Processes*, *56*(7), 530–552. https://doi.org/10.1080/0163853X.2018.1519358

Landow, G.P. (2006). *Hypertext 3.0: Critical theory and new media in an era of globalization*. Baltimore, MD: Johns Hopkins University Press.

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*, Article 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*, Article 101870. https://doi.org/10.1016/j.cedpsych.2020.101870

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

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

Leu, D.J., Kinzer, C.K., Coiro, J., Castek, J., & Henry, L.A. (2013). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. In D.E. Alvermann, N.J. Unrau, & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 1150–1181). Newark, DE: International Reading Association.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*(4), 778–784. https://doi.org/10.1037/0022-0663.94.4.778

Magliano, J.P., Higgs, K., Santuzzi, A., Tonks, S.M., O'Reilly, T., Sabatini, J., … Parker, C. (2020). Testing the inference mediation hypothesis in a post-secondary context. *Contemporary Educational Psychology*, *61*, Article 101867. https://doi.org/10.1016/j.cedpsych.2020.101867

Magliano, J.P., & Millis, K.K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251–283. https://doi.org/10.1207/S1532690XCI2103_02

Magliano, J.P., Millis, K., Ozuru, Y., & McNamara, D.S. (2007). A multidimensional framework to evaluate reading assessment tools. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). Mahwah, NJ: Erlbaum.

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*(4), 512–519. https://doi.org/10.1002/acp.2930

Mayer, R.E. (Ed.). (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

McNamara, D.S. (Ed.). (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. Mahwah, NJ: Erlbaum.

McNamara, D.S., & Magliano, J.P. (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

Milosevic, I., & R.E. McCabe (Eds.). (2015). *Phobias: The psychology of irrational fear*. Santa Barbara, CA: Greenwood.

Mizrachi, D. (2015). Undergraduates' academic reading format preferences and behaviors. *Journal of Academic Librarianship*, *41*(3), 301–311. https://doi.org/10.1016/j.acalib.2015.03.009

Parris, S.R., & K. Headley (Eds.). (2015). *Comprehension instruction: Research-based best practices* (3rd ed.). New York, NY: Guilford.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Pressley, M., & Harris, K.R. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P.A. Alexander & P.H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 265–286). Mahwah, NJ: Erlbaum.

Salmerón, L., Strømsø, H.I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, M. Planck, & P. van den Broek (Eds.), *Learning to read in a digital world* (pp. 99–120). Amsterdam, Netherlands: John Benjamins.

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, *19*(3), 285–307. https://doi.org/10.1007/s10648-007-9046-3

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, *47*(4), 427–463. https://doi.org/10.1002/RRQ.030

Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, *45*(4), 258–266. https://doi.org/10.1080/00461520.2010.515936

Singer, L.M., & Alexander, P.A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *Journal of Experimental Education*, *85*(1), 155–172. https://doi.org/10.1080/00220973.2016.1143794

Singer Trakhman, L.M., Alexander, P.A., & Berkowitz, L.E. (2019). Effects of processing time on comprehension and calibration in print and digital mediums. *Journal of Experimental Education*, *87*(1), 101–115. https://doi.org/10.1080/00220973.2017.1411877

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

Strømsø, H.I., Bråten, I., & Samuelstuen, M.S. (2003). Students' strategic use of multiple sources during expository text reading: A longitudinal think-aloud study. *Cognition and Instruction*, *21*(2), 113–147. https://doi.org/10.1207/S1532690XCI2102_01

Swanson, H.L., & Alloway, T.P. (2012). Working memory, learning, and academic achievement. In K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (Vol. 1, pp. 327–366). Washington, DC: American Psychological Association.

Swanson, H.L., & Trahan, M.F. (1992). Learning disabled readers' comprehension of computer mediated text: The influence of working memory, metacognition, and attribution. *Learning Disabilities Research & Practice*, *7*(2), 74–86.

Tabachnick, B.G., & Fidell, L.S. (2014). *Using multivariate statistics* (6th ed.). Harlow, UK: Pearson.

Trabasso, T., & Magliano, J.P. (1996). Conscious understanding during comprehension. *Discourse Processes*, *21*(3), 255–287. https://doi.org/10.1080/01638539609544959

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*(5977), 453–456. https://doi.org/10.1126/science.1182594

van den Broek, P., Fletcher, C.R., & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, *16*(1/2), 169–180. https://doi.org/10.1080/01638539309544835

van den Broek, P., Lorch, R.F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087. https://doi.org/10.3758/BF03206376

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading*, (pp. 71–98). Mahwah, NJ: Erlbaum.

Vinje, F.E. (1982). *Journalistspråket* [The journalist language]. Fredrikstad, Norway: Institute for Journalism.

Walton, E.W. (2014). Why undergraduate students choose to use e-books. *Journal of Librarianship and Information Science*, *46*(4), 263–270. https://doi.org/10.1177/0961000613488124

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

Submitted October 8, 2020
Final revision received March 25, 2021
Accepted April 7, 2021

NATALIA LATINI is a doctoral student in the Department of Education at the University of Oslo, Norway; email natalia.latini@iped.uio.no.

IVAR BRÅTEN (corresponding author) is a professor in the Department of Education at the University of Oslo, Norway; email ivar.braten@iped.uio.no.

APPENDIX A

English Version of the Text

Each of the 10 paragraphs was presented on a separate page in the experiment. For this article, we translated the text from Norwegian to English.

PHOBIAS

by
Psychologist Andrea Torgersen
Published in Journal of
Norwegian Psychological Association, no. 4, 2019

Fear is a natural reaction to real dangers, for example a fire, a dangerous animal, or an assault. The physiological changes that occur with fear sets the body in a state of alarm. A fear reaction can be characterized by anger and aggression and is then called a fight reaction. But a fear reaction can also be characterized by fearful flight and is then called a flight reaction. Many get scared if they see a poisonous snake or are threatened by other people. This type of fear is rational and does not prevent people from carrying out their everyday activities. Although all phobias are fears, not all fears are phobias.

(*continued*)

Phobias are intense and enduring fears of specific objects, activities, or situations. It is a form of fear that is exaggerated in relation to the real danger, that does not decrease with rational explanations or persuasion, and that is beyond voluntary control. Phobias interfere with adjustment, cause significant distress, and inhibit necessary actions toward goals. Many everyday activities may become nearly insurmountable for people who suffer from phobias because they will do everything to avoid what they fear. If they are not able to avoid it, fearful flight will characterize the reaction to phobic fear.

There are two main types of phobias: social phobias and specific phobias. A social phobia is a serious and persistent fear that arises when people anticipate a public situation in which they may be observed by others, such as when anticipating speaking to an audience or attending parties. People with a social phobia fear that they will behave in a way that is humiliating or embarrassing. They may also fear that other people will notice signs of anxiety. A specific phobia usually occurs as a response to certain animals and insects, injury and illness, natural phenomena (e.g., thunderstorms), or certain places (e.g., closed spaces).

Some less familiar phobias are phonophobia and trypophobia. Phonophobia involves fear of certain sounds. Often, this fear is based on the possibility that some sounds may damage hearing. Phonophobia will often arise in connection with kitchen activities, such as loading or emptying a dishwasher, and with environmental sounds such as loud speech and car traffic. Trypophobia involves fear of patterns of small holes, such as hole patterns that exist on the skin, seed capsules, corals, or the beeswax plates in a bee cube. Trypophobia is distinguished from many other phobias because the fear concerns objects that are completely harmless and that apparently have no connection to anything dangerous at all.

A possible explanation for why phobias arise is that they have genetic causes that increase the vulnerability of the autonomous nervous system, which is the part of the nervous system that humans cannot control directly and voluntarily. Because of this, humans overreact more easily and experience strong physiological activation, such as in the form of heart pounding and shivering. Genetic factors may also reduce the amount of an inhibitory chemical substance in the brain (GABA) that controls emotional responses. Some researchers think that abnormally low levels of this inhibitory substance may cause some people to have oversensitive nervous systems that easily produce anxiety reactions during stress.

Phobias may also be learned. When something occurs together with a very unpleasant or frightening situation, people may start to fear it, even though it was not considered dangerous at the outset. A person who has experienced being bitten by a dog or locked in a small room, for example, might later develop cynophobia or claustrophobia. Just thinking about a dog or a closed room may now generate strong fear. Such learning of phobias is a form of classical conditioning. One may also develop phobias indirectly by watching someone else panicking in certain situations or by being told that something is very dangerous. Such learning of phobias is a form of observational learning.

Another possibility is that humans are born with a tendency to fear things that were very dangerous in the distant past. This may be why certain phobias, such as fear of spiders or heights, are more common than other phobias. At one time in human history, certain fears enhanced our ancestors' chances of survival. This indicates that we carry around a tendency to respond quickly and unconsciously to objects or situations that have been a threat to humans in prehistoric times. Trypophobia may have such an evolutionary origin because the feared hole patterns remind people of poisonous organisms with similar visual characteristics.

The prevalence of phobias varies with type of phobia. Further, phobias are two to three times more common among women than among men. This may be associated with gender role patterns. Boys are more often encouraged to confront dangerous situations, which can reduce the risk of phobias because they become more accustomed to such situations. To express fear and avoid dangerous situations are more in accordance with girls' gender roles and something that girls are commonly encouraged to do, which is not the case for boys. Moreover, women are better at identifying emotional states in others by means of nonverbal signals such as facial expressions and body language.

Most people with a specific phobia do not seek treatment because they quite simply learn to avoid what they fear. Those who do seek treatment for phobias will in some cases get medicine to take right before or in feared situations. The most common medicines are anxiety-reducing substances (benzodiazepines) that alleviate muscle tensions and other physical symptoms of anxiety. Another drug that has shown promising results is cortisol, which is a hormone that is naturally secreted in the adrenal cortex. Although cortisol is a stress hormone, a right dosage may actually reduce phobias because it inhibits the effect of another stress hormone, namely, adrenaline.

People with phobias try to reduce their fear by staying clear of situations that scare them. Psychological treatment often consists of exposing them gradually to what they try to avoid in a safe environment. This is called exposure therapy, where the goal is to keep the person within the situation until he or she learns to endure the fear and experience that the feared consequences do not occur. Different types of exposure are used in the treatment, such as that persons are exposed to what they fear in real life, by imagining what they fear, or by interacting with what they fear by means of VR technology.

APPENDIX B

Coding System for Scoring the Written Responses

Definitions of scores 0, 1, and 2	
Score	**Definition**
0	Neither main idea nor elaboration of main idea correctly represented
1	Main idea or elaboration correctly represented
2	Main idea and elaboration correctly represented
Paragraph 1	
Score	**Example**
***Main idea**: Fear is a natural reaction to real dangers.*	
0	Fear and phobias are natural states.
1	As it was said in the article, it is fear that is a natural reaction to what happens in the surroundings and can be perceived as something dangerous.
***Elaboration**: A fear reaction can involve either fight or flight.*	
0	Fear is with us in situations and make us able to fight when we meet a dangerous situation.
1	There are two fear reactions, namely "fight or flight." With the first, the person reacts with anger and aggression, with the other, one escapes from what was dangerous.
Paragraph 2	
Score	**Example**
***Main idea**: Phobias involve irrational fear.*	
0	Having a phobia means that one is really afraid of something.
1	A phobia concerns fear that is not rational, something one shouldn't be afraid of.
***Elaboration**: Phobias interfere with adaptive functioning.*	
0	Phobias do not prevent people from functioning normally in their everyday lives as most of us quickly learn to avoid the things we fear.
1	Phobia is a form of strong fear that hinders persons in performing everyday tasks because they would rather avoid being exposed to their fear.
Paragraph 3	
Score	**Example**
***Main idea**: The main types of phobias are social phobias and specific phobias.*	
0	The two most common types are social phobia and trypophobia.
1	We usually distinguish between social phobias and specific phobias.
***Elaboration**: Description of both social and specific phobias and mentioning of at least one category of specific phobias*	
0	Mental phobia will be a fear one has developed in the form of anxiety before presenting in front of class, for example.
1	Social phobia, which is an irrational fear of social situations in which one is afraid that others will notice that one has anxiety. A specific phobia concerns more specific situations and is often directed toward a specific thing, such as heights, dogs, spiders etc.

(*continued*)

Coding System for Scoring the Written Responses (*continued*)

Paragraph 4	
Score	**Example**
***Main idea**: Some less familiar phobias are phonophobia and trypophobia.*	
0	We have several types, such as phonytropia and tropophobia.
1	Phonophobia and trypophobia are two less familiar phobias.
***Elaboration**: Description of both phonophobia and trypophobia*	
0	Tryponomia is a special type of phobia. This is a phobia of things with small holes, like skin, for example.
1	There are different types of phobias, e.g. phonophobia (afraid of certain sounds, e.g. the sound of a dishwasher), trypophobia (afraid of hole patterns, e.g. holes in the skin).
Paragraph 5	
Score	**Example**
***Main idea**: Phobias may have genetic causes.*	
0	The reason some get such strong reactions is a neurological disorder in the brain that makes one particularly afraid of things.
1	There may be different causes of phobias. First, there may be genetic causes....
***Elaboration**: The genetic explanation is related to the (autonomous) nervous system/a chemical substance in the brain/GABA.*	
0	There may be reactions in the nervous system that trigger too much hormones that lead to immense emotions and reactions.
1	Some researchers believe that phobias are caused by a vulnerable autonomous nervous system that makes the body overreact in situations where most would only be a little scared, or not scared at all.
Paragraph 6	
Score	**Example**
***Main idea**: Phobias may be learned.*	
0	They also talked about causes of phobias that concern the environment.
1	Another reason may be that phobias are acquired.
***Elaboration**: Learning of phobias may occur by means of classical conditioning and by means of observational learning (at least one form must be explained).*	
0	One theory is that one has experienced situations as traumatizing when quite young, and therefore brought this phobia into adulthood.
1	Some phobias seem to stem from previous experiences that have been perceived as particularly unpleasant, for example having been locked up (claustrophobia). Then, a classical conditioning occurs - that one associates a specific object or situation with very strong distress.
Paragraph 7	
Score	**Example**
***Main idea**: Phobias may be caused by an innate tendency to fear things that were dangerous in the distant past (i.e., have an evolutionary origin).*	
0	Phobias concern an inherent fear that people have of certain things and occurrences.
1	Another explanation may be evolutionary, namely that we get phobias of objects or situations that were life-threatening a long time ago (e.g., spiders), and that this fear therefore is innate as a defense mechanism.
***Elaboration**: Fear enhanced our chances of survival.*	
0	We are scared of heights because most people who are alive have not fallen from high heights.
1	Those human beings that managed to avoid the dangers of animals such as these (e.g., spiders), to a larger extent managed to survive and reproduce themselves in the past.

(*continued*)

Coding System for Scoring the Written Responses (*continued*)

Paragraph 8	
Score	**Example**
***Main idea**: Phobias are more common among women than among men.*	
0	Phobias may be different depending on one's gender.
1	There are more females than males who have phobias.
***Elaboration**: This may be associated with typical gender role patterns, with boys confronting dangers and girls avoiding dangers and/or being better at identifying emotional states in others.*	
0	It can also be because women display emotions more easily and one can more easily notice that women are scared of something.
1	This may be because boys are taught that one should be tough and face the fear, whereas girls have internalized that one should avoid fear and potentially dangerous situations.
Paragraph 9	
Score	**Example**
***Main idea**: Phobias may be treated with medicines.*	
0	All 0 points reflect lack of mentioning (no one misrepresented this idea).
1	An alternative is to get medicines before doing something one fears because of a phobia.
***Elaboration**: These medicines are anxiety reducing and/or cortisol (i.e., a stress hormone that inhibits another stress hormone, adrenaline).*	
0	Drugs may be used to treat this, usually antidepressants.
1	In the treatment of phobias one can use benzodiazepines, which are anxiety reducing. Or cortisol may have a positive effect by preventing the secretion of large amounts of adrenaline, which could lead to a fight or flight experience.
Paragraph 10	
Score	**Example**
***Main idea**: Phobias may be treated with exposure therapy.*	
0	Phobias can be cured by psychological help.
1	There are different methods for treating phobias. One is exposure therapy.
***Elaboration**: Exposure therapy involves gradually exposing persons to what they fear in a safe environment.*	
0	One must jump into the situation to face one's fear and not be scared of the incident.
1	One has individuals with phobias face what they fear in a safe environment, one who is afraid of spiders can hold a spider that is not poisonous for small periods to see that it is not dangerous.

Note. In translating the examples in this table, we tried to retain participants' original expressions rather than transforming them into (more) correct grammar.

APPENDIX C

Coding System for Scoring the Verbal Protocols

Processing	Definition	Examples
Associations	Comments made in response to text content that referred to prior knowledge or experience without facilitating understanding or creating coherence	It is not easy to be young. I'm not sure how difficult it will be to be young in the future, for my children and their children, because we have quite another, quite another view of what is private and not today than earlier. So we'll see. Soon privacy will be gone (paragraph 3). While thinking of it, I was thinking about my cousin, because she has phonophobia, so she has stopped visiting our house, so she just keeps to herself, it's so funny (paragraph 4).
Paraphrases	Comments that repeated or reworded text content	And then, there are two forms of phobias, social phobia and specific phobia (paragraph 3). That it is two to three times more common for women to have phobias than it is for men (paragraph 8).
Backward inferences	Comments that connected content in the current paragraph with content in one or more preceding paragraphs	I was thinking that this paragraph was about, about phobias, so it's building on the first paragraph, because in the first paragraph fear was explained (referring back to content in paragraph 1 after reading paragraph 2). …boys might be a bit more, like, fight mode, whereas girls are a bit more flight mode, it seems like there is such a relationship (referring back to content in paragraph 1 after reading paragraph 8).
Elaborative inferences	Comments that connected text content with relevant prior knowledge and experience that facilitated understanding and increased coherence	I am thinking that struggling with social phobias must be the worst. Yeah, because if you need treatment you kind of have to see a human being, so then you must break a barrier, which you have inside (paragraph 3). This gave me a kind of flashback to when I took psychology in high school and we learned about mice, for example, that was given a shock each time it tried to take food and then it learned that it couldn't go there, so then it started to stay away from it. I was, like, imagining that mouse and that psychology class (paragraph 6).
Predictive inferences	Comments that anticipated content in one or more forthcoming paragraphs based on content in the current paragraph	…but I reckon that what I'm going to read later is that phobias are not necessarily that rational (paragraph 1). …but what it didn't mention at all, was exposing oneself to the phobia in a way, which I find a little interesting. It is possible that it is described somewhere else in the text (paragraph 9).
Monitoring	Comments that involved reflections on one's own thinking about text content or one's own (lack of) understanding or knowledge of text content	I didn't realize that one could have a phobia of speaking in public places or of how others will perceive you, so that was a bit, I didn't know that in advance. I thought it was only, «only» in quotation marks, something that people regarded as anxiety, that those were not the same, but they actually are in a way (paragraph 3). I haven't really understood this text, but it has something to do with genes, I think. The hole pattern for example, um, yes, um. Do not quite understand what this text is about, the last one (paragraph 7).
Evaluation	Comments that were opinions about or affective responses to the text content	It was interesting what they said about some people being more vulnerable to phobias in relation to their nervous system (paragraph 5). Yes, it makes sense that people in a way can fear things that have been dangerous before, but I still think it was a bit strange (paragraph 7).
Other	Comments that did not fit into any of the categories above and did not facilitate understanding or contribute to coherence	And the last one was something about animals (paragraph 3). I don't have that many thoughts about that, the only thing I'm thinking is that it's quite good for those struggling with fear of the phobias (paragraph 9).

Note. In translating the examples in this table, we tried to retain participants' original expressions rather than transforming them into (more) correct grammar.