Received: 25 April 2023 | Revised: 20 October 2023 | Accepted: 22 October 2023

DOI: 10.1111/jcal.12903

ARTICLE

Journal of Computer Assisted Learning WILEY

Highlighting and highlighted information in text comprehension and learning from digital reading

Lucia Mason[1] | Angelica Ronconi[1] | Barbara Carretti[2] | Sara Nardin[2] | Christian Tarchi[3]

[1]Department of Developmental Psychology and Socialisation, University of Padua, Padua, Italy

[2]Department of General Psychology, University of Padua, Padua, Italy

[3]Department of Education, Languages, Intercultures, Literatures, and Psychology, University of Florence, Florence, Italy

Correspondence
Lucia Mason, Department of Developmental Psychology and Socialisation (DPSS), University of Padova, via Venezia 8 35131 Padova, Italy.
Email: lucia.mason@unipd.it

Abstract

Background: Digital texts are progressively becoming the medium of learning for students, but research has indicated that students tend to process information more superficially while reading on screen. It is therefore relevant to examine what strategies can support digital text comprehension.

Objectives: This study aimed to investigate the effects of highlighting—both learner generated and experimenter provided—when reading digitally.

Methods: University students ($N = 170$) were randomly assigned to the condition of learner-generated highlighting, experimenter-provided highlighting, or control. Reading outcomes were measured as literal and inferential text comprehension, transfer of knowledge, and metacognitive calibration of comprehension performance at immediate and delayed post-tests. Individual differences in prior knowledge, cognitive reflection, and reading self-efficacy were taken into account. The quality of the information highlighted by students in the condition of active highlighting was also measured.

Results: From linear mixed-effects models, the main effect of condition did not emerge for any of the outcomes. However, an interactive effect of condition and cognitive reflection emerged for literal text comprehension that favoured readers in the condition of experimenter-provided highlighting with higher ability to resist automatic thinking. Inferential text comprehension, transfer of knowledge, and calibration of performance were only predicted by cognitive reflection or reading self-efficacy. Finally, the quality of information highlighted significantly contributed to students' literal text comprehension and transfer of knowledge in the learner-generated highlighting condition.

Takeaways: Active highlighting is not effective per se during digital reading. The "amplification" effect of already highlighted text and higher cognitive reflection suggests that readers who are more able to resist automatic thinking may also invest more effort in the task, taking more advantage of the provided support. Even if active highlighting may not be effective per se compared to other reading conditions, what students highlight contributes to literal text comprehension and their learning from text.

KEYWORDS
digital reading, highlighting, inferential text comprehension, literal text comprehension, metacognitive calibration, transfer of knowledge

1 | INTRODUCTION

Reading strategically with an intentional and goal-directed approach is essential for comprehension and learning. Research has documented the positive relation between strategic processing and reading performance and, in particular, the quality of strategy use (Leopold & Leutner, 2015; Parkinson & Dinsmore, 2018). For today's students, especially in higher education, reading on the screen of a digital device for learning tasks is very common. The transition from reading on paper to reading on screen has led to the flourishing of investigations regarding the role of reading medium in text comprehension (Ben-Yehudah & Eshet-Alkalai, 2018; Florit et al., 2022; Ronconi et al., 2022). Three meta-analyses comparing the effects of traditional and digital reading, mostly in college students, are available (Clinton, 2019; Delgado et al., 2019; Kong et al., 2018). They share the issue that text comprehension is greater when reading traditionally, in particular for expository text and when reading occurs within a constrained timeframe. Moreover, looking at the process of reading in terms of reading time, no differences emerged in the two meta-analyses that addressed this outcome (Clinton, 2019; Kong et al., 2018). By contrast, in terms of metacognitive calibration such as the accuracy of judgements of one's own performance (Schraw, 2009), a meta-analysis indicated that self-assessment is more accurate when reading on paper than on a screen (Clinton, 2019).

A plausible explanation for the screen inferiority effect is the “shallowing hypothesis” (Annisette & Lafreniere, 2017), which assumes that students tend to process information superficially while reading on screen because they are used to quick and immediately rewarding interactions with digital devices. If this approach is also used when reading lengthy and complex texts for learning, interactions with the digital medium are inevitable superficial and passive, negatively affecting engagement with the text. As a consequence, the required deep processing cannot occur and comprehension of the text is damaged (Delgado & Salmerón, 2021). It is also worth noting that when reading print, readers' tendency is to read line by line or paragraph by paragraph, while when reading on screen they tend ‘to jump’ more from one place to another, which results in a shallowing processing that is not conducive to complex learning (Ben-Yehudah & Eshet-Alkalai, 2018; Zhang, 2013).

In light of this background, one question is particularly relevant: How can students' more active approach to digital text comprehension be supported? To provide an answer to this question and advance our knowledge of possible contextual factors that can promote a more productive approach to learning from text, the current study sought to investigate the effects of a simple and popular learning strategy for use during digital reading, that is, highlighting.

1.1 | Highlighting as a learning strategy

Learning strategies are techniques that students can use to improve learning and achievements across a variety of content domains. A learning strategy can be applied at various levels of quality, and it is always possible that a student applies it ineffectively. The quality of its implementation is related to learning performance (Leopold & Leutner, 2015; Parkinson & Dinsmore, 2018). However, learning strategies are considered to be more or less effective regarding the specific goals they can fulfil. According to generative learning theory, learning involves the active construction of meaning from informational material, which is cognitively processed, manipulated, and integrated with one's prior knowledge (Wittrock, 1974, 1989). Because we are interested in learning strategies for comprehending written texts, we should thus first consider that reading is a complex cognitive process that relies on both lower (e.g., decoding) and higher level of information processing (McNamara & Magliano, 2009). Multiple models of text comprehension have been proposed. A review of such models—such as, for example, the construction–integration model (Kintsch, 1998), the constructionist model (Graesser et al., 1994), or the landscape model (van den Broek et al., 1999)—is beyond the scope of our study. However, it is relevant to emphasize that, despite their differences, each of the models highlights the role of the reader, who can comprehend a text at different levels. While engaged in meaning making to form a coherent model of the information read, readers should mentally represent not only what is explicitly said in the text (literal comprehension), but also what is implicitly conveyed by making appropriate inferences (inferential comprehension) in connecting information across sentences and with prior knowledge (Sinatra & Broughton, 2011).

Readers can use strategies to support their text comprehension. Some strategies are considered to be more effective, as they contribute to generative learning and support deeper processing through the selection of relevant information, its organization in a coherent mental representation, and its integration with relevant prior knowledge. According to Fiorella and Mayer (2014, 2016), strategies such as summarizing, mapping, drawing, imagining, self-testing, self-explaining, teaching, and enacting support, although to varying degrees, active engagement in constructive meaning from the learning material.

Fiorella and Mayer (2014, 2016) did not include highlighting among such strategies, because it could better target rote, not deep, learning. However, highlighting was considered in previous reviews on learning techniques (Dunlosky et al., 2013), as students report using it frequently (Gurung et al., 2010). Second-hand books are usually marked by multicolour highlights. We focused our study on this strategy for five main reasons: (a) it is very commonly used while studying with the intention of improving learning; (b) technically, it is easy to implement while reading on the screen of a computer or tablet; (c) it has still been scarcely investigated for digital reading; (d) it can be a starting point to counteract the tendency toward a superficial approach to digital reading as foreseen in the "shallowing hypothesis" (Annisette & Lafreniere, 2017), which emerged from previous research; and (e) it has the potential to support text comprehension in university students based on the evidence from a recent meta-analytic study (see below, Ponce et al., 2022).

Essentially, highlighting (or underlining) means "marking potential important portions of to-be-learned materials while reading" (Dunlosky et al., 2013, p. 6). Based on generative learning theory, Fiorella and Mayer (2014, 2016) focused on three main processes to be activated to generate meaningful learning: selecting the relevant information to focus on, organizing it into a coherent mental representation, and integrating the new information with knowledge activated from long-term memory. Highlighting is assumed to activate the cognitive process of selecting relevant incoming information to process it in working memory, a fundamental step toward acquiring the new information and storing it in long-term memory (Fiorella & Mayer, 2014). Thus, highlighting is intended to sustain memory of textual information rather than assisting text comprehension, which requires deeper cognitive processing. However, it is possible that engagement in selecting information may also help in organizing reading material and connecting it with prior knowledge (Dunlosky et al., 2013).

It is also worth noting that the quality of highlighting contributes to the effectiveness of this strategy in supporting greater learning from text (Winchell et al., 2020). When a student highlights too little or highlights almost everything, the strategy is not productive. When highlighting everything, for example, students are making too little effort to process the text, as they remain uninvolved in selection of the relevant parts; they are therefore less likely to remember what is not distinctive (Dunlosky et al., 2013).

1.2 | Effects of highlighting on learning outcomes

Research on the effects of highlighting dates back to the 70s. Interestingly, active or learner-generated highlighting was compared to passive or instructor-provided highlighting. In the former condition, students are explicitly asked to highlight relevant information, while in the latter condition, the important parts are already highlighted in the text to be read. Instructor-provided highlighting can be conceived as a form of signalling or verbal cueing, because it drives learners' attention toward relevant material in the text (van Gog, 2021). Learner-generated highlighting and instructor-provided highlighting are two reading conditions that are usually compared to a control condition in which students only read the learning material. Results from a pioneer study by Fowler and Barker (1974) showed that the two highlighting conditions did not lead to better outcomes on a final test compared to the control condition. However, it also emerged that active highlighting was superior than passive highlighting. Interestingly, the quantity of text highlighted by the learner was negatively correlated with test performance.

Subsequent research has confirmed that students remember better marked text (Lorch, 1989). Learner-generated highlighting, however, is not always more effective than material already highlighted by an experimenter, likely because the latter is usually more competent in selecting relevant information (Nist & Hogrebe, 1987). For two decades, the literature has provided mixed findings about the learning benefits of highlighting or underlining while reading on paper (e.g., Johnson, 1988; Peterson, 1992). Compared to other learning strategies, however, highlighting remains one of the least valued (Fiorella & Mayer, 2015). Nevertheless, Miyatsu et al. (2018) more recently argued that highlighting might be productive in some circumstances, although past research has suggested that the strategy is not effective in educational contexts. In this regard, Yue et al. (2015) indicated that highlighting was more beneficial for students who did not perceive the strategy as productive or were unsure about its advantages compared to students who were in favour of highlighting.

It has also been demonstrated that if students highlight texts ineffectively by, for example, not selecting enough information or not selecting the critical information, they can also be quickly taught to implement the strategy more productively (e.g., Leutner et al., 2007). Highlighting was also used by List and Lin (2023), who combined it with annotations in a very recent study on learning from multiple digital texts. Highlighting and annotations were examined in relation to different task instruction conditions: to identify important information from the texts, to make connections across texts, to facilitate the evaluation of the texts, and to monitor one's own comprehension of information from a difficult text. The results revealed no significant differences across task instruction conditions for comprehension and integration of multiple texts. However, highlighting and annotating to monitor comprehension monitoring and evaluation, by noticing source information and statistical evidence, contributed to the comprehension and integration of multiple texts, unlike highlighting and annotating relevant information in the texts. The study suggested the potential of combining multiple strategies, such as highlighting and annotation, and task instructions for comprehending a set of digital texts on the same topic.

A very recent meta-analysis on the outcomes of highlighting documented that learner-generated highlighting improves memory but not comprehension, while instructor-provided highlighting improved both memory and comprehension. Moreover, learner-generated highlighting increased learning for college students, but not for school students, while instructor-provided highlighting led to greater learning for both college and school students (Ponce et al., 2022). The authors interpreted their findings as a confirmation that, generally, highlighting supports students in more superficial text processing, which is sufficient to remember the text. However, the findings also indicated that learner-generated highlighting is more effective for college students, who likely have a better sense of what to highlight, as they are more able to distinguish the more relevant information from the less relevant.

Of note, is that in all but one of the studies meta-analysed by Ponce et al. (2022) participants read on paper. In the only study that focused on active highlighting during digital reading, the strategy was found not to be as effective for learning from digital texts as it was from printed texts (Ben-Yehudah & Eshet-Alkalai, 2018). More recently, Goodwin et al. (2020) indicated that active digital highlighting and rereading supported text comprehension in fifth to eighth graders, while Winchell et al. (2020) found that college student-generated highlighting predicted their comprehension and interest. Thus, the literature indicates, on the one hand, that the benefit of active highlighting on paper may not automatically transfer to highlight on screen (Ben-Yehudah & Eshet-Alkalai, 2018). On the other hand, however, the potential of this strategy during digital reading for students' at different educational levels also emerged (Goodwin et al., 2020; Winchell et al., 2020). To make a significant contribution with our study, we moved from taking into account the mixed results for the fundamental academic learning activity of reading, which is increasingly performed on digital devices. We therefore considered it worthwhile to continue investigating the role of highlighting by comparing both of its forms, active and provided, in comprehension and learning from digital reading. Our investigation also took into account individual differences that could potentially moderate the role played by highlighting.

2 | THE CURRENT STUDY

This study aimed to gain new knowledge of the effects of both learner-generated and experimenter-provided highlighting in digital reading, specifically reading to study complex informational material on the screen of a computer. The overall purpose was to contribute to research on strategic and goal-oriented learning from digital texts. We also considered that cognitive and motivational individual differences might play a role in the relationship between highlighting and reading outcomes. Specifically, we focused on prior knowledge, cognitive reflection, and reading self-efficacy as possible moderators. Prior knowledge was considered, as it influences comprehension by favouring students who have more relevant prior knowledge to activate from long-term memory as decades of research have documented (e.g., Mason et al., 2020; McNamara & Kintsch, 1996; Ozuru et al., 2009). Cognitive reflection refers to the ability to think rationally rather than intuitively, avoiding easier and automatic responses and allowing them to be overridden by further reflection (Toplak et al., 2014). In other words, cognitive reflection implies the ability to use System 2—slow, deliberate, rationale, and effortful thinking—instead of System 1, or fast, automatic, intuitive, and effortless thinking (Kahneman, 2011). A recent meta-analysis has documented that cognitive reflection highly correlates with cognitive abilities, including verbal ability (Otero et al., 2022). As such, it likely supports the construction of high-quality mental representations of content of a complex text. Finally, self-efficacy for reading was considered, as this motivational variable is a well-known resource for reading processes and outcomes (e.g., Bråten et al., 2013; Chen et al., 2021). Controlling for these variables ensured that any effects of our manipulation of the independent variable, highlighting, was independent of them. Thus, to the best of our knowledge, this study is the first to seek to contribute to current knowledge by taking into account simultaneously: (a) both forms of highlighting, learner generated and experimenter provided, during digital reading; (b) not only text comprehension at different levels but also metacomprehension, that is, the metacognitive calibration of one's own comprehension performance; and (c) cognitive and motivational factors that may play a role in text comprehension and metacognitive calibration. To pursue these purposes, this study was guided by the following research questions:

> **RQ1.** In digital reading, do differences for literal and inferential text comprehension, transfer of knowledge, and metacognitive calibration of comprehension performance emerge when comparing learner-generated highlighting, experimenter-provided highlighting, and no highlighting while controlling for prior knowledge, cognitive reflection, and reading self-efficacy?

> **RQ2.** Does the overall quality of learner-generated highlighting predict text comprehension?

For RQ1, we hypothesized that active highlighting by college students, who are supposed to have the ability to select relevant information, would be at least as effective as provided highlighted text in sustaining text processing (Ben-Yehudah & Eshet-Alkalai, 2018) and thus immediate and delayed comprehension, in particular literal comprehension, if not inferential comprehension. Both highlighting conditions would be superior to the control condition for the lower level of text comprehension (H1) (Fiorella & Mayer, 2014, 2016; Ponce et al., 2022). For transfer of knowledge, which requires the application of newly learned knowledge going beyond the text, we hypothesized no differences across the three reading conditions (H2). Given the lack of prior research, we took an explorative approach to the effects of active and provided highlighting on metacognitive calibration. We did not expect a moderating effect for the possible contribution of prior knowledge, because the complex text topic would be unfamiliar to participants, so there would be very little differences among participants for this variable. In contrast, we expected that both cognitive reflection and reading self-efficacy could contribute to text comprehension and moderate the effects of condition in favour of those with higher perception of competence and ability to resist automatic responses and think rationally. More specifically, we expected that readers with higher cognitive reflection, relative to those with lower ability to use System 2, would benefit more from active highlighting, which required them to process information less superficially, showing higher reading outcomes. At the same time, we also expected that readers with higher cognitive reflection would benefit more than readers low in this ability, as their tendency to use System 2 and put effort in processing information would lead them to take real advantage from passive highlighting by processing text content more deeply. Similarly, we considered that the advantage for reading outcomes of high ability in cognitive reflection could also emerge for the motivational variable, that is, readers with higher reading self-efficacy would benefit either from active or passive highlighting. In the former case, the motivational variable supports deeper processing and persistence in face of difficulties. In the case of experimenter-provided highlighted, higher self-efficacy would sustain effortful content processing, as when a task is facilitated to some extent.

For RQ2, we hypothesized that the quality of the information selected by university students would predict their reading outcomes for comprehension and learning (H3), as they depend, at least to some extent, on the selection of relevant information, which can also help readers to organize the text content coherently (Leopold & Leutner, 2015; Miyatsu et al., 2018). Our hypotheses were not pre–registered, unfortunately.

3 | METHOD

3.1 | Participants

Initially 233 university students completed the online study on the Qualtrics platform. The data for 23 students from the control condition (see next section) were excluded from the analysis, as they spontaneously highlighted or took notes while reading. The data for 17 students were excluded because their reading time was too short or too long for their performance to be considered reliable and acceptable, or they did not complete all the tasks. Therefore, the final sample consisted of 192 students ($F = 154$, $M = 35$, and non-binary $= 3$; $M_{age} = 25.40$, $SD = 8.72$). The sample was smaller at the delayed post-test, which was completed by 170 students, but still with sufficient power. The sample size was appropriate according to a priori power analysis performed with G*power (Faul et al., 2007), based on $\alpha = 0.05$, $1 - \beta = 0.80$, and an estimated medium effect size ($f = 0.25$). The statistical test selected for power analysis using G*Power was ANOVA: the repeated–measures ANOVA model with the inclusion of a between-within interaction mostly overlaps with linear mixed-effects models (LMMs) with an interaction term. The analysis with G*Power thus made it possible to quantify the sample size required to detect an interaction effect of a between-subject factor (reading condition) and a within-subject factor (post-test time).

Participants were recruited during regular lectures and agreed to take part in the study in exchange for partial course credit. The study was approved by the pertinent ethics committee and all participants signed an informed consent form. Almost all participants (98%) were native speakers of the country language. Participants were enrolled in a three-year bachelor's program (53.12%), in a five-year program (23.44%), or in a two-year master's program (23.41%). Most of participants studied psychology (61.7%), while the remaining participants were mainly prospective teachers of primary school education (31.8%). All of the participants had passed an entrance examination that included a reading comprehension test. Of note is that the participants generated a code at the beginning of the first session and stored it, as they were informed that they would be asked to insert the code at the beginning of the second session. On the basis of the generated code, we were able to match the data from the first and second sessions (immediate post-test and delayed post-test) for each participant.

3.2 | Experimental reading conditions

We used a between-subjects design with an immediate post-test and delayed post-test design. Participants were randomly assigned to one of three conditions: Learner-generated highlighting, experimenter-provided highlighting, or a control condition with no highlighting. In the learner-generated highlighting condition, students were given the following instructions: "Read the text as you would when you read to study and be able to answer a series of questions we will ask you after reading. While reading, please highlight the important parts of the text to help you understand it better." The Qualtrics platform allowed them to use only one colour (yellow) to highlight the text using the mouse.

In the experimenter-provided highlighting condition, the following instructions were given: "Read the text as you would when you read to study and be able to answer a series of questions we will ask you after reading. In the text, the important parts are already highlighted to help you understand it better." In the control condition, the participants were instructed to: "Read the text as you would when you read to study and be able to answer a series of questions we will ask you after reading." As a first manipulation check, at the end of the experiment we asked the participants in the control condition whether they used any strategies while reading. As introduced in the previous session, a number of them reported that they had highlighted (which we could verify) or taken notes on a sheet of paper. We therefore did not consider the data from these students.

In sum, there were 76 participants in the learner-generated highlighting condition (68 completed both immediate and delayed post-tests), 66 in the condition of experimenter-provided highlighting (59 at both post-tests), and 50 in the control condition (43 at both post-tests). The unequal subsamples were due to the exclusion of different numbers of participants from the three conditions for the reasons already mentioned.

As a second manipulation check, we computed participants' time spent reading the text in each condition to ensure that in the learner-generated highlighting condition, participants spent longer with the text than in the other two conditions.

3.3 | Learning text

All participants read a 1028-word long text about the complex topic of stem cells. The text was taken and adapted from a previous study

(Tarchi et al., 2021). As an indication of text difficulty, we used the Gulpease index for Italian texts (Lucisano & Piemontese, 1988). The readability score (maximum = 100) was 47, which indicates that the text was challenging, as some effort was needed to comprehend its content well, even for students with a high school diploma. The text introduced information regarding the origins of scientific research on stem cells, their types, names, structure, functions, and development; current and future advances; and the regulated use of stem cells in some European countries in light of the debate on embryonic stem cells.

Following Leopold and Leutner (2015), in the experimenter-provided highlighting condition, the text was highlighted according to three levels of information relevance identified by a colleague who is an expert in reading and text comprehension: (a) the first or highest level included the fundamental ideas of the theme: there were eight across the paragraph; (b) the most important concepts were identified at the second level: there were 17; and (c) less important but still relevant pieces of information were identified at the third level: there were nine throughout the text. Literal and inferential questions in the post-tests addressed aspects at all the three levels. A picture of the highlighted text is shown in Figure S1. All of the questions but one in the immediate and delayed post-tests (see below) required information that was highlighted in the text used in the experimenter-provided highlighting condition.

3.4 | Measures

3.4.1 | Dependent variables

Text comprehension

The same knowledge test measured literal (17 questions) and inferential (eight questions) text comprehension in both immediate and delayed post-tests using multiple-choice questions that presented four options, of which only one was correct. Information to answer the literal questions were provided in the text, while the answers for the inferential had to be constructed by connecting information. Students received one point for every correctly answered question (maximum score = 17 for literal questions and 8 for inferential questions). An example of the questions to measure literal comprehension is: "Embryonic stem cells are called pluripotent cells because (a) they replicate in all types of human cells; (b) they replicate in 250 types of human cells*; (c) they are in all types of human tissues; and (d) they differentiate only in the cell types of the tissues they belong to." An example of the questions to measure inferential text comprehension is: "Adult stem cells … (a) must be collected in a sufficient amount to be implanted in the tissue to be regenerated; (b) must be collected in a sufficient amount in order for them to be able to create a new organ; (c) must be cultivated in order for them to grow and to be implanted in the tissue to be regenerated*; (d) must be implanted in the tissue to be regenerated and them they replicate in that tissue." As measured by McDonald's ω, the reliability for the literal questions at the immediate and delayed post-tests was 0.68 and 0.58, respectively; for inferential question, it was 0.60 and 0.54, respectively. Reliability indices were below the benchmark for reliability typically required for standardized tests. However, considering that the acceptability of a test's reliability also depends on how the measure is used and what type of decision is based on it (e.g., Kerlinger & Lee, 2000; Nunnally, 1978), our intention was to compare group, not individual, means across the three conditions using a test devised for the purpose of research.

Transfer of knowledge

Readers' ability to apply the new learned knowledge to different situations at both the immediate and delayed post-tests was measured using four open–ended questions. Responses were assigned 0–3 points according to their level of correctness and completeness (maximum score = 12). An example of these questions is: "In the past, the umbilical cord was thrown away as waste material, but today it is a well-known source of useful blood stem cells. What might be some of the possible uses of blood stem cells?" Responses were coded by two independent raters (second and fourth authors) with a mean inter-rater agreement of 0.97 and 0.96, as measured by intraclass correlation coefficient (ICC), for immediate and delayed text comprehension, respectively.

Quality of learner-generated highlighting

We assessed the quality of the learner-generated highlighting according to the aforementioned three levels of information relevance identified in the text and used for the already highlighted text given to readers in the experimenter-provided highlighting condition. Specifically, for each level of information relevance, we first computed the proportion of information units actively highlighted. These proportions were computed on the total number of information units already highlighted at each level—according to the expert model—in the text used in the experimenter-provided highlighting condition. We then considered the overall mean proportion in the analysis (see Table 1).

Metacognitive calibration of performance

The accuracy of participants' judgements of their comprehension performance was measured by asking participants "How well did you answer the questions on the text you read?" Participants expressed their judgement using a slider that moved from 0 to 100. Calibration bias was computed as the difference between the self-evaluation of comprehension performance and actual performance.

3.4.2 | Control variables

Prior knowledge

This was measured by 14 multiple-choice questions, scored 0 if incorrect and 1 if correct. The reliability for this task was very low (McDonald's $\omega = 0.26$), but this may be explained by the fact that before exposure to learning material about unfamiliar and complex content, which had never been introduced to the participants, readers may have only very small and fragmented pieces of information and may also guess their answers. A high reliability value cannot be

TABLE 1 Descriptive statistics for dependent variables by condition ($N = 192$).

Experimenter-provided highlighted ($n = 66$)	*M* (SD)	Skewness	Kurtosis
Reading time[a]	487.69 (282.84)	1.75	3.70
Reading time (log-transformed)	6.05 (0.54)	−0.34	2.47
Immediate literal comprehension	12.52 (2.83)	−0.77	−0.14
Immediate inferential comprehension	6.12. (1.31)	−1.48	5.25
Immediate transfer of knowledge	6.20 (1.68)	−0.76	1.65
Calibration of performance	−21.36 (18.74)	0.03	−0.15
Delayed literal comprehension ($n = 59$)	11.03 (2.30)	0.24	−0.89
Delayed inferential comprehension ($n = 59$)	5.61 (1.35)	−0.15	−0.59
Delayed transfer of knowledge ($n = 59$)	5.34 (1.92)	−0.77	0.76
Learner-generated highlighting ($n = 76$)			
Reading time[a]	763.78 (385.71)	1.69	4.52
Reading time (log-transformed)	6.53 (0.48)	−0.27	0.97
Proportion of highlighted information units [Level 1]	0.59 (0.23)	−0.23	−0.23
Proportion of highlighted information units [Level 2]	0.65 (0.19)	−0.76	0.55
Proportion of highlighted information units [Level 3]	0.62 (0.25)	−0.24	−0.71
Proportion of highlighted information units [Average]	0.62 (0.17)	−0.25	−0.64
Immediate literal comprehension	12.61 (2.69)	−0.40	−0.27
Immediate inferential comprehension	6.18 (1.35)	−0.81	0.75
Immediate transfer of knowledge	5.95 (1.88)	−0.56	0.33
Calibration of performance	−23.23 (18.96)	−0.43	−0.73
Delayed literal comprehension ($n = 68$)	10.16 (2.48)	−0.61	0.68
Delayed inferential comprehension ($n = 68$)	5.51 (1.34)	−0.21	−0.39
Delayed transfer of knowledge ($n = 68$)	5.88 (1.92)	−0.58	0.76
Control ($n = 50$)			
Reading time[a]	549.97 (360.59)	1.84	3.60
Reading time (log-transformed)	6.15 (0.57)	0.28	−0.15
Immediate literal comprehension	12.50 (2.64)	−0.58	−0.27
Immediate inferential comprehension	5.82 (1.35)	−0.26	−0.25
Immediate transfer	6.36 (1.90)	−0.63	1.22
Calibration of performance	−17.41 (21.79)	−0.21	−0.60
Delayed literal comprehension ($n = 43$)	10.42 (2.91)	−0.33	−0.64
Delayed inferential comprehension ($n = 43$)	5.37 (1.46)	−0.06	−0.46
Delayed transfer of knowledge ($n = 43$)	6.37 (1.83)	−0.34	−0.48

[a]In seconds.

expected for a knowledge test devised for research and for assessing unfamiliar and complex content in a non-redundant way with different items (Taber, 2018).

Cognitive reflection

This was measured using the 6-item Italian Cognitive Reflection Test previously validated with Italian university students (Primi et al., 2015). The original test consisted of three items (Frederick, 2005). An example item is: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" (Intuitive answer = 10 cents; correct answer = 5 cents). Primi et al. (2015) added three new items to the original test by Frederick (2005), for example: "Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?" (intuitive answer = 30; correct answer = 29). This test involves cognitive reflection to override the automatic but incorrect responses and produce the correct one based on rational rather than intuitive thinking (Toplak et al., 2014). McDonald's ω for this test was 0.74.

Reading self-efficacy

This was measured using a 9-item scale rated on a 10-point Likert-type scale (1 = not at all; 10 = very much), adapted by Tarchi et al. (2021) from Anmarkrud and Bråten (2009). An example item is: "I will not have problems understanding even the most difficult part of the texts I should read this year." McDonald's ω for this measure was 0.88.

3.5 | Procedure

Participants first completed a short demographic questionnaire, the prior knowledge task, and the reading self-efficacy scale. They then carried out the cognitive reflection task. Afterwards, they read the text about stem cells with instructions appropriate to their assigned reading condition. After reading, they answered the multiple-choice questions assessing literal and inferential comprehension, the open-ended questions assessing the transfer of knowledge, and the question on metacognitive calibration to measure the accuracy of their judgement of their comprehension performance. Students in the control condition also responded to a question asking whether they had used any strategy during reading the text on cell stems and, if so, to indicate the strategy. The delayed post-test was given 2 weeks after the immediate post-test.

3.6 | Analytical plan

To answer RQ1 we used LMMs with a random intercept to analyse the effect of both condition and time on immediate and delayed literal and inferential comprehension, as well as transfer of knowledge, taking into account only the data of participants who carried out both post-tests ($N = 170$). LMMs allow consideration of both fixed and random effects. In all fitted models, we included the time (immediate and delayed post-tests) and condition (learner-generated highlighting, experimenter-provided highlighting, and control) as fixed effects and the variable of student as the random effect. As covariates, we considered prior knowledge, cognitive reflection, and reading self-efficacy; all were mean centered. Based on correlations, to answer RQ1 we fitted an LMM that included a 3-way interaction between condition, time, and cognitive reflection to test whether the effect of condition over time was moderated by the tendency to override automatic but incorrect responses based on rational rather than intuitive thinking. It is noteworthy that, for the sake of parsimony, we did not consider the interaction terms for prior knowledge and reading self–efficacy as fixed effects, because they did not improve the fit of the models. To examine the effect of condition on performance calibration, we used a multiple linear regression model, as this dependent variable was only measured at immediate post-test, taking into account the data of the participants considered for this post-test ($N = 192$).

To answer RQ2 about the contribution of the quality of participants' highlighting to text comprehension, transfer of knowledge, and metacognitive calibration, we used similar LMMs for the learner-generated highlighting condition ($N = 68$ in both post-tests), but we also included the quality of highlighted information and the interaction between quality and time.

To ensure unbiased variance estimations, we calculated parameters using the restricted maximum likelihood procedure (Zuur et al., 2009). Both 95% confidence interval and statistical significance are reported for each parameter. The goodness of fit of the mixed-effects model is indicated by the marginal R, which represents the variance explained by the fixed factors, the conditional R^2, which represents the variance explained by both the fixed and random factors (Nakagawa & Schielzeth, 2013), and the intra-class correlation coefficient (ICC). We also report the variance explained by the random effect of the students ($\tau_{00\text{-ID}}$) and the variance of the residuals (σ^2). Based on their nature, the time and condition variables were categorical and dummy coded. Effect coding allows comparison of each category to the overall mean and estimation of the deviation of each category from the mean. The term overall mean here refers to the unweighted grand mean, as the sample size across three conditions was not equal. Their reference levels were immediate post-test for time, and control for condition; p-values were two-sided, and the statistical significance level was defined as $p < 0.05$. Statistical analyses were performed using the R software (R Core Team, 2021), version 4.2.1; we used the "contr.sum" function ("stats" package, version 4.2.1) to effect-code our categorical variables and then fitted LMMs using the "lmer" function ("lme4" package, version 1.1-30; Bates et al., 2015). We tested for the overall significance of condition and its interaction with cognitive reflection using "Anova" (type3, "car" package, version 3.1.1; Fox & Weisberg, 2019), which, by default, applies F-tests for linear regression models and Wald chi-square tests for the LMMs. Finally, we performed post-hoc pairwise comparisons using "emmeans" and "emtrends" functions ("emmeans" package, version 1.8.1-1, Lenth, 2020) with Bonferroni's adjusted p values. Data visualization was performed using the "ggplot2" package (version 3.3.6; Wickham, 2016).

4 | RESULTS

We report the results for the two research questions, starting from the preliminary analyses.

4.1 | Preliminary analyses

We considered reading time as a second manipulation check to ensure that the readers in the learner-generated highlighting condition took more time to read the text than readers in the other two conditions (see Table S1). To correct for skewness of the data distribution, reading time was log-transformed. A linear regression model with the covariates of prior knowledge, cognitive reflections, and reading self-efficacy revealed that condition had a significant main effect on reading time, $F(2,164) = 12.38$, $p < 0.001$. Participants who actively highlighted the text spent 36% more time on the task than readers in the experimenter-provided highlighting condition (ratio = 0.64, 95% CI [0.511–0.811], $p < 0.001$), and 32% more time than readers in the control condition (ratio = 0.68, 95% CI [0.532–0.880], $p = 0.001$).

TABLE 2 Correlations between control and dependent variables.

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
1. Prior topic knowledge	7.62	1.76	–									
2. Reading self-efficacy	54.96	8.36	0.00	–								
3. Cognitive reflection	2.98	1.50	−0.01	0.13	–							
4. Immediate literal comprehension	12.55	2.71	0.08	0.20**	0.16*	–						
5. Immediate inferential comprehension	6.07	1.34	−0.00	0.11	0.31**	0.42**	–					
6. Immediate transfer of knowledge	6.14	1.82	0.03	0.15*	0.12	0.29**	0.17*	–				
7. Performance calibration	−21.07	19.69	−0.06	0.21**	0.00	−0.20**	−0.15*	−0.17*	–			
8. Delayed literal comprehension	10.53	2.55	0.04	0.18*	0.22**	0.61**	0.37**	0.29**	−0.14	–		
9. Delayed inferential comprehension	5.51	1.37	0.03	0.03	0.28**	0.27**	0.50**	0.14	−0.08	0.30**	–	
10. Delayed transfer of knowledge	5.82	1.93	0.02	0.13	0.04	0.26**	0.29**	0.52**	0.01	0.28**	0.16*	–

*$p < 0.05$; **$p < 0.01$.

TABLE 3 Direct and interactive effects on literal (left) and inferential (right) text comprehension ($N = 170$).

	Literal comprehension			Inferential comprehension		
Predictors	**Estimates**	**95% CI**	***p***	**Estimates**	**95% CI**	***p***
(Intercept)	11.54	11.19–11.89	**<0.001**	5.75	5.58–5.92	**<0.001**
Condition [Highlighted]	0.31	−0.18 to 0.79	0.214	0.07	−0.17 to 0.32	0.549
Condition [Highlighting]	−0.23	−0.70 to 0.24	0.339	0.08	−0.16 to 0.31	0.515
Post-test [T2]	−0.96	−1.14 to −0.78	**<0.001**	−0.23	−0.33 to −0.12	**<0.001**
Prior knowledge	0.10	−0.09 to 0.30	0.306	0.01	−0.08 to 0.11	0.770
Reading self-efficacy	0.05	0.01–0.09	**0.020**	0.01	−0.01 to 0.03	0.390
Cognitive reflection	0.25	0.02–0.48	**0.033**	0.25	0.14–0.36	**<0.001**
Condition [Highlighted] * post-test [T2]	0.11	−0.14 to 0.36	0.393	−0.00	−0.15 to 0.14	0.972
Condition [Highlighting] * post-test	−0.16	−0.40 to 0.08	0.193	−0.07	−0.21 to 0.07	0.320
Condition [Highlighted] * cognitive reflection	0.33	0.02–0.64	**0.038**	0.00	−0.15 to 0.16	0.960
Condition [Highlighting] * cognitive reflection	−0.43	−0.74 to −0.11	**0.009**	−0.05	−0.21 to 0.11	0.555
Post-test [T2] * cognitive reflection	0.07	−0.05 to 0.18	0.268	−0.00	−0.07 to 0.07	0.968
Condition [Highlighted] * post-test) * cognitive reflection	−0.08	−0.24 to 0.08	0.333	−0.00	−0.10 to 0.09	0.931
Condition [Highlighting] * post-test * cognitive reflection	0.05	−0.11 to 0.22	0.540	−0.06	−0.15 to 0.04	0.248
Random effects						
σ^2	2.74			0.94		
$\tau_{00\text{-ID}}$	3.73			0.80		
ICC	0.58			0.46		
N_{ID}	170			170		
Observations	340			340		
Marginal R^2/conditional R^2	0.210/0.665			0.120/0.523		

Note: Highlighted, experimenter-provided highlighting, highlighting, learner-generated highlighting. Bold values indicate statistical significance.

Descriptive statistics for the dependent variables are reported in Table 1 by reading conditions. Overall, participants' performance was relatively good for literal questions, good for inferential questions, and modest for transfer questions.

Means, standard deviations, and correlations between control and dependent variables are reported in Table 2 for the entire sample.

RQ1. Effects of highlighting on dependent variables.

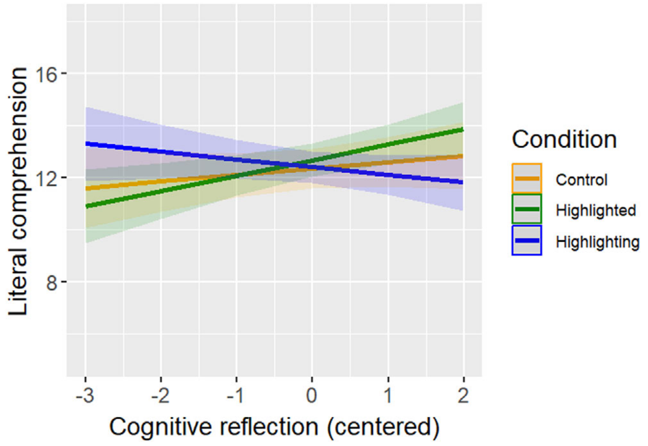

FIGURE 1 Interaction by condition and cognitive reflection in literal text comprehension.

4.1.1 | Literal text comprehension

For literal text comprehension measured at immediate and delayed post-tests, an LMM revealed the main effect of time ($\beta = -0.96$, 95% CI [−1.14 to −0.78], $p < 0.001$), that is, students had lower scores on the delayed post-test than at the immediate post-test (see Table 3, left panel). The interactive effect of condition × cognitive reflection also emerged, χ^2 (2) = 7.99, $p = 0.018$.

When considering the interaction between condition and cognitive reflection, post-hoc comparisons revealed the effect of cognitive reflection on literal comprehension significantly differed between the two conditions involving highlighting ($\beta = 0.75$, $SE = 0.27$, $p = 0.018$). The interactive effect did not emerge when comparing active ($\beta = -0.52$, $SE = 0.29$, $p = 0.228$) and passive ($\beta = 0.24$, $SE = 0.28$, $p = 1.00$) highlighting conditions with the control condition. Specifically, participants in the experimenter-provided highlighting condition with higher ability to use slow and rational rather than fast and intuitive thinking, showed better comprehension of literal information than those with lower ability in the same condition ($\beta = 0.58$, $SE = 0.19$, 95% CI [0.21–0.95]). By contrast, for students who actively highlighted the text ($\beta = -0.18$, $SE = 0.20$, 95% CI [−0.57–0.21]) and for those in the control condition ($\beta = 0.34$, $SE = 0.21$, 95% CI [−0.08–0.76]), no statistically significant differences emerged related to higher or lower cognitive reflection (see Figure 1). The model explained 66% of the variance (marginal R^2 explained by fixed effects = 21%).

4.2 | Inferential text comprehension

For inferential text comprehension at the immediate and delayed post-tests, an LMM only revealed the main effects of time ($\beta = -0.23$, 95% CI [−0.33 to −0.12], $p < 0.001$) and cognitive reflection ($\beta = 0.25$, 95% CI [0.14–0.36], $p < 0.001$; see Table 3, right panel). Again, participants had lower scores on the delayed post-test than on the test administered immediately after reading the text. In addition, the higher the participants' ability to override automatic and intuitive responses in favour of more reflective ones, the greater their text comprehension when inferential processes were required. The model explained 52% of the variance (marginal R^2 explained by fixed effects = 12%).

4.3 | Transfer of knowledge

For transfer of knowledge—that is, the ability to go beyond the text and apply the new learned information to new questions—an LMM revealed the main effect of time ($\beta = -0.16$, 95% CI [−0.31 to −0.02], $p = 0.025$) and the interactive effect of condition and time, $\chi^2(1) = 12.33$, $p = 0.002$ (see Table 4).

The significant interaction showed that only in the experimenter-provided highlighting condition ($\beta = -1.03$, $SE = 0.24$, $p < 0.001$) did participants' transfer of knowledge decrease more at the delayed post-test compared to the immediate post-test. Both in the learner-generated highlighting ($\beta = -0.04$, $SE = 0.22$, $p = 0.847$) and in the control ($\beta = -0.10$, SE = 0.28, $p = 0.719$) conditions no significant differences emerged between the two post-tests (see Figure 2). The model explained 54% of the variance (marginal R^2 explained by fixed effects = 7%).

4.4 | Metacognitive calibration of performance

We used calibration bias as a measure of metacognitive calibration to reflect the difference between participants' self-evaluation of their performance and their actual performance. In general, participants tended to underestimate their performance (see Table 5). A linear regression model only revealed the main effect of reading self-efficacy ($\beta = 0.49$, 95% CI [0.15–0.83], $p = 0.005$). The overall main effect of the highlighting condition and the interaction with cognitive reflection were not significant, $F(2,184) = 1.22$, $p = 0.297$, and $F(2,184) = 0.21$, $p = 0.812$, respectively).

The motivational variable significantly positively contributed to calibration bias, which means that readers with higher self-confidence in their ability to comprehend scientific texts were more accurate in evaluating their own performance, that is, they underestimated themselves to a lesser degree than did those with less self-efficacy. The model explained only 6% of the variance (adjusted R^2 = 3%).

RQ2. Contribution of active highlighting to text comprehension.

To answer the second research question, we only considered the reading condition involving learner-generated highlighting. Descriptive statistics for proportion of highlighted information at the three levels and the overall mean are reported in Table 1. A first LMM revealed that the quality of the highlighted information according to the expert model significantly predicted literal text comprehension ($\beta = 5.42$,

TABLE 4 Effects on transfer of knowledge ($N = 170$).

	Transfer		
Predictors	**Estimates**	**95% CI**	***p***
(Intercept)	6.02	5.77–6.27	**<0.001**
Condition [Highlighted]	−0.26	−0.61 to 0.09	0.144
Condition [Highlighting]	−0.10	−0.44 to 0.24	0.552
Post-test	−0.16	−0.31 to −0.02	**0.025**
Prior knowledge	0.02	−0.12 to 0.17	0.731
Reading self-efficacy	0.03	−0.00 to 0.06	0.076
Cognitive reflection	0.08	−0.08 to 0.25	0.310
Condition [Highlighted] *post-test	−0.36	−0.56 to −0.16	**0.001**
Condition [Highlighting] *post-test	0.14	−0.05 to 0.34	0.146
Condition [Highlighted] * cognitive reflection	−0.03	−0.26 to 0.19	0.765
Condition [Highlighting] * cognitive reflection	−0.15	−0.38 to 0.08	0.207
Post-test [T2] * cognitive reflection	−0.04	−0.14 to 0.05	0.378
Condition [Highlighted] *post-test * cognitive reflection	−0.05	−0.18 to 0.08	0.449
Condition [Highlighting] * post- test [T2] * cognitive reflection	0.06	−0.08 to 0.19	0.399
Random effects			
σ^2	1.75		
$\tau_{00\text{-ID}}$	1.78		
ICC	0.50		
N_{ID}	170		
Observations	340		
Marginal R^2/conditional R^2	0.067/0.538		

Note: Highlighted, experimenter-provided highlighting; highlighting, learner-generated highlighting. Bold values indicate statistical significance.

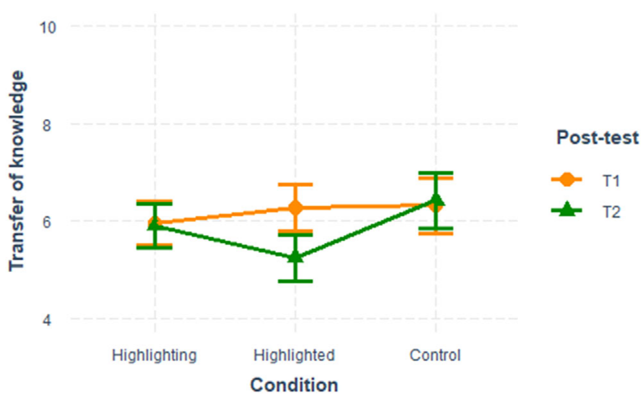

FIGURE 2 Interaction by condition and time in delayed transfer of knowledge.

95% CI [1.88–8.95], $p = 0.003$) regardless of the post-test timing (see Table 6, left panel). The model explained 64% of the variance (marginal R^2 explained by fixed effects = 29%).

From a second LMM for inferential comprehension, only the contribution of cognitive reflection ($\beta = 0.21$, 95% CI [0.01–0.41], $p = 0.041$) emerged, regardless of the post-test timing (see Table 6, right panel). The model explained 49% of the variance (marginal R^2 explained by fixed effects = 10%).

Finally, a third LMM for transfer of knowledge showed, again, the predictive role of the information highlighted by the students ($\beta = 3.17$, CI [0.46–5.89], $p = 0.022$), regardless of the post-test timing (Table 7). The model explained 52% of the variance (marginal R^2 explained by fixed effects = 7%).

5 | DISCUSSION

This study was motivated by the idea of testing whether processing and comprehension of text content can be effectively sustained during digital reading using a popular student strategy when reading on paper, that is, highlighting, which is also easily implemented, from a technical point of view, with digital texts (Dunlosky et al., 2013; Fiorella & Mayer, 2014). Specifically, we examined the effects of both learner-generated and experimenter-provided highlighting compared to a control condition in which participants were simply instructed to read the text. Considering the documented screen inferiority effect, probably determined by a shallower approach to reading on screen than on paper

TABLE 5 Effects on metacognitive calibration of performance ($N = 192$).

	Calibration of performance		
Predictors	**Estimates**	**95% CI**	***p***
(Intercept)	−20.69	−23.51 to −17.88	**<0.001**
Condition [Highlighted]	−1.00	−4.92 to 2.92	0.615
Condition [Highlighting]	−2.22	−6.02 to 1.58	0.250
Prior knowledge	−0.58	−2.17 to 1.01	0.472
Reading self-efficacy	0.49	0.15–0.83	**0.005**
Cognitive reflection	−0.32	−2.19 to 1.55	0.738
Condition [Highlighted] * cognitive reflection	0.77	−1.80 to 3.34	0.555
Condition [Highlighting] * cognitive reflection	−0.04	−2.66 to 2.59	0.977
Observations	192		
R^2/R^2 adjusted	0.063/0.027		

Note: Highlighted, experimenter-provided highlighting; highlighting, learner-generated highlighting. Bold values indicate statistical significance.

TABLE 6 Contribution of the quality of highlighted text information to literal (left) and inferential (right) text comprehension in the condition of learner-generated highlighting ($N = 68$).

	Literal comprehension			Inferential comprehension		
Predictors	**Estimates**	**CI**	***p***	**Estimates**	**CI**	***p***
(Intercept)	9.17	6.94–11.40	**<0.001**	5.51	4.27–6.74	**<0.001**
Quality of highlighted information	5.42	1.88–8.95	**0.003**	1.02	−0.94 to 2.97	0.306
Post-test	−2.19	−4.38 to 0.01	0.051	−0.51	−1.80 to 0.77	0.431
Prior knowledge	0.13	−0.14 to 0.40	0.348	−0.04	−0.19 to 0.11	0.589
Reading self-efficacy	0.09	0.03–0.15	**0.005**	0.01	−0.03 to 0.04	0.715
Cognitive reflection	−0.17	−0.55 to 0.20	0.354	0.21	0.01–0.41	**0.041**
Quality of highlighted information * post-test	−0.13	−3.59 to 3.33	0.941	0.26	0.03–0.49	0.906
Random effects						
σ^2	2.99			1.03		
$\tau_{00\text{-ID}}$	2.96			0.79		
ICC	0.50			0.43		
N_{ID}	68			68		
Observations	136			136		
Marginal R^2/conditional R^2	0.290/0.643			0.102/0.491		

Note: Bold values indicate statistical significance

(Annisette & Lafreniere, 2017), we sought to understand whether this effect could be reduced, at least to some extent, by either asking students to actively selected important information in a text or by providing already highlighted information.

5.1 | Does highlighting affect reading outcomes?

The first research question asked whether the effects of highlighting would emerge on a number of outcome variables such as literal and inferential text comprehension, transfer of newly learned knowledge, and metacognitive calibration of comprehension performance. Based on the most recent literature (Ponce et al., 2022), we hypothesized that active highlighting by university students, who are supposed to have the ability to select relevant information, would be at least as effective as providing a highlighted text, especially for literal comprehension, which was the lower level of text representation at both testing times, if not for inferential comprehension (H1). We also hypothesized there would be no significant differences for transfer of knowledge, which implies going beyond the texts and applying new learned knowledge—in other words, learning from the text (H2). Moreover, we asked whether individual differences in terms of prior knowledge, cognitive reflection, and reading self-efficacy would moderate the possible effects. From the results of LMMs considering both fixed and random effects, H1 was not confirmed, as condition did not significantly differentiate either level of text comprehension. This

TABLE 7 Contribution of the quality of highlighted text information to transfer of knowledge ($N = 68$).

	Transfer of knowledge		
Predictors	**Estimates**	**CI**	***p***
(Intercept)	4.02	2.30–5.73	**<0.001**
Quality of highlighted information	3.17	0.46–5.89	**0.022**
Post-test	0.76	−0.94 to 2.45	0.380
Prior knowledge	0.07	−0.14 to 0.28	0.512
Reading self-efficacy	0.04	−0.00 to 0.09	0.074
Cognitive reflection	−0.06	−0.34 to 0.23	0.699
Quality of highlighted information * post-test	−1.31	−3.99 to 1.37	0.335
Random effects			
σ^2	1.80		
$\tau_{00\text{-ID}}$	1.71		
ICC	0.49		
N_{ID}	68		
Observations	136		
Marginal R^2/conditional R^2 0.070/0.524			

Note: Bold values indicate statistical significance

finding aligns with those that indicate the ineffectiveness of highlighting (e.g., Ben-Yehudah & Eshet-Alkalai, 2018; Peterson, 1992).

However an interactive effect of condition and cognitive reflection emerged, which showed that, only in the experimenter-provided highlighting condition, readers with a higher ability to override easier and intuitive responses in favour of slower and more rational ones had higher scores for the literal or factual level of text representation. Although very partially, this finding parallels the results of the meta-analysis by Ponce et al. (2022), which underlined the benefits of instructor-provided highlighted for both memory and text comprehension, without considering individual differences. We also took into consideration the possible effect of cognitive and motivational differences in the use of this learning strategy. Our results add to the literature on the moderating role that the ability to use more System 2 than System 1 can play. The interesting interaction of condition and cognitive reflection showed that students with a stronger tendency toward slow and rational thinking are those who take more advantage of having relevant information already available to proceed on further elaboration of the text. However, this advantage was only evident when factual understanding was considered. Not surprisingly, delayed text comprehension was lower than comprehension immediately after reading, regardless of condition.

Concerning the deeper level of inferential representation, condition did not play a role, only the individual difference of cognitive reflection had a main effect, which confirms that inferential processes are sustained by engagement in thinking (e.g., Dai & Wang, 2007). As for the literal questions, delayed inferential comprehension was lower than immediate comprehension. It is interesting to note here that, in a very recent study combining active highlighting with another pretty popular student strategy (i.e., annotating) using different task instructions, overall text comprehension and integration did not benefit, either (List & Lin, 2023). However, unlike highlighting and annotating relevant information in the text, task instructions for comprehension monitoring, source evaluation, and corroboration predicted multiple text comprehension. The integration of more than one learning strategy with specific task instructions to elicit specific forms of content processing may also potentially be a more effective combination for the comprehension of a single, complex, and lengthy text. Undoubtedly, digital reading, of either single or multiple texts, requires more investigation to gain deeper knowledge of the role of readers' strategies and other contextual factors, such as specific task instructions or reading goals that can be supportive when acquiring challenging conceptual content from reading on a screen.

We also hypothesized that the effect of highlighting would not lead to greater transfer of knowledge, as the function of the strategy is to select relevant information, while the ability to apply newly learned knowledge implies deep content processing that might not be supported by the cognitive process related to highlighting. Our H2 was confirmed, as transfer of knowledge did not differentiate across conditions, which means that neither passive nor active highlighting supported the deepest level of comprehension that would reflect being able to go beyond text information. Although we cannot exclude the possibility of detecting an effect of condition on transfer of knowledge in a larger sample, we can reasonably state that, in our data, learning from the text was not related to reading condition.

Another relevant result is that at the delayed post-test, only learners in the experimenter-provided highlighting condition showed a significant decrease from immediate post-test in their transfer performance compared to those in the other reading conditions. Taking together this outcome and that regarding literal comprehension, we can say that experimenter-provided highlighting is not more effective per se, but only in combination with higher ability to reflect cognitively, and that it might also have a negative impact on long-term learning from a text. We can thus speculate that, if the given support does not lead to a less superficial approach to text reading, the transfer of knowledge that implies deep learning decreases over time.

Overall, individual differences—specifically cognitive reflection and self-efficacy—contributed to the differing levels of text comprehension uniquely or in interaction with the highlighting strategy, which indicates that participants' approach to the reading task, which is determined by cognitive and motivational components, plays a major role. It should also be pointed out that in the active highlighting condition, students took more time reading than did those in the other conditions, but the effects for comprehension and transfer of knowledge were not significantly different across the three conditions. Learner-generated highlighting may, therefore, be considered to be less effective in this study. However, we also emphasize that, when considering school learning, speed is not always a crucial outcome. In many situations, accuracy is more important than speed, especially when the difference in time is only a matter of a few minutes during the execution of a task demanding the comprehension of a complex and lengthy text with science content.

We took an explorative approach regarding the effect of condition on the calibration of comprehension performance. The findings revealed that only the motivational variable of reading self-efficacy contributed to the metacognitive ability, which means that regardless of the reading condition, participants who had a stronger belief that they would be able to comprehend scientific texts were also more accurate in self-evaluating how well they responded to post-test questions, avoiding large underestimation, in our case. This finding is aligned with research on the individual factors that may contribute to calibration and self-regulated learning (Alexander, 2013; Stone, 2000).

5.2 | Does the quality of learner-generated highlighting contribute to reading outcomes?

The second research question asked whether the quality of the information actively highlighted by the participants in one of the three reading conditions would predict their text comprehension and transfer of knowledge. We hypothesized that the quality of highlighted content would be associated with the outcome variables (H3). Even if reading condition per se did not differentiate the results across conditions, but only in combination with cognitive reflection for literal text comprehension, it is relevant to know whether the quality of the information actively highlighted by the participants while reading predicts the outcomes of their reading. Our H3 was substantially confirmed, as literal text comprehension and transfer of knowledge were significantly predicted by quality scores. Inferential text comprehension, however, was only predicted by the individual difference in cognitive reflection, which means that the process of connecting information from the text with prior knowledge involves thinking as activated by System 2.

5.3 | Educational implications

Although not significant in relation to reading condition, the results have implications for implementing the strategy of highlighting during digital reading. The effective use of this strategy probably requires explicit instructions, practice, or training. Research involving high school students reading on paper has shown that when they were trained to highlight, they outperformed the control group without training in comprehension. Importantly, the training focused on providing an overview of the strategy and then elaborating each of its steps, while providing an opportunity for practice. In addition, the training incorporated a self-regulation component, as students were taught that self-regulated learning involves checking their learning from an external perspective that incorporates self-observation, self-evaluation, and how the strategy is applied to act consequentially (Leopold & Leutner, 2015, Exp. 1). With older students it may also be useful to review the nature of the highlighting strategy and create or refine their awareness of the conditions under which the use of the strategy is really effective. Although highlighting is very popular, its application is often suboptimal, and even university students may not consider that the quality of strategy use makes a difference in their text comprehension and learning from text.

It is also worthwhile to combine this simple strategy with other deeper, meaning-making strategies that are intended to contribute to generative learning by sustaining not only the selection of relevant information, but also its organization and integration, such as annotating relevant information or constructing an outline of the text, which has proven to be effective for reading on paper (Fiorella & Mayer, 2015). It is also important to consider what and how much support should be provided for modelling the strategy and optimizing its potential benefits in relation to individual differences. The interesting amplification effect of higher cognitive reflection and already highlighted text suggests that readers who are better able to resist automatic thinking may also invest more effort in the task and thus take more advantage of the provided support (Toplak et al., 2014). Importantly, the quality of active highlighting contributes to text comprehension. Students should be aware that quality matters, so the selection of information to be highlighted implies being able to distinguish the relevance of information in the text.

5.4 | Limitations

As with any study, this one is not without limitations. First, as aforementioned above, the reliability scores for post-test knowledge are lower than desirable. We pointed out that, in our case, we devised both the text and the tests for the study based on a previous investigation, and this may have resulted in suboptimal reliability for the tests, even if all materials were previously tested for readability and comprehensibility with a few readers who then did not participate in the study. Undoubtedly, stronger data would require assessment instruments with higher internal consistency. Second, we considered three control variables for individual differences that may contribute to the comprehension of complex scientific tests. For keeping the experiment within an acceptable time frame, we sacrificed the assessment of reading comprehension, given that all participants had passed a test on comprehension of informative texts during the admission examination, which ensured that all they had sufficient ability to comprehend texts like the one used in this study. Nevertheless, their reading comprehension ability may have differed and taking this into consideration would make it possible to know if and to what extent the variable still differentiates reading outcomes even at higher levels of education.

Third, participants in the control condition were not explicitly recommended not to use any strategies when reading the text. We opted for not constraining they too much and demotivating their reading of a lengthy complex text about an unfamiliar topic. However, at the end of the experiment, we asked whether they had used any strategy besides reading and rereading, as well as checking what they had done digitally. Given that it was online study, however, it is still possible that they used a strategy without admitting it. We did not consider, however, the data of students in the control condition who stated that they had used strategies. For this reason we had an unbalanced number of participants across the three reading conditions. Fourth, and related to the previous limitation, is the number of participants who did not return for the delayed post-test, which also contributed to the unbalanced number of participants across conditions.

5.5 | Conclusions

Despite these limitations, this study advances current knowledge about highlighting, both active and passive, as a strategy for digital reading. This study confirms that the simple and technically easy to implement strategy of learner-generated highlighting is not effective per se. Passive or experimenter-provided highlighted is more effective than active highlighting but only in combination with the ability to reflect and use System 2, and this only for the lower level of text comprehension (i.e., literal or factual text comprehension). For the deeper levels of text comprehension, only individual differences such as cognitive reflection and reading self-efficacy were significantly associated with inferential comprehension and transfer of knowledge, respectively. The interesting and positive amplification effect of already highlighted text and higher cognitive reflection suggests that readers who are better able to resist automatic thinking may also invest more effort in the task, thus taking more advantage of the provided support (Toplak et al., 2014). Importantly, the quality of active highlighting contributes to literal text comprehension and transfer of knowledge. Even if learner-generated highlighting may not be effective per se compared to other reading conditions, what students highlight matters for their literal text comprehension and learning from the text. Finally, our findings can help in deriving specific recommendations for the use of the popular strategy of highlighting to aid learning from digital texts.

AUTHOR CONTRIBUTIONS

Lucia Mason: Conceptualization; methodology; supervision; writing – original draft; writing – review and editing. **Angelica Ronconi:** Data curation; formal analysis; investigation; writing – original draft. **Barbara Carretti:** Conceptualization; methodology; supervision; validation. **Sara Nardin:** Data curation; investigation; software. **Christian Tarchi:** Conceptualization; methodology; resources; validation.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/jcal.12903.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Lucia Mason https://orcid.org/0000-0001-7134-0510
Angelica Ronconi https://orcid.org/0000-0001-8944-8765
Barbara Carretti https://orcid.org/0000-0001-5147-7544
Christian Tarchi https://orcid.org/0000-0003-4013-4794

REFERENCES

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. https://doi.org/10.1016/j.learninstruc.2012.10.003

Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences*, *19*, 252–256. https://doi.org/10.1016/j.lindif.2008.09.002

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Ben-Yehudah, G., & Eshet-Alkalai, Y. (2018). The contribution of text-highlighting to comprehension: A comparison of print and digital reading. *Journal of Educational Multimedia and Hypermedia*, *27*, 153–178.

Bråten, I., Ferguson, L. E., & Anmarkrud, O. (2013). Prediction of learning and comprehension when adolescents read multiple texts: The roles of word-level processing, strategic approach, and reading motivation. *Reading and Writing: An Interdisciplinary Journal*, *26*(3), 321–348. https://doi.org/10.1007/s11145-012-9371-x

Chen, F., Sakyi, A., & Cui, Y. (2021). Linking student, home, and school factors to reading achievement: The mediating role of reading self-efficacy. *Educational Psychology*, *41*(10), 1260–1279. https://doi.org/10.1080/01443410.2021.1953445

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

Dai, D. Y., & Wang, X. (2007). The role of need for cognition and reader beliefs in text comprehension and interest development. *Contemporary Educational Psychology*, *32*, 332–347. https://doi.org/10.1016/j.cedpsych.2006.05.002

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

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2019). 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.cedpsych.2020.101870

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students'learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

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

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, *39*(2), 75–85. https://doi.org/10.1016/j.cedpsych.2014.01.001

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity*: *Eight learning strategies that promote understanding*. Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*, 717–741. https://doi.org/10.1007/s10648-015-9348-9

Florit, E. D., Carli, P., Lavelli, E., & Mason, L. (2022). Digital reading in beginner readers: Advantage or disadvantage for comprehension of narrative and informational linear texts? *Journal of Computer Assisted Learning*, *39*, 432–445. https://doi.org/10.1111/jcal.12754

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology*, *59*, 358–364. https://doi.org/10.1037/h0036750

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42. https://doi.org/10.1257/089533005775196732

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., 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

Gurung, R. A. R., Weidert, J., & Jeske, A. (2010). Focusing on how students study. *Journal of the Scholarship of Teaching and Learning*, *10*, 28–35.

Johnson, L. L. (1988). Effects of underlining textbook sentences on passage and sentence retention. *Reading Research and Instruction*, *28*, 18–32. https://doi.org/10.1080/19388078809557955

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Harcourt College Publishers.

Kintsch, W. (1998). *Comprehension: A paradigm or for cognition*. 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

Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. R Package Version 1.4.8. https://CRAN.R-project.org/package=emmeans

Leopold, C., & Leutner, D. (2015). Improving students' science text comprehension through metacognitive self-regulation when applying learning strategies. *Metacognition and Learning*, *10*, 313–346. https://doi.org/10.1007/s11409-014-9130-2

Leutner, D., Leopold, C., & den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy: A training experiment. *Journal of Psychology*, *215*, 174–182. https://doi.org/10.1027/0044-3409.215.3.174

List, A., & Lin, C.-J. (2023). Content and quantity of highlights and annotations predict learning from multiple digital texts. *Computers & Education*, *199*, 104791. https://doi.org/10.1016/j.compedu.2023.104791

Lorch, R. F. (1989). Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, *1*, 209–234. https://doi.org/10.1007/BF01320135

Lucisano, P., & Piemontese, M. E. (1988). GULPEASE: Una formula per la predizione della difficoltà Dei testi in lingua italiana [GULPEASE: A formula to predict the difficulty of texts in Italian]. *Scuola e Città*, *3*, 110–124.

Mason, L., Zaccoletti, S., Scrimin, S., Tornatora, M. C., Florit, E., & Goetz, T. (2020). Reading with the eyes and under the skin: Comprehending conflicting digital texts. *Journal of Computer Assisted Learning*, *36*, 89–101. https://doi.org/10.1111/jcal.12399

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288. https://doi.org/10.1080/01638539609544975

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

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science*, *13*(3), 390–407. https://doi.org/10.1177/1745691617710510

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Nist, S. L., & Hogrebe, M. C. (1987). The role of underlining and annotating in remembering textual information. *Reading Research and Instruction*, *27*, 12–25. https://doi.org/10.1080/19388078709557922

Nunnally, J. C. (1978). *Psychometric theory*. McGraw Hill.

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence*, *90*, 101614. https://doi.org/10.1016/j.intell.2021.101614

Ozuru, Y., Dempsey, K., & Mcnamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Parkinson, M. M., & Dinsmore, D. L. (2018). Multiple aspects of high school students' strategic processing on reading outcomes: The role of quantity, quality, and conjunctive strategy use. *British Journal of Educational Psychology*, *88*, 42–62. https://doi.org/10.1111/bjep.12176

Peterson, S. E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction*, *31*, 49–56. https://doi.org/10.1080/19388079209558078

Ponce, H. R., Mayer, R. E., & Méndez, E. E. (2022). Effects of learner-generated highlighting and instructor-provided highlighting on learning from text: A meta-analysis. *Educational Psychology Review*, *34*, 989–1024. https://doi.org/10.1007/s10648-021-09654-1

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2015). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*, 453–469. https://doi.org/10.1002/bdm.1883

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ronconi, A., Veronesi, V., Mason, L., Manzione, L., Florit, E., Anmarkrud, Ø., & Bråten, I. (2022). Effects of reading medium on the processing, comprehension, and calibration of adolescent readers. *Computers & Education*, *185*, 104520. https://doi.org/10.1016/j.compedu.2022.104520

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). Routledge.

Sinatra, G. M., & Broughton, S. W. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, *46*, 374–393. https://doi.org/10.1002/RRQ.005

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, *12*(4), 437–475. https://doi.org/10.1023/A:1009084430926

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education*, *160*, 104034. https://doi.org/10.1016/j.compedu.2020.104034

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*, 147–168. https://doi.org/10.1080/13546783.2013.844729

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). Erlbaum.

van Gog, T. (2021). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 221–230). Cambridge University Press.

Wickham, H. (2016). *Elegant graphics for data analysis*. Springer.

Winchell, A., Lam, A., & Mozer, M. (2020). Highlights as an early predictor of student comprehension and interests. *Cognitive Science*, *44*, e12901. https://doi.org/10.1111/cogs.12901

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, *11*(2), 87–95. https://doi.org/10.1080/00461520903433554

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2

Yue, C. L., Storm, B. C., Kornell, N., & Bjork, E. L. (2015). Highlighting and its relation to distributed study and students' metacognitive beliefs. *Educational Psychology Review*, *27*(1), 69–78. https://doi.org/10.1007/s10648-014-9277-z

Zhang, M. (2013). Supporting middle school students' online reading of scientific resources: Moving beyond cursory, fragmented, and opportunistic reading. *Journal of Computer Assisted Learning*, *29*(2), 138–152. https://doi.org/10.1111/j.1365-2729.2012.00478.x

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* (Vol. 574). Springer.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mason, L., Ronconi, A., Carretti, B., Nardin, S., & Tarchi, C. (2024). Highlighting and highlighted information in text comprehension and learning from digital reading. *Journal of Computer Assisted Learning*, *40*(2), 637–653. https://doi.org/10.1111/jcal.12903