INTERNATIONAL LITERACY ASSOCIATION
READING RESEARCH QUARTERLY

Reading Text When Studying in a Second Language: An Eye-Tracking Study

Nicolas Dirix
Heleen Vander Beken
Ellen De Bruyne
Marc Brysbaert
Wouter Duyck
Ghent University, Belgium

ABSTRACT

The authors investigated how eye movements are influenced by different reading goals in participants' first (L1) and second language (L2). Participants read or studied the contents of texts while their eye movements were recorded. One group was asked to read L1 and L2 texts as they would read any expository text (informational reading). Another group was asked to study L1 and L2 texts for subsequent tests involving true/false questions (study condition). After reading, all participants, including those in the informational reading condition, completed the true/false tests without being able to further consult the texts, which allowed the authors to investigate the extent to which reading goal and text language affect recognition memory for texts. In general, more reading time was spent on studying than on informational reading, which also resulted in higher test scores in the study condition. The L2-processing cost was larger in the study condition than in the informational reading condition: Participants needed approximately 20% more time to study L2 texts. The results of various eye movement measures suggest that this is caused by slower word recognition processes and a smaller amount of information that can be processed simultaneously in L2. This was true not only for the first reading of the text but also for the rereadings in the study condition. Interestingly, the additional time for L2 studying seemed to compensate for the less efficient processing, as the recognition test scores were the same in L2 as in L1.

In countries with a native language other than English, there has been an increase in the use of English as a Medium of Instruction (Dafouz & Camacho-Miñano, 2016; Doiz, Lasagabaster, & Sierra, 2013). One of the consequences in higher education is an increase in the use of English textbooks. Bilinguals, defined by Grosjean (2008) as "people who use two or more languages (or dialects) in their everyday lives" (p. 10), are thus expected to understand and remember the content of textbooks in a second language (L2) in which they are less proficient.

Our focus in the present study was on the reading processes for academic texts written in the native language (L1) or L2, as well as memory for the text contents. Specifically, we investigated the effect of language on two distinct reading goals: informational reading, particularly reading out of interest, and studying, which we define in the current study as memorizing factual information to pass a subsequent recognition test.[1] We used eye tracking to identify similarities and differences in L1 and L2 text processing.

Models of purposeful reading try to explain how the reading process is affected by reading goals. According to the reading as problem solving

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 0(0)
pp. 1–27 | doi:10.1002/rrq.277

model (RESOLV; Britt, Rouet, & Durik, 2018; Rouet, Britt, & Durik, 2017), the reader constructs a context model and a task model. The context model consists of several features, such as the nature of the reading material (e.g., academic texts) and the reading goal. For example, an academic text can be read to keep up to date with new developments or for pleasure (i.e., informational reading), but it is also often material for subsequent students' exams (i.e., it involves studying). The task model entails the subjective interpretation of the reading goal, which drives the reading approach (e.g., a quick skim of the text, a thorough analysis of its contents).

Because of the lack of research investigating whether (and how) reading processes changes when studying in L1 or L2, the current study is an investigation of the interaction between two types of contextual features: the specific reading goal and the language of the text. This is inevitable, given that multiple theoretical accounts in the field of bilingualism assume that L2 processing is less efficient than L1 processing. For example, according to the resource hypothesis, L2 processing is more demanding for working memory (Sandoval, Gollan, Ferreira, & Salmon, 2010). This higher cognitive load results in a smaller capacity for higher order processing (e.g., building a mental model, inference making). Another example is the weaker links hypothesis (Gollan, Montoya, Cera, & Sandoval, 2008), which assumes that representations of L2 words function as low-frequency L1 words: weaker and less detailed. These hypotheses can also be related to the bilingual interactive activation plus model (BIA+; Dijkstra & van Heuven, 2002). According to this model of bilingual reading, the linguistic networks of the two languages are shared, but the L2 representations at the early (orthographic/phonological) stages have a lower resting state level, so it is more time and resource demanding to activate them.

Empirical evidence supports these assumptions: L2 words and sentences are indeed read more slowly than their L1 counterparts (Cop, Drieghe, & Duyck, 2015; Whitford & Titone, 2012), and students' vocabulary retention is worse for L2 than L1 words (Gablasova, 2014). Furthermore, general L2 text comprehension seems to be somewhat weaker, as L2 readers show reduced inference making (Horiba, 1996; Pérez, Hansen, & Bajo, 2018). Questions then arise as to whether these L2 costs have an influence on the study process of academic texts, relative to L1, and whether study language affects the memory for the contents of the text.

A few studies have investigated the matter. First, two research teams found that participants spend more time studying texts in L2 than in L1 (Chen & Donin, 1997; Donin, Graves, & Goyette, 2004). In both studies, participants read short L1 and L2 texts and orally recalled the contents in the same language as they had read them. The results on this subsequent free-recall test showed inconsistencies between the studies. In Chen and Donin's (1997) study, the participants were Chinese–English bilingual university students (enrolled in a science program at an English-speaking university). Here, no L1–L2 difference was found in recall, although Vander Beken and Brysbaert (2018) remarked that there was a trend toward an L2 recall cost in Chen and Donin's study, which may have failed to reach statistical significance because of a lack of power. In Donin et al.'s (2004) study, English-speaking male officers enrolled in a French-language program in the Canadian Forces took part and were divided into high and low L2 proficiency groups. This time, an L2 recall cost appeared, but this was mainly caused by a higher L1 recall performance for the low L2 proficiency group. The proficiency level was possibly confounded with prior knowledge, as the participants of the low L2 proficiency group were more acquainted with the contents of some of the texts. It is not surprising that L2 studying takes longer, for example, given the slower processing rate in L2 reading (e.g., Cop, Drieghe, & Duyck, 2015; Whitford & Titone, 2012). These studies, however, investigated differences between L1 and L2 reading with a reading goal more related to informational reading, whereas differences in the time course of the reading pattern while studying in L1 and L2 remain largely unexplored. Furthermore, it is interesting that this slower L2 processing does not always result in a diminished memory performance for the contents of the texts.

A second important finding was that the test type matters. In a free-recall test, information must be retrieved from the text without memory cues, whereas recognition tests such as true/false statements provide very strong memory cues to the facts mentioned in the text. In Vander Beken and Brysbaert's (2018) study, bilingual Dutch–English undergraduate students studied two single-page expository texts in L1 or L2 and received a free-recall test for one of the texts and a true/false judgment test for the other. Participants' performance on the recall test was much better in L1 than in L2, but there was no difference on the true/false recognition test. In a follow-up study, retaining only the recognition test, Vander Beken, Woumans, and Brysbaert (2018) found that recognition performance remained comparable in L1 and L2 even after a delay of up to 30 days. This suggests that the memorization of facts is similar in L1 and L2 and that the memory traces encoded in L2 do not decay at a faster rate than in L1. However, the L1–L2 difference on the recall test suggests that there is an L2 cost when access to the encoded information is not supported by memory cues.

Vander Beken and Brysbaert (2018) provided two possible explanations for this recall cost. One possibility is that it may be more difficult for participants to reproduce their thoughts in L2 than in L1 (e.g., Joh, 2006). In language production tasks, bilinguals indeed make fewer and slower correct responses and show delayed retrieval in L2 (e.g., Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007; Sandoval et al., 2010).

The other explanation involves the mental model of the text. It has been proposed that to comprehend the contents of a text and retain them in the memory, the reader forms a mental model (e.g., Britton, 1994; Goldman & Varnhagen, 1986; Kintsch, 1974; van den Broek, Young, Tzeng, & Linderholm, 1999). In van den Broek et al.'s (1999) landscape model, it has been suggested that each proposition in a text is translated into a specific pattern of activation related to the concepts involved. Although the information is assumed to be independently stored language at the conceptual level, an important aspect of van den Broek et al.'s model is that the propositions are enriched by concepts that are coactivated together with the text words and help in linking the individual propositions. Vander Beken and Brysbaert (2018) proposed that a difference in richness between the propositions created in L1 and L2, caused by weaker (co)activation of concepts in L2, could lead to the L2 recall cost because there are fewer memory cues connecting the propositions. The recall cost could thus (at least partially) be situated at the encoding stage. Bilingual frameworks such as the weaker links hypothesis indeed suggest that the links between word forms and the underlying concepts are weaker in L2 than in L1 (Gollan et al., 2008). Therefore, the spread of activation through the conceptual network could be less extensive in L2, resulting in a less organized landscape.

The building of a landscape model of the text could also account for the findings of Donin et al. (2004), where prior knowledge turned out to be important. If the reader is already familiar with the concepts in a text, his or her larger conceptual network about the topic may help the reader build a richer landscape. At present, it is not clear how this would interact with text language, although it might be hypothesized that it would be particularly helpful for the harder L2 condition.

The early stages of word recognition and sentence comprehension may already be a bottleneck in the development of a rich mental model in L2. According to the BIA+ model (Dijkstra & van Heuven, 2002), L2 lexical representations have a lower resting state level, so it is more time and resource demanding to activate them. It has indeed been found that readers look longer at L2 words than at L1 words, from the first encounter onward, which arguably signals more effortful lexical access (Cop, Drieghe, & Duyck, 2015; Whitford & Titone, 2012). As in most word recognition models, lexical access precedes semantic activation, so it can be assumed that it will take longer to activate a word's meaning in L2, which could lead to a poorer development of the mental model of a text.

To summarize, two main findings showed up in studies comparing studying texts in L1 and L2. First, performance on a yes/no recognition test was equal in L1 and L2, but an L2 cost appeared in more demanding recall tests (Donin et al., 2004; Vander Beken & Brysbaert, 2018). Second, participants need more time to study in L2 than in L1 (Chen & Donin, 1997; Donin et al., 2004).[2] This indicates that a specific feature, text language, influences the reading process (in line with the RESOLV model; Rouet et al., 2017). Bilingual frameworks such as the weaker links hypothesis (Gollan et al., 2008) and the BIA+ model (Dijkstra & van Heuven, 2002) provide an explanation for the finding that L2 processing is more effortful compared with L1. Yet, questions remain about the differences between L1 and L2; that is, it is not clear where this longer L2 study time originates from. For instance, we do not know whether the longer L2 study time is entirely caused by the longer time needed to identify the individual words (processing the so-called shallow structure of the text) or also by extra time needed to integrate the sentences into a mental model of the text (building the deep structure of the text). A technique that can provide a more detailed view is eye tracking.

Eye Movement Research

Eye tracking is a noninvasive technique with which the eye movements can be registered while a person is reading a text. It has a high precision in time and space and allows the researcher to see where in the text the person moves his or her eyes to (with so-called saccades) and where and how long they stay there (measured as a fixation). In this way, a lot of valuable information becomes available about how people process a text (Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Rayner, 1998, 2009; Rayner, Chace, Slattery, & Ashby, 2006). Eye tracking can, for example, show the researcher how reading differs when people read a text for pleasure and when they study it for a test.

A distinction is usually made between measures of early and late processing. *Early processing* refers to word recognition and is typically measured with dependent variables such as fixation duration and saccadic amplitude. First, there is fixation duration: How long does it take before the reader moves his or her eyes? This is particularly relevant when there is only one fixation on a word, because then the fixation duration can be taken as a measure of word-processing time. The second measure is the amplitude of the eye movement, the so-called saccade size. In particular, the size of forward saccades is important for early processing, as it indicates how much information can be processed simultaneously (typically seven to nine characters; Liversedge & Findlay, 2000). These measures give a good idea of word recognition and the processing of the shallow text structure. In general, texts with easy words will have short fixation durations and longer forward saccades. Similarly, proficient readers will have short fixation durations and longer forward saccades than less competent readers will.

The late processing eye movement measures involve regressions to earlier parts of the text and rereading of (parts of) the text. These include measures such as the regression rate or time (the number or summed time of fixations in the regression), and total reading time (the total time the eyes spent on a word or part of a text). These measures are more related to understanding the propositions in the text and linking them to the mental model (i.e., the deep structure), although short-range regressions could also indicate that one of the previously read words was not well understood. Because the late processing measures all involve rereading of text, the stage in which the early processing is measured is often called first-time reading or first-pass reading.

A possible limitation is that for detailed eye movement analysis, reading from a screen with the head fixed in a head- and chinrest is needed. Several studies investigated whether there are differences in reading behavior and text comprehension between reading printed texts and reading on a screen. Whereas the eye movement patterns and reading speed are similar for print and digital reading (Noyes & Garland, 2008; Zambarbieri & Carniglia, 2012), it was found that recall for key points is slightly better for print than for digital reading (Singer & Alexander, 2017) and that text comprehension is somewhat better for reading on paper than on screen (Delgado, Vargas, Ackerman, & Salmerón, 2018). It seems that the specific task circumstances play an important part because, for example, the text subject is of importance: A comprehension advantage for digital reading was found for English language arts and social studies, but comprehension was better with printed text for mathematics (Kingston, 2008).

L2 Reading

Eye tracking has been applied to investigate word-level characteristics such as word frequency. It has been found that the word frequency effect is larger in L2 than in L1, resulting in longer fixation times on low-frequency words in L2 compared with low-frequency words in L1 (Cop, Keuleers, Drieghe, & Duyck, 2015; Whitford & Titone, 2012). This is relevant for the present study because academic textbooks often contain difficult, domain-specific, low-frequency words, so we can expect longer fixation times when studying in L2.

Concerning the text level, a study by Cop, Drieghe, and Duyck (2015) revealed a difference in processing efficiency between L1 and L2 reading for pleasure: Participants made more and longer fixations and smaller saccades and skipped fewer words in L2. The authors attributed these differences to slower word activation and verification processes in L2 and suggested that a general reduction in the rate of lexical processing could explain why readers are less good at processing text in L2. Note that in Cop, Drieghe, and Duyck's study, the material was a murder mystery. As academic texts often contain complex syntactic sentences, this could be an additional challenge in L2, both when reading and when studying texts. The exact changes in text demands for L2 readers are largely unknown, leading to the need for the current study.

Reading Goal and Information Centrality

Next to the language of the text, two other contextual features have been investigated with eye tracking. These are to what extent text processing depends on information centrality (i.e., how important the proposition is within the text; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Miller & Keenan, 2009; Yeari, Oudega, & van den Broek, 2017) and on the reading goal (Yeari, van den Broek, & Oudega, 2015). For example, it takes participants longer to read a sentence for the first time (i.e., the first-pass reading time) if it contains central information than when it contains peripheral information (Hyönä & Niemi, 1990). In Yeari et al.'s (2015) study, this was true for various reading goals (informational reading, reading to prepare a presentation, reading to prepare for a closed-ended question test, and reading to prepare for an open-ended question test). However, the difference between central and peripheral information disappeared for the total reading time of the texts in the two conditions with test purposes, arguably because the readers realized that questions were unlikely to be limited to the central information. Furthermore, participants had longer total reading times and made more fixations when reading to prepare for tests compared with their informational reading. These results indicate differences in text processing based on the reading goal. Regarding studying, a text is more thoroughly read when a test is expected, and more attention is directed toward peripheral ideas when rereading the text for test preparation. A remaining question that we addressed in this study is whether this information centrality effect for different reading goals is similar in L1 and L2. As previous research has suggested that L2 reading is more demanding (Cop, Drieghe, & Duyck, 2015; Pérez et al., 2018; Whitford & Titone, 2012), this might induce changes in attention to either central or peripheral information.

The processing differences in Yeari et al.'s (2015) study also resulted in different accuracy scores on subsequent recognition tests: Test scores were higher in the two test purpose conditions than for informational reading, and participants' accuracy was higher on central information questions than on peripheral information questions. This was true for all reading goals. Finally, Yeari et al. reported relatively low correlations between total reading times and test accuracy, indicating that the strength of a memory representation is only partially determined by the time spent on fixating the relevant information.

In the current study, we extended the investigation of different reading goals to the bilingual domain, as no eye movement study currently exists of reading to memorize information in L2. In reading models such as RESOLV (Rouet et al., 2017), text language is often not explicitly mentioned as a context feature, although the previously discussed research suggested that it could have a major impact on reading processes. Furthermore, Yeari et al.'s (2015) study showed differences in information centrality effects, as shown in reading times, depending on the reading goal, with more importance attached to peripheral information when the reader expects a test. In the current study, we investigated whether a similar pattern of results as a function of reading goals would be found in L2.

The Current Study

We presented participants with four texts that are representative for materials students have to study in higher education: two in L1 (Dutch) and two in L2 (English). As such, language was manipulated within subjects. The reading goal was manipulated between subjects: Half of the participants had to memorize the content of the texts to prepare for a true/false judgment test, and the other half did an informational reading of the texts. Yet, all participants received the same tests afterward. We monitored the eye movements of the participants during their reading of the texts to gather information on both early, shallow and late, deeper processing (e.g., first-time reading vs. rereading) so we could look in detail at how the initial processing of the texts differed as a function of language and reading goal and how subsequent processing differed.

We coded the information units in our texts according to their information centrality so we could assess the effect of this variable on text processing under the influence of the different context features (language and reading goal). Information units according to Yeari et al. (2015) consist of a "main predicate, its arguments…and the adjectives and/or adverbs of these arguments" (p. 1076).

The design allowed us to address the various outstanding issues in the domain of bilingual reading and studying. In particular, we addressed four research questions:

1. What are the differences between the eye movement patterns in L1 and L2 studying?

When preparing for a test, L2 studying takes longer than L1 studying (e.g., Chen & Donin, 1997; Donin et al., 2004), but it is not clear where this longer study time comes from. Studies on reading, such as Cop, Drieghe, and Duyck (2015), have suggested that the difference may originate from the initial reading (encoding the shallow structure), so we expected longer and more fixations for the initial reading of L2 text. Further, it is interesting as to whether there are still differences between L1 and L2 studying after first-pass reading (i.e., once the information is initially encoded).

2. Is the effect of reading goal on the eye movement pattern similar in L1 and L2?

In L1 eye movement research, it was found that a specific reading goal (informational reading vs. preparation for a test) affects the reading strategy for a text and that this is particularly true for peripheral information units (Yeari et al., 2015). Our L1 results serve as a replication of this finding. Furthermore, the L2 pattern can reveal whether participants show the same adaptation of their reading process to the reading goal in L2. In L1, we expected a pattern similar to that found by Yeari et al. (2015): (a) longer reading times and more fixations for the study condition compared with reading, (b) a longer first-pass time for central than peripheral information in both reading goals, and (c) an interaction between reading goal and information centrality for the total reading time (the central–peripheral difference remains for reading but disappears for studying). We expected a similar pattern in L2 and assumed that each possible deviation could be an indication of less optimal processing.

3. Is the memory trace, as tested by a cued recognition test, affected by language, reading goal, and/or information centrality?

Several studies showed that memory performance is not always equal in L1 and L2, particularly in recall tests (Donin et al., 2004; Roussel, Joulia, Tricot, & Sweller, 2017; Vander Beken & Brysbaert, 2018). In the current study, we investigated performance on a true/false judgment task, so we expected that the accuracy scores would be equal in the L1 and L2 study conditions, following the results of Vander Beken and Brysbaert (2018). In combination with our expectations for the reading measures, this would mean that although the memorization of L2 information comes with an encoding cost, the less efficient and slower L2 text processing does not harm later recognition processes. New in the present study is that we could examine whether this is true for both central and peripheral information units in the texts. Furthermore, we expected higher test accuracy in the studying condition than in reading and better scores for central information than for peripheral information (cf. Yeari et al., 2015).

4. Are fixation times of an information unit an adequate predictor of test accuracy scores of that unit?

Finally, a lot is known about the underlying (cognitive) processes that influence eye movements (e.g., Boston et al., 2008; Rayner, 1998), but to date, there has been only one study (to our knowledge) that investigated whether longer fixation times correspond to an improved memory trace (Yeari et al., 2015). In general, the researchers found low correlations, but in a separate analysis of the closed-ended question condition (which is similar to our testing condition), the correlation between total reading time and accuracy score was statistically significant. As the true/false statements of our tests correspond to the individual information units, our experiment was perfectly suited to further investigate the relation between fixation time and test score. We expected to obtain a similar result as Yeari et al. (2015).

To answer the research questions properly, we decided to include several covariates in the analysis of the eye movement measures and the accuracy scores. In eye movement research, there are well-established findings of variables influencing reading times, such as word frequency (Cop, Keuleers, et al., 2015; Whitford & Titone, 2019), word length (Drieghe, Brysbaert, Desmet, & De Baecke, 2004; Rayner, Slattery, Drieghe, & Liversedge, 2011), the number of words in the unit (or unit length; Cop, Drieghe, & Duyck, 2015), and language proficiency (both in L1 and L2; Cop, Keuleers, et al., 2015; Whitford & Titone, 2012). As we did not specifically control these variables in our experimental design or test material, we found it important to include them in the analyses.

Importantly, some of the abovementioned studies showed language differences in the effects of these variables on reading behavior, such as a larger frequency effect in L2 reading than in L1 reading (Cop, Keuleers, et al., 2015; Whitford & Titone, 2019). The BIA+ model (Dijkstra & van Heuven, 2002) offers an explanation for this specific effect: The frequency effect is driven by the amount of exposure to a word, leading to lower resting state levels for low-frequency words, which in turn requires more time to recognize these words. For a bilingual reader, language differences occur in the amount of exposure to the words in both languages (there is less exposure to L2 words), causing a larger frequency effect in L2 than in L1. Furthermore, the RESOLV model (Rouet et al., 2017) suggests that contextual (e.g., language) and task (e.g., reading goal) specifications influence reading behavior. As such, we decided to include interactions among language, reading goal, and the control variables.

Following the same reasoning, for the accuracy score analysis, we included factors that could influence the memorization of information: text perception (Vander Beken & Brysbaert, 2018), language proficiency (Droop & Verhoeven, 2003), reading motivation (Andreassen & Bråten, 2010), and prior knowledge about the topic (Coiro, 2011; Kendeou & van den Broek, 2007).

Method

Participants

Eighty first-year undergraduate students in psychology or education sciences at Ghent University took part in the experiment (mean [*M*] age = 19.39 years, standard deviation [*SD*] = 4.66 years; 67 females). They were all Dutch native speakers who received formal English education from age 14 onward and were exposed to English regularly through (online) media. Hence, they were proficient in English but L1-dominant and late bilinguals (see Table 1 for language proficiency ratings). Their average score on the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) was 73, which corresponds to Common European Framework of Reference for Languages level B2 (upper intermediate proficiency). As the majority of their handbooks are in English, and they regularly need to consult academic articles for coursework, these students read academic texts in L2 on a daily to weekly basis. The participants were asked to sign an informed consent and received course credit, as well as an additional payment of €5 for their participation. All participants had 20/20 or corrected-to-normal vision.

Materials

Texts

We used four texts in the current study: two expository texts ("Sea Otters" and "The Sun") and two academic texts ("Metacognition" and "Problem Solving"). We did not include the distinction of text type in our analyses, as this was not a topic of interest for the current study. The length of the texts varied between 248 and 432 words in Dutch and between 285 and 421 words in English. All

TABLE 1
Descriptive Statistics of the Participants' Language Proficiency

	Dutch (first language)	English (second language)
Lexical Test for Advanced Learners of English score	88.92 (5.41)	72.89 (9.75)
Self-ratings		
Reading	4.91 (0.33)	3.48 (0.78)
Listening	4.96 (0.19)	3.79 (0.67)
Writing	4.79 (0.47)	3.11 (0.78)
Speaking	4.98 (0.16)	3.48 (0.78)

Note. Standard deviations are in parentheses. The range for the Lexical Test for Advanced Learners of English scores is 0–100, and the range for the self-ratings is 1–5.

texts were presented in Arial, 18-point, black letters on a white background with 1.5-line spacing. Each text fitted on one screen.

The expository texts about the sun and sea otters were taken from Roediger and Karpicke (2006). Vander Beken and Brysbaert (2018) translated the English texts to Dutch and matched the different language versions on semantics, word frequency, and word prevalence (for a detailed description, see Vander Beken & Brysbaert, 2018). The academic texts were taken from peer-reviewed research articles (metacognition: Efklides, 2006; problem solving: Mayer, 1998), as these constitute typical learning materials for university students. The original English texts were translated to Dutch by a content expert and checked by a native speaker. Both language versions were compatible in style, content, structure, and length.

The four texts were divided into information units, which were the specific interest areas for which the eye movements were monitored (see Table 2 for unit characteristics). Roediger and Karpicke (2006) divided the texts about the sun and sea otters into 30 propositions or ideas. In a few cases, we further divided propositions that contained several units of factual information (e.g., a name and an event). Similar to the expository texts, the academic texts were also divided into information units. Take the following sentence from "The Sun" as an example: "About 5 billion years from now, the core of the Sun will shrink and become hotter." This sentence contains three information units: (1) "About 5 billion years from now," (2) "the core of the Sun will shrink," and (3) "and [the core will] become hotter."

To obtain information centrality ratings, the units were rated on a 5-point Likert-type scale. The instruction for this task was the following: "Indicate how important you think the content of every unit is in this text by circling the corresponding number from 1 (totally unimportant) up to 5 (very important)." The units were rated by 10 to 12 experts (PhDs or PhD students in marine biology and astronomy or educational sciences) and six to 10 undergraduate students (psychology or education sciences). The average of all ratings was taken per unit.

Tests

Ten true/false statements were created for each text, five of which related to the most central units and five to the most peripheral units. We already had this information for "The Sun" and "Sea Otters" (Vander Beken & Brysbaert, 2018). For "Metacognition" and "Problem Solving," De Bruyne, Aesaert, and Valcke (2017) developed knowledge mastery tests to measure domain-specific content knowledge. The true/false statements were created out of almost literal sentences from the knowledge tests. Two example statements are "Sea otters keep warm by means of their double-layered fur" from the "Sea Otters" text, and "Metacognitive experiences are based on the outcome of monitoring task-processing features" from the "Metacognition" text. The reliability coefficient ω_t for the four tests ranges between .61 and .75.

Participants had to complete the tests on the open-source survey software tool LimeSurvey (version 2.05; https://www.limesurvey.org/). Students made true/false judgments after the following instruction: "Are the following statements true according to the text? Answer with yes (true) or no (false)."

Additional Questionnaires

The participants completed a questionnaire concerning their perception of the text and their reading motivation in Dutch and English. All of these single-item questions involved ratings on a 7-point Likert-type scale, on which only the extreme values were labeled, ranging from "not at all" (1) to "very much" (7). For text perception, participants were asked the following questions about each of the texts: how interesting they found them (text interest), how difficult they found the content of the text (content difficulty), and to which degree they were already familiar with the content (prior knowledge). For reading motivation, participants were asked for each language how much they like to read (reading motivation), their personal judgment of their reading capability in that language (reading self-efficacy), and how important they thought it was to be able to understand texts in that language (perceived reading importance).

TABLE 2
Descriptive Statistics of the Texts, Averaged Over Information Units per Language and Information Centrality Type

	Dutch (first language)			English (second language)		
Information centrality type	Unit length[a]	Average word length[b]	Average word frequency[c]	Unit length[a]	Average word length[b]	Average word frequency[c]
Central	11.55 (3.54)	6.09 (1.78)	5.30 (0.59)	11.86 (4.48)	5.23 (1.63)	5.51 (0.54)
Peripheral	9.9 (4.56)	5.48 (1.00)	5.38 (0.57)	9.68 (4.22)	5.09 (0.79)	5.37 (0.76)

Note. Standard deviations are in parentheses.
[a]The number of words in the information unit. [b]The average word length of the words in the unit. [c]The average Zipf SUBTLEX frequency of the words in the unit: SUBTLEX-NL for Dutch words (Keuleers, Brysbaert, & New, 2010) and SUBTLEX-UK for English words (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

Finally, participants completed the Dutch and English versions of the LexTALE (a language proficiency test; Lemhöfer & Broersma, 2012), as well as self-ratings of their L1 and L2 proficiency on a 5-point Likert-type scale (see Table 1). The LexTALEs were programmed in C with Tscope (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006), and the additional questionnaires were presented on LimeSurvey. The reliability coefficient ω_t for the LexTALEs was .61 for the Dutch version and .78 for the English version. The low reliability of the Dutch LexTALE is due to a ceiling effect. The test is meant for advanced L2 speakers and intended to be of comparable difficulty as the English LexTALE. Indeed, we see that the scores on the Dutch LexTALE as L1 are considerably higher (M = 89) than those on the English LexTALE as L2 (M = 73). We report no reliabilities for the other measure, as these are all based on single items.

Apparatus

The eye movements were monitored with an EyeLink 1000+ (desktop mount version; SR Research). Only the dominant eye was recorded at a sampling rate of 1000 Hz, and the spatial accuracy lies between 0.25 and 0.5 degrees of visual angle. The texts were presented on a 24-inch screen (1920 × 1080). The participants were seated in a comfortable chair at approximately 95 cm from the screen. During the reading or studying of the texts, their head rested on a chin- and headrest mounted to the table. The questionnaires were presented on a Dell Latitude E5550 laptop with a 15.6-inch monitor.

Procedure

The participants were welcomed by the experimenter and received oral instructions about the experiment. The experimenter was a faculty staff member who did not have a teaching assignment in the relevant participant pool. Participants in the studying condition were asked to study the text to be optimally prepared to complete the true/false test afterward, as they would do for a regular university exam. Participants in the informational reading condition were unaware about the subsequent tests and were asked to read the texts like they would encounter them in a magazine or on a website. Participants were tested individually, and the entire experiment took place in the same room.

After these initial instructions, participants sat down on a chair in front of the eye tracker. Before the presentation of the first text, a 9-point calibration was performed.[3] Participants were then given a maximum of 10 minutes to read or study the first text. The text was removed from the screen after the 10 minutes expired, or participants could press the space bar if they finished earlier. Between texts, they were allowed to take a small break. Before the presentation of each new text, a new 9-point calibration procedure was run. There was always a switch of language between texts: If the first text was an L2 text, the following would be an L1 text, and so on. Text and language order were counterbalanced across participants, so each text could be the first, second, third, or fourth text, and it could be presented either in L1 or in L2. The maximum interval of 10 minutes remained the same for all texts.

After studying or reading the four texts, participants had to complete the four comprehension tests with true/false questionnaires. Participants in the informational reading condition were only informed about the tests at this point in the experiment. The order and language of the tests were the same as the presentation of the texts in the studying/reading phase.

After completing the tests, the participants were first presented with the additional questionnaires and then the Dutch and English LexTALEs. The order of the LexTALEs was determined by the language of the first text read: If it was an L1 (Dutch) text, the Dutch LexTALE was completed first. Finally, the participants were presented with the proficiency self-ratings. The entire session lasted approximately 60 minutes for reading and 90 minutes for studying.

Data Analysis

All data analysis was performed in R (version 3.4.1; R Core Team, 2017). For the generalized linear mixed-effects models, we used lme4 (1.1-13) and lmerTest (for computation of p-values; 2.0-33) packages. Effect sizes (Cohen's d) were calculated using Westfall, Kenny, and Judd's (2014) procedure for linear mixed-effects models. For the interpretation of this effect size, it is important to know that it is much lower than d values based on condition means (as happens in analyses of variance) because Westfall et al.'s effect size is calculated relative to the variability between participants and the variability between items. Brysbaert and Stevens (2018) reported that a d value of 0.09 calculated with Westfall et al.'s equation turned into an effect size of 0.30 when calculated in an analysis of variance over participants on condition means based on 40 items and to 0.80 when the condition means were based on 350 items.

Eye Movements

To assess the effects of language, reading goal, and information centrality, we made a distinction between five eye movement measures. Two of these relate to shallow text processing: first-pass reading time (the time it took to read an information unit for the first time), which involves the time needed to correctly identify the words and familiarize oneself with the content of the text, and saccadic amplitude (the size of forward eye movements that readers make, progressing through the text). The other three variables relate to deep structure processing. The total reading time (the summed duration of all passages) and the fixation count (the total number of fixations made on all passages) are informative for text comprehension difficulty (what is difficult needs to be fixated more and longer). Finally, the regression count (the number of times readers go back to a section of the text) shows how easily parts of the text are integrated: The need to look back to a word or part of the text is higher when it is not very clear how it fits in the rest of the text. A regression can also be made when a word is misidentified and needs revisits for a correct identification.

All eye movements were analyzed with linear mixed models. These are regression analyses that are run across participants and stimuli and allow researchers to generalize to both other participants and other stimuli. Only data of the 10 units in each text with the highest and lowest information centrality were included in the analysis. (An additional analysis including all information units yielded no statistically significant differences with the results presented here.) The dependent variables were the five variables discussed previously. The fixed-effects structure consisted of our factorial design: Language (Dutch/L1 or English/L2) × Reading Goal (studying or reading) × Information Centrality (central or peripheral). In addition, we included the covariates' unit length (the number of words in the unit), average word frequency (the average Zipf word frequencies of the unit; Dutch frequencies from SUBTLEX-NL: Keuleers, Brysbaert, & New, 2010; English frequencies from SUBTLEX-UK: van Heuven, Mandera, Keuleers, & Brysbaert, 2014), average word length (the average number of letters of the words in the unit), and L1 and L2 proficiency of the participants (Dutch and English LexTALE scores). As mentioned earlier, we also included the three-way interactions amon language, reading goal, and the covariates. All continuous predictors were centered to reduce correlations between main effects and interactions. Using linear mixed-effects models allowed us to model both participant and information unit as random intercepts to (a) account for genetic, developmental, or social differences between participants and (b) be able to generalize to other information units and texts, as the current ones are not an exhaustive list of the space of possible words and sentences (Baayen, Davidson, & Bates, 2008). As each participant read texts in multiple languages, and each information unit could be presented in two reading goals, language was fitted as a random slope for participants and reading goal as a random slope for unit.

Accuracy Scores

Because of the dichotomous nature of the true/false statements (they are answered either correctly or incorrectly), we analyzed the accuracy scores with a generalized linear mixed-effects model. The dependent variable was the score for each true/false statement (1 = *correct*, 0 = *incorrect*). The factorial design was again included in the fixed effects (Language × Reading Goal × Information Centrality), as well as covariates for reading motivation (self-efficacy, perceived reading importance, and reading motivation), text perception (prior knowledge, text interest, and content difficulty), and L1–L2 proficiency. Random intercepts for participant and question were included, and the same random slopes were fitted as in the eye movement analysis.

To test whether the accuracy scores could be predicted by reading times, we calculated the correlations between the accuracy score and the timed eye movement measures (first-pass time and total reading time). We only formally investigated this relation (by running statistical models with the timed measures included as the predictor) for statistically significant correlations.

Results

The data of four participants (three students in the studying condition and one student who read informatively) could not be included in the final data set due to recording issues (mainly caused by head movements during the recording).

The tables with all the final model outcomes are presented in the Appendix. We conducted pairwise comparisons (Tukey's multiple comparison tests) if there was a statistically significant interaction between two or all of the main predictors (language, reading goal, and information centrality). In the case of a statistically significant three-way interaction of a control variable with language and reading goal, we performed a separate analysis for each language and for each reading goal. For each of the four resulting models, we ran contrasts if the relevant interaction was statistically significant, to identify its specific pattern. Statistically nonsignificant main effects or interactions including the control variables are not reported.[4]

First, however, we briefly discuss the findings related to the questionnaires to verify whether the two between-group conditions are comparable.

Questionnaires

We first ran a between-group analysis of the variables of the questionnaires to check whether we could assume that the groups were equal (see Table 3). We analyzed the Dutch and English LexTALEs with two-sample *t*-tests and the text perception and motivation measures with Wilcoxon signed-rank tests (with continuity correction). A Dunn–Šidák correction for multiple testing was applied to determine statistically significant differences ($\alpha = .00465$).

There were no statistically significant differences between groups except for text interest: The studying

TABLE 3
Between-Group Comparison of the Reading Goal Groups on Proficiency, Text Perception, and Motivation Measures

	Reading group	Studying group	Test statistic	*p*
Dutch Lexical Test for Advanced Learners of English score (max = 100)	88.72 (5.54)	89.49 (5.08)	*t* = 0.633	.529
English Lexical Test for Advanced Learners of English score (max = 100)	71.86 (9.73)	74.51 (9.66)	*t* = 1.193	.237
Text interest (max = 7)	3.81 (1.00)	4.50 (0.91)	*W* = 1,027	.002
Content difficulty (max = 7)	3.99 (0.70)	3.94 (0.92)	*W* = 695	.787
Prior knowledge (max = 7)	2.03 (0.73)	1.81 (0.49)	*W* = 860	.151
L1 reading motivation (max = 7)	5.43 (1.61)	5.87 (1.30)	*W* = 624	.292
L2 reading motivation (max = 7)	4.92 (1.52)	5.14 (1.64)	*W* = 636	.363
L1 reading self-efficacy (max = 7)	5.68 (1.16)	6.11 (0.88)	*W* = 560	.074
L2 reading self-efficacy (max = 7)	4.49 (1.24)	4.67 (1.36)	*W* = 645	.413
L1 perceived reading importance (max = 7)	6.65 (0.59)	6.87 (0.52)	*W* = 566	.017
L2 perceived reading importance (max = 7)	6.41 (0.80)	6.51 (0.91)	*W* = 637	.308

Note. L1 = first language; L2 = second language. Standard deviations are in parentheses. The significance level is α = .00465.

group rated the texts as more interesting in comparison with the reading group. It could be that the studying group found the texts more interesting because they analyzed the content more thoroughly. Furthermore, and interestingly, reading motivation was higher for L1 ($M = 5.66$, $SD = 1.47$) than for L2 ($M = 5.03$, $SD = 1.57$), which is in accordance with the findings of Vander Beken and Brysbaert (2018) and Vander Beken et al. (2018; Wilcoxon test: $V = 1{,}352$, $p < .001$).

Overall Processing Times

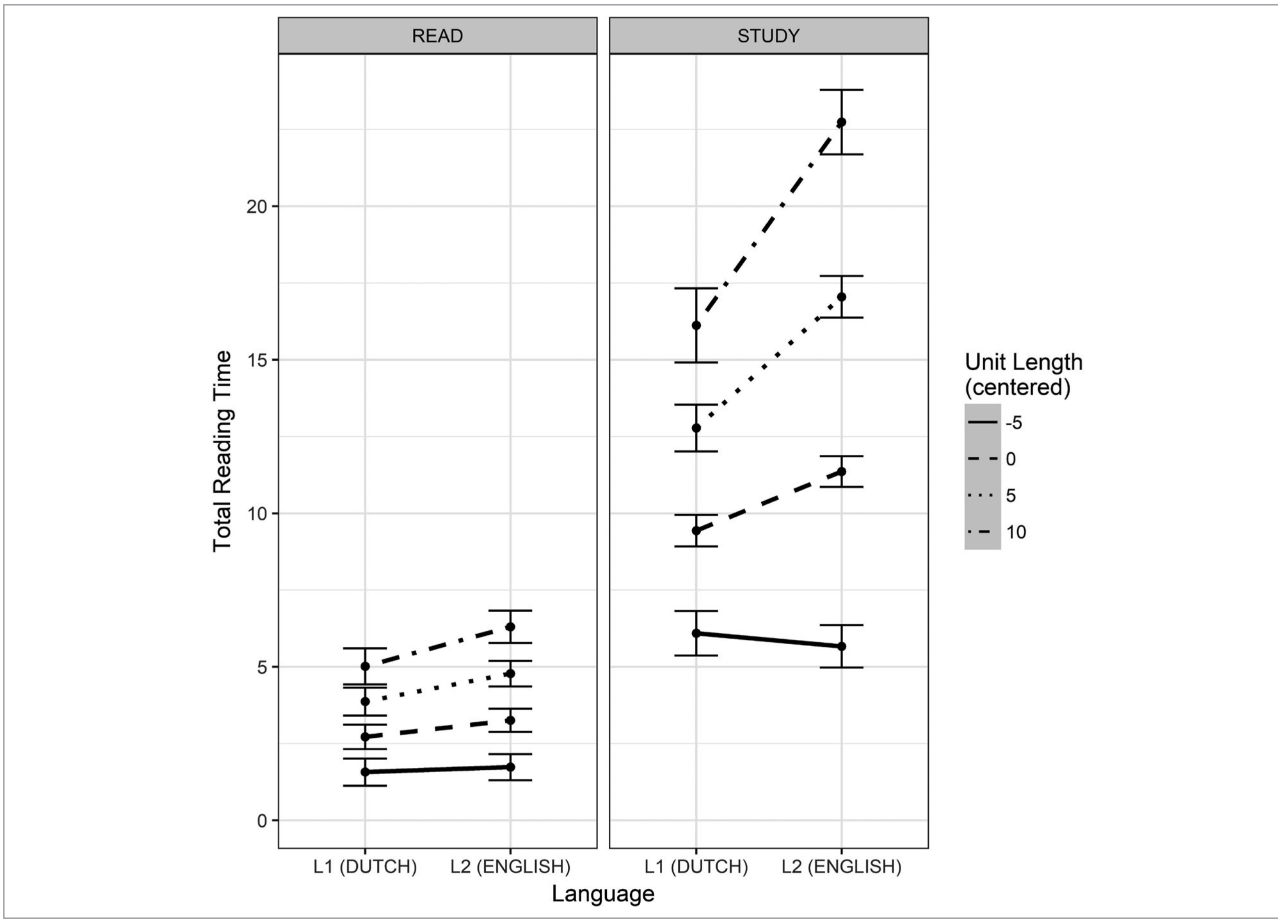
On average, participants spent 1 minute 58 seconds on reading the texts and 7 minutes 20 seconds on studying them. Times were shorter in L1 than in L2: 1 minute 52 seconds versus 2 minutes 4 seconds for reading and 7 minutes 9 seconds versus 7 minutes 31 seconds for studying. This translates into processing speeds of 189 words per minute (WPM) for reading in L1, 174 WPM for reading in L2, 54 WPM for studying in L1, and 50 WPM for studying in L2. So, participants took on average more than 3 times longer to study the texts than to read them.

As indicated earlier, we limited the more detailed eye movement analyses to the 10 target information units: the five most central and the five most peripheral. The results of the dependent variables are presented in light of the four research questions. The model outcomes for the eye movement measures are presented in Table A1 and for the test scores in Table A2. Note that for regression count, reading goal could not be added as a random slope because this resulted in convergence errors.

What Are the Differences Between the Eye Movement Patterns in L1 and L2 Studying?

Results relevant for this research question are interactions between language and reading goal. The three-way interaction among language, reading goal, and unit length was statistically significant for total reading time ($\beta = 0.395$, standard error [SE] = 0.143, $t = 2.757$, $p < .01$, $d = 0.080$; see Figure 1). In the separate analyses per reading goal, the interaction between language and unit length was statistically significant for studying ($\beta = 0.472$, $SE = 0.141$, $t = 3.333$, $p < .01$, $d = 0.073$). Post hoc contrasts revealed that total reading times were shorter for L1 than L2 if the unit contained at least 10 words ($\chi = 5.20$, degrees of freedom [df] = 1, $p < .05$). The three-way interaction among language, reading goal, and unit length was again statistically significant for fixation count ($\beta = 1.470$, $SE = 0.527$, $t = 2.788$, $p < .01$, $d = 0.078$; see Figure 2). In the separate analyses for reading goal, the interaction between language and unit length was statistically significant for studying ($\beta = 1.734$, $SE = 0.517$, $t = 3.352$, $p < .01$, $d = 0.070$). Post hoc contrasts revealed that the fixation count was lower for L1 than L2 if the unit contained at least 10 words ($\chi = 5.40$, $df = 1$, $p < .05$). Finally, the three-way interaction among language, reading goal, and L2 proficiency was statistically significant for saccadic amplitude ($\beta = 0.013$, $SE = 0.005$, $t = 2.448$, $p < .05$, $d = 0.006$). In the separate analysis for studying, the interaction between language and L2 proficiency was statistically significant ($\beta = 0.013$, $SE = 0.003$, $t = 4.119$, $p < .001$, $d = 0.006$). Post hoc contrasts revealed

FIGURE 1
The Interaction Among Language (*x*-Axis), Reading Goal (Panels), and Unit Length (Lines) for Total Reading Time (*y*-Axis, in Seconds)



Note. L1 = first language; L2 = second language. Error bars represent standard errors.

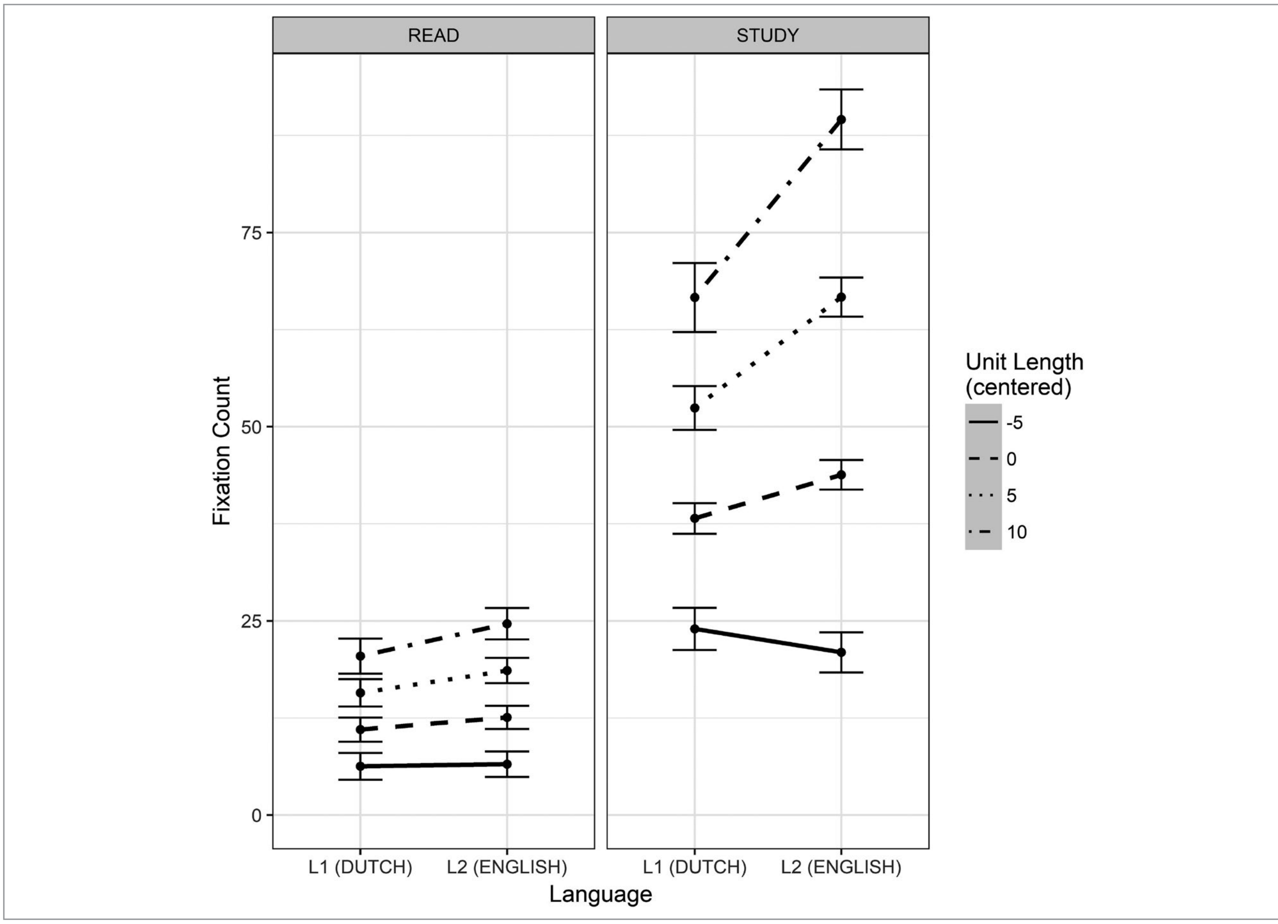
that the saccadic amplitude was smaller for L2 compared with L1 if the LexTALE score was lower than 80.16 ($\chi = 3.85$, $df = 1$, $p < .05$), which corresponds to an L2 proficiency effect on L2 studying but not L1 studying.

Is the Effect of Reading Goal on the Eye Movement Pattern Similar in L1 and L2?

To answer this research question, we report the results of interactions with reading goal, which indicate differences in processing between the studying and informational reading conditions. For the first-pass time, there was a statistically significant three-way interaction among language, reading goal, and unit length ($\beta = 0.055$, $SE = 0.024$, $t = 2.286$, $p < .05$, $d = 0.043$; see Figure 3). In the L1 analysis, there was a statistically significant interaction between reading goal and unit length ($\beta = -0.052$, $SE = 0.020$, $t = -2.605$, $p < .05$, $d = 0.042$): post hoc contrasts showed that the first-pass time was shorter when studying compared with reading if the unit contained at least 10 words ($\chi = 4.63$, $df = 1$, $p < .05$).

In the total reading times, reading goal interacted with average word length ($\beta = 1.336$, $SE = 0.388$, $t = 3.441$, $p < .001$, $d = 0.271$). Post hoc contrasts showed that total reading times were shorter for reading than studying when the average word length exceeded 3.2 letters ($\chi = 10.01$, $df = 1$, $p < .01$; the difference between the reading goals became larger with increasing average word length). Furthermore, the three-way interaction among language, reading goal, and unit length was statistically significant ($\beta = 0.395$, $SE = 0.143$, $t = 2.757$, $p < .01$, $d = 0.080$; see Figure 1). The interaction between reading goal and unit length was statistically significant in both the L1 ($\beta = 0.438$, $SE = 0.062$, $t = 7.033$, $p < .001$, $d = 0.093$) and L2 analyses ($\beta = 0.835$, $SE = 0.052$, $t = 16.189$, $p < .001$, $d = 0.184$). Post hoc contrasts showed that the total reading time was shorter for informational reading compared with studying if the unit contained at least two words in L1 ($\chi = 13.84$, $df = 1$, $p < .001$) or three words in L2 ($\chi = 7.90$, $df = 1$, $p < .01$).

FIGURE 2
The Interaction Among Language (*x*-Axis), Reading Goal (Panels), and Unit Length (Lines) for Fixation Count (*y*-Axis)



Note. L1 = first language; L2 = second language. Error bars represent standard errors.

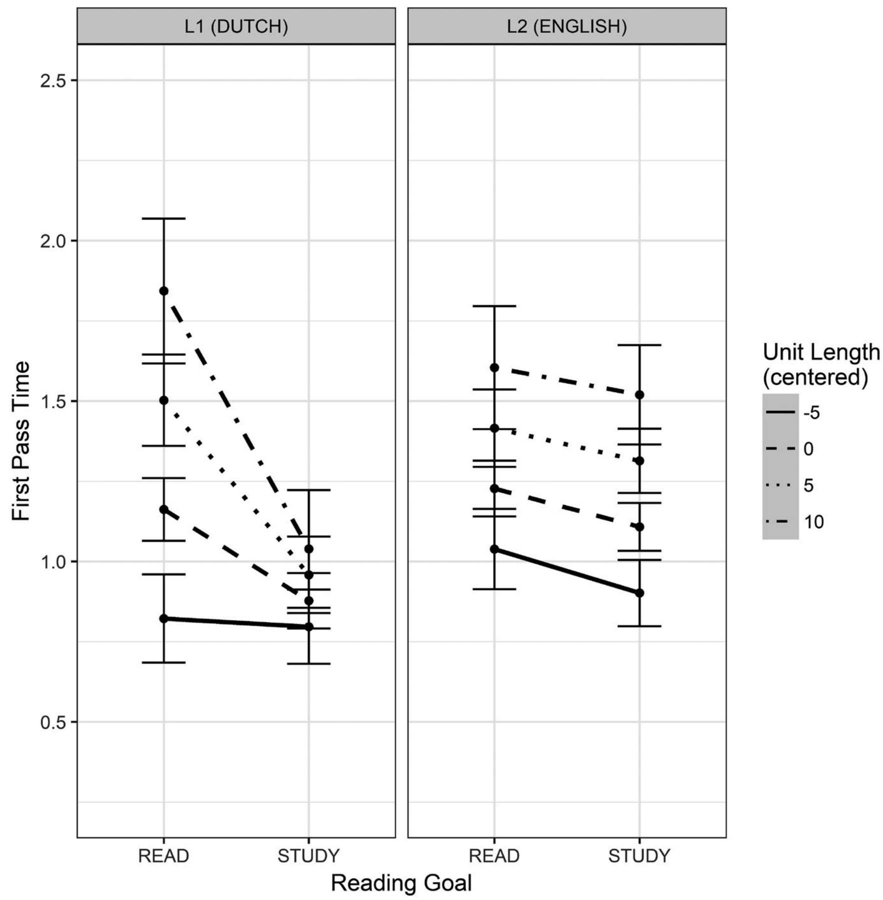
The difference between reading and studying became larger with increasing unit length.

For the fixation count, there was a similar pattern as in the total reading times. Reading goal interacted with average word length (β = 5.312, *SE* = 1.432, *t* = 3.710, *p* < .001, *d* = 0.282). Post hoc contrasts showed that the fixation count was higher for reading than studying when the average word length exceeded 3.2 letters (χ = 11.53, *df* = 1, *p* < .001; the magnitude of this difference increased when the average word length increased). There was again a statistically significant three-way interaction among language, reading goal, and unit length (β = 1.470, *SE* = 0.527, *t* = 2.788, *p* < .01, *d* = 0.078; see Figure 2). The interaction between reading goal and unit length was statistically significant in both the L1 (β = 1.902, *SE* = 0.234, *t* = 8.141, *p* < .001, *d* = 0.106) and L2 analyses (β = 3.382, *SE* = 0.189, *t* = 17.931, *p* < .001, *d* = 0.199). Post hoc contrasts showed that the more fixations were made while reading compared with studying if the unit contained at least two words in L1 (χ = 11.78, *df* = 1, *p* < .001) and three words in L2 (χ = 5.50, *df* = 1, *p* < .05). The difference between reading and studying became larger with increasing unit length.

The interaction between reading goal and unit length was statistically significant for regression count (β = 0.234, *SE* = 0.037, *t* = 6.354, *p* < .001, *d* = 0.101). Post hoc contrasts showed that the regression count was higher for studying than reading when the unit length was at least two words (χ = 4.79, *df* = 1, *p* < .05); this effect became larger with increasing unit length. Furthermore, the three-way interaction among language, reading goal, and average word length was statistically significant (β = −0.469, *SE* = 0.187, *t* = −2.501, *p* < .05, *d* = 0.203; see Figure 4). In the separate analyses for L1, the interaction between reading goal and average word length was statistically significant (β = 0.805, *SE* = 0.142, *t* = 5.653, *p* < .001, *d* = 0.265). Post hoc contrasts revealed that the regression count was lower for reading than studying if the unit had an average word length of 4.252 letters or more (χ = 31.52, *df* = 1, *p* < .001); the difference became larger with increasing average word length.

Finally, there was a statistically significant three-way interaction among language, reading goal, and average word frequency (β = −1.346, *SE* = 0.441, *t* = −3.053, *p* < .01, *d* = 0.582; see Figure 5). In the separate analyses

FIGURE 3
The Interaction Among Reading Goal (*x*-Axis), Language (Panels), and Unit Length (Lines) for First-Pass Time (*y*-Axis, in Seconds)



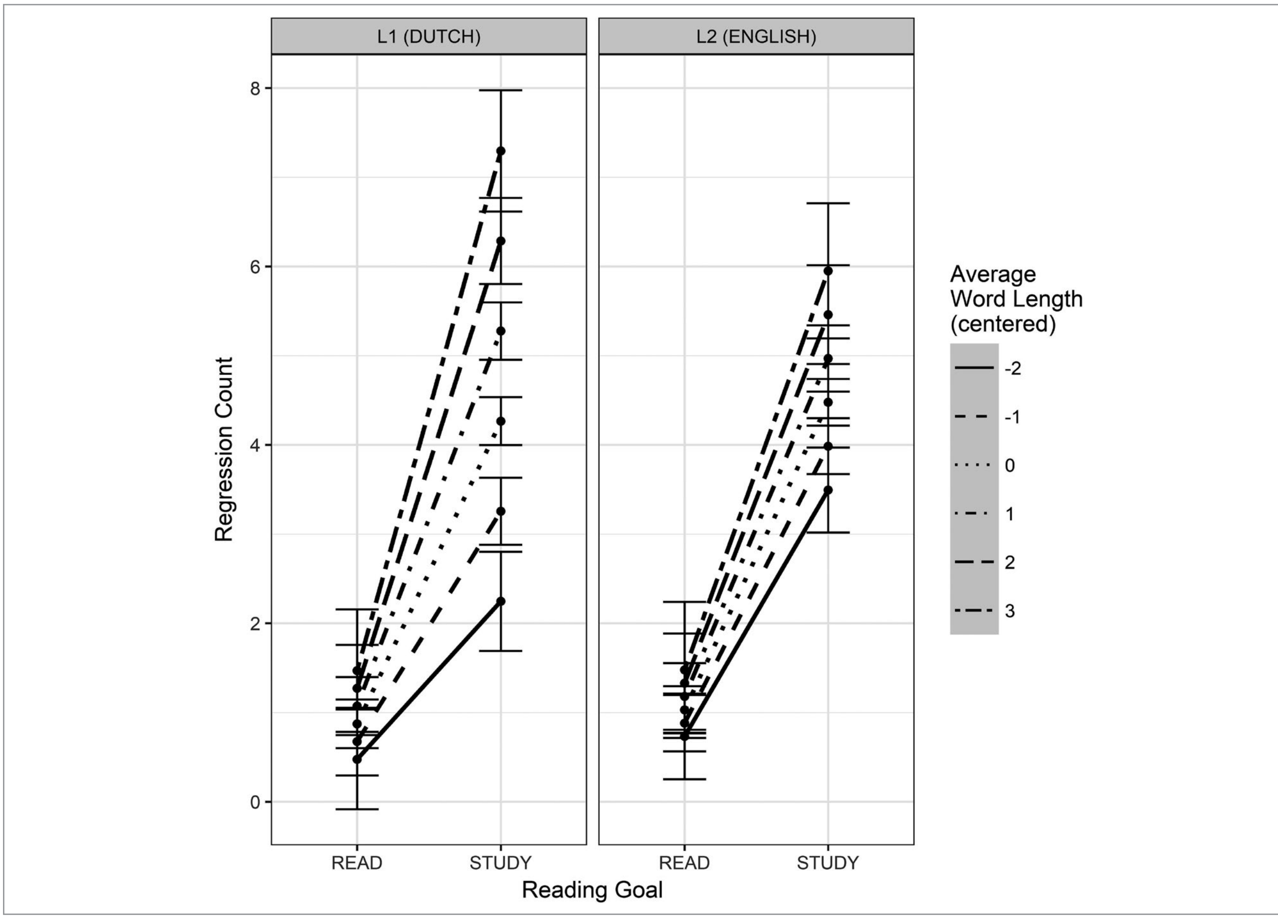
Note. L1 = first language; L2 = second language. Error bars represent standard errors.

for L1, the interaction between reading goal and average word frequency was statistically significant (β = 1.087, *SE* = 0.359, *t* = 3.032, $p < .01$, *d* = 0.358). Post hoc contrasts revealed that the regression count was lower for reading than studying if the average Zipf word frequency of the unit was 4.639 or more (χ = 6.72, *df* = 1, $p < .01$). The difference between reading and studying became larger with increasing word frequency. For the saccadic amplitude, reading goal interacted significantly with average word length (β = 0.092, *SE* = 0.041, *t* = 2.259, $p < .05$, *d* = 0.045) and average word frequency (β = 0.220, *SE* = 0.108, *t* = 2.033, $p < .05$, *d* = 0.107). Post hoc contrasts showed that the saccadic amplitude was smaller for reading than studying when the average word length of the unit was at least 5.125 letters (χ = 3.85, *df* = 1, $p < .05$) or when the average Zipf word frequency was at least 5.575 (χ = 3.84, *df* = 1, $p < .05$).

To specifically address the question of whether the longer study times in L2 than in L1 were limited to the initial reading, we divided the total reading times into passages. A passage was defined as the reading of an information unit after an incoming forward saccade or after a regression. All participants had at least three passages for the critical information units. Figure 6 shows the results of the first three passages.

We clearly see that the longer reading times for L2 are not limited to the first passage, but they are still very present for the third passage. Post hoc paired sample *t*-tests showed a statistically significant longer first passage time for L2 than L1 (L1: *M* = 0.886 second, *SD* = 0.254; L2: *M* = 1.042, *SD* = 0.301; *t*(38) = 4.184, $p < .001$, *d* = 0.560), which was similar for the second passage time (L1: *M* = 0.842, *SD* = 0.181; L2: *M* = 0.983, *SD* = 0.261; *t*(38) = 4.671, $p < .001$, *d* = 0.628) and for the third passage time (L1: *M* = 0.758, *SD* = 0.186; L2: *M* = 0.888, *SD* = 0.225; *t*(38) = 4.748, $p < .001$, *d* = 0.630). Although there is some decrease in the difference between the first passage and the third, L2 readers do not catch up after a slower first passage. Even when we limited the analysis to the last passage of each participant, there was still a statistically

FIGURE 4
The Interaction Among Reading Goal (*x*-Axis), Language (Panels), and Average Word Length (Lines) for Regression Count (*y*-Axis)



Note. L1 = first language; L2 = second language. Error bars represent standard errors.

significant difference between languages (L1: $M = 0.573$, $SD = 0.137$; L2: $M = 0.688$, $SD = 0.222$; $t(38) = 4.012$, $p < .001$, $d = 0.623$).[5] The language effect thus seems present in all readings of the information units.

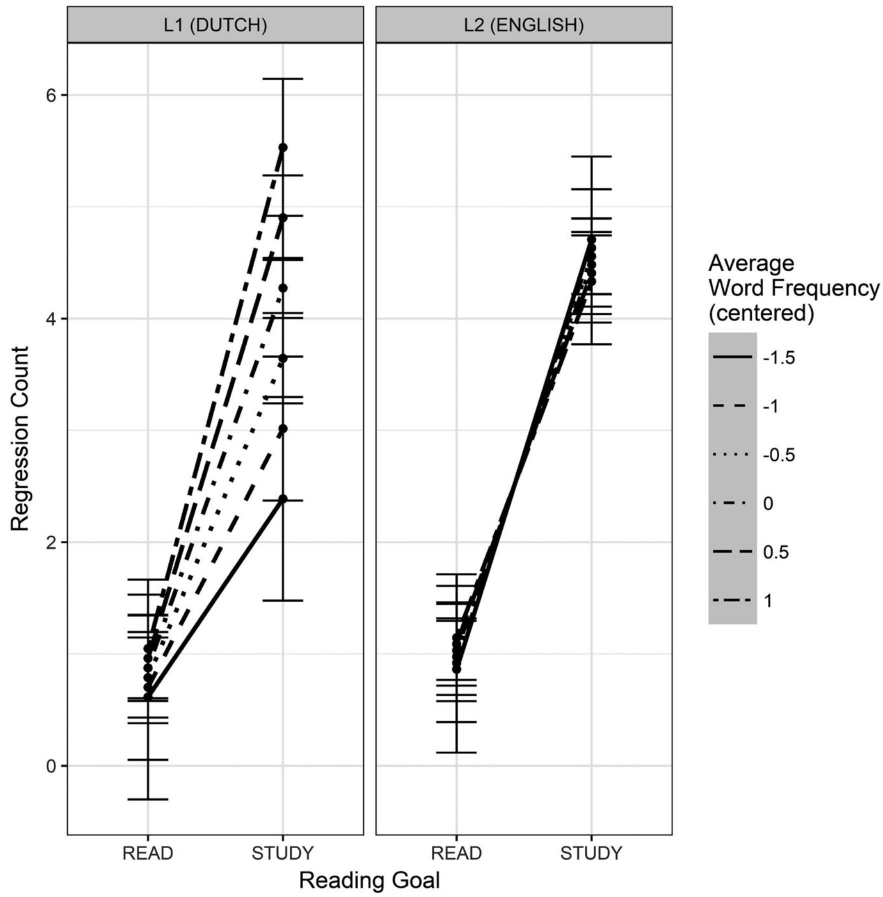
Is the Memory Trace, as Tested by a Cued Recognition Test, Affected by Language, Reading Goal, and/or Information Centrality?

For the model of the test scores, no random slopes could be fitted due to convergence errors. There was a main effect of reading goal ($\beta = 0.935$, $SE = 0.187$, $z = 5.005$, $p < .001$, $d = 1.048$; see Figure 7); accuracy scores were higher for studying than for reading. Furthermore, the main effects of content difficulty ($\beta = -0.106$, $SE = 0.038$, $z = -2.809$, $p < .01$, $d = 0.119$) and reading motivation ($\beta = 0.087$, $SE = 0.044$, $z = 2.010$, $p < .05$, $d = 0.098$) were statistically significant. Accuracy scores were lower when participants perceived the texts as more difficult or when their motivation score for reading in the target language was lower.

Are Fixation Times of an Information Unit an Adequate Predictor of Test Accuracy Scores of That Unit?

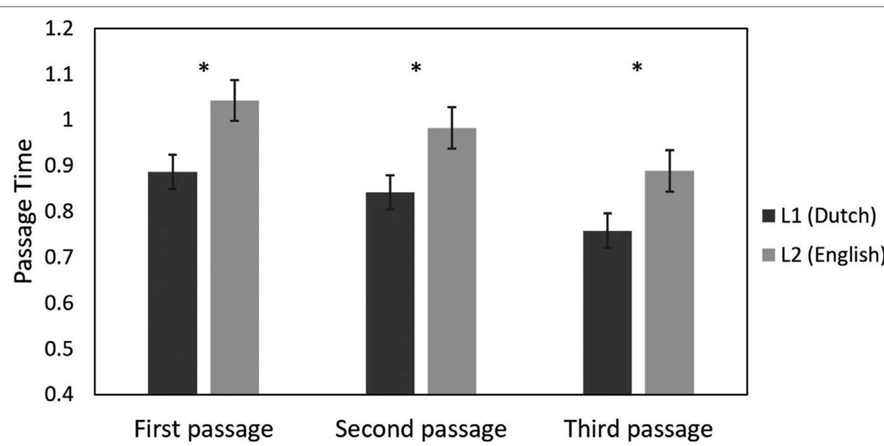
We calculated the correlations between the two timed measures (first-pass and total reading time) and accuracy scores to further investigate the relation between reading times and accuracy scores. The correlation of test accuracy with first-pass time was not statistically significant ($r < .01$, $p = .866$), whereas the correlation with total reading time was very small but statistically significant ($r = .07$, $p < .001$). This association was, however, confounded by reading goal, as none of the correlations between total reading times and accuracy scores remained statistically significant when we calculated them separately for each group (reading: $r = .02$, $p = .378$; studying: $r = -0.04$, $p = .125$). Hence, we did not further investigate the influence of reading time on test scores.

FIGURE 5
The Interaction Among Reading Goal (*x*-Axis), Language (Panels), and Average Zipf Word Frequency (Lines) for Regression Count (*y*-Axis)



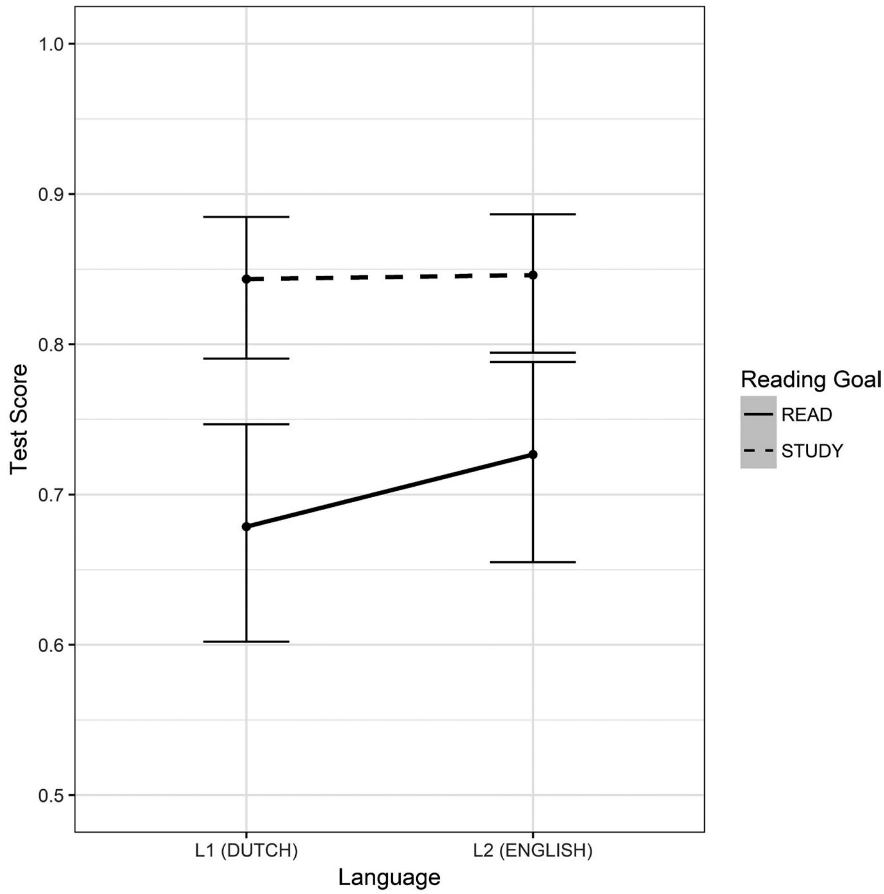
Note. L1 = first language; L2 = second language. Error bars represent standard errors.

FIGURE 6
The Interaction Between Language and Passage Number (*x*-Axis) for Passage Times (*y*-Axis, in Seconds)



Note. L1 = first language; L2 = second language. Error bars represent standard errors.

FIGURE 7
The Interaction Between Language (*x*-Axis) and Reading Goal (Lines) for Accuracy Score (*y*-Axis)



Note. L1 = first language; L2 = second language. Error bars represent standard errors.

Discussion

In the current study, we investigated the effects of two context features (reading goal and text language) on the reading process. Participants had to read (informational reading) or study (to memorize the text content as preparation for a recognition test) four texts (two in L1 and two in L2) while their eye movements were monitored. Afterward, all participants received tests with true/false judgments to examine their memory for the contents of the texts. Using this experimental design, we could thoroughly investigate the differences between L1 and L2 reading while studying texts on the one hand and the effect of different reading goals within and across languages on the other hand. To our knowledge, this is the first study that has compared the L2 reading pattern under different reading goals. The technique that we applied, eye tracking, also allowed us to gain insight the nature of the previously reported longer L2 studying time in comparison with L1 (e.g., Chen & Donin, 1997; Donin et al., 2004) and to make inferences on the underlying cognitive processes under our experimental conditions. Indeed, eye movements can, for example, be related to early and late reading processes (Boston et al., 2008; Rayner, 1998; Rayner et al., 2006).

More specifically, we set out to answer four research questions: (1) to investigate whether there are differences in the eye movement patterns for different text languages while studying; (2) to examine the effects of different reading goals on the eye movements within and across each language; (3) to determine whether the memory trace for the text content is affected by reading goal, language, or information centrality; and (4) whether it is possible to predict accuracy scores of the test based on reading times. The results are highly relevant to English as a Medium of Instruction in higher education settings, as our participant sample, materials, and recognition tests are representative to such a setting and can be related to models of purposeful reading, such as RESOLV (Britt et al., 2018; Rouet et al., 2017), as we investigate how several context features affect the reading process. We discuss the results in light of the research questions.

The Influence of Language on Eye Movements While Studying

We found that total reading times were about 20% longer and that 15% more fixations were made when studying in L2, analogous to the L2 cost observed in previous eye-tracking studies of reading (e.g., Cop, Drieghe, & Duyck, 2015). This effect interacted with unit length: If the number of words increased, the additional time needed to study the unit increased more in L2 than in L1. Interestingly, in the sentence-reading analysis in Cop, Drieghe, and Duyck's (2015) study, a similar interaction between language and sentence length in reading times was reported. It seems that especially the processing of longer sentences, which are syntactically more complex, results in an additional difficulty for L2 studying. As Cop, Drieghe, and Duyck reported, the effect of sentence complexity on reading times seems to be characteristic for the reading pattern in a less proficient language, similar to that of children reading in their L1 (Rayner, 1986; Warren, White, & Reichle, 2009).

We were particularly interested in whether the longer period to memorize the content in L2 was caused by the first reading of the text or would be distributed throughout the learning. Figure 7 clearly shows that the latter is the case: Participants still needed more time to read an information unit in L2 than in L1 when the unit was revisited for the third time (or even more), which is surprising. This shows that the processing cost imposed by reading in L2 continues throughout the reading process.

There was no effect of language on regression counts, suggesting that the reading process for memorization is not more error prone in L2. This agrees with the observation that the reading rate in L2 is slower overall, rather than less accurate. The saccade length was influenced by L2 proficiency but only in L2 studying: The average saccadic amplitude was larger for participants with a higher L2 proficiency level. This suggests that they can process more information (characters or words) from a single fixation, resulting in larger jumps through the text.

We also expected an information centrality effect on the first-pass time in accordance with the results of Yeari et al.'s (2015) study, but the interaction between reading goal and information centrality was not statistically significant. In fact, there was no significant effect of information centrality on any measure (which is subsequently discussed in more detail).

To summarize, text language is a contextual feature that affects reading for memorization throughout the entire study time, not only during initial encoding. Especially for complex text, the reading process is compromised, as students seem to have more difficulties with processing longer sentences in L2, as shown in the corresponding measures reflecting deeper processing (total reading time and fixation count). This is relevant for English as a Medium of Instruction in higher education, as academic textbooks often contain complex sentences. The L2–L1 differences in eye movement patterns closely resemble those of the previous sentence-reading studies, showing more and longer fixations and smaller saccades in L2.

The Influence of Reading Goal on Eye Movements

In Yeari et al.'s (2015) study, an effect of reading goal as a contextual feature was found on the eye movement patterns in informational reading of L1 texts. For our second research question, we attempted to replicate this L1 effect and wanted to investigate whether the influence of reading goal would be similar, or rather more pronounced, in L2. The many statistically significant differences in dependent variables between the studying and reading conditions confirm that our manipulation to induce distinct reading goals was successful. Nevertheless, in terms of effect size, the majority of the effects can be categorized as small, even when we take into account that effect sizes in linear mixed-effects models are smaller than those in analyses of variance based on condition means (Brysbaert & Stevens, 2018).

The first-pass time was shorter when participants studied the text contents than when they read out of interest, showing a difference in the initial, shallow processing of the text, which was performed more quickly in the studying condition. A higher order interaction revealed that this effect was mainly driven by L1; it was not statistically significant in L2. Interestingly, Yeari et al. (2015) also reported a shorter first-pass time when participants read the texts in preparation for a closed-ended question test compared with informational reading. The most likely explanation of this difference is that readers studying a text first browse through it to get familiar with the structure and the contents, and subsequently examine the text more thoroughly by rereading (parts of) the text. Interestingly, the shorter first-pass time in studying than in reading was not present in L2. In L2, participants either adopted another strategy or simply were not able to go faster through the text because of their lower processing capabilities. This pattern of results shows that context features such as language and reading goal not only have main effects on the reading process but can also interact.

In the total reading time and fixation count, there was a large difference between reading and studying, showing a more comprehensive deeper processing of the text in the latter condition: Participants spent about 70% longer and made about 70% more fixations on the information units when they had to memorize the text content compared with informational reading. This reading goal effect was larger for L2 than L1 and was modulated by unit length: An increase in the number of words in the unit led to a larger increase in total reading times and number of fixations in L2 than in L1.

More regressions were made while studying the texts than reading them. This is consistent with the finding that longer reading times for studying than reading only emerged on total reading times, not on first-pass times. It informs us that participants who memorize information do so by reading the text multiple times or look back in the text often, instead of spending longer on the first-pass reading of information units. Reading goal interacted with word length and frequency (in L1 only) and unit length (for both languages). A higher number of regressions were made toward longer, more complex units, which indicates that integration of these units in the surrounding content was more difficult. This effect was larger for studying than informational reading. Furthermore, in L1 studying, the increase in regressions toward units containing longer and more frequent words was higher than in L1 reading. Whitford and Titone (2012) also reported higher regression rates toward higher frequent content words. As these words were also skipped more often, the researchers hypothesized that this could result in a higher need to revisit them.

In general, there does not seem to be a large language difference in the number of regressions. Yeari et al. (2015) pointed out that regressions are also made for a failure of or difficulties with sentence comprehension (Frazier & Rayner, 1982; Vauras, Hyönä, & Niemi, 1992). This suggests that after the initial longer processing time, comprehension is not more erroneous or more difficult in L2 compared with L1. (Note that there was no statistically significant difference either on the subjective ratings of text difficulty between L1 [M = 3.79, SD = 1.18] and L2 [M = 4.16, SD = 1.18], Wilcoxon test: V = 1,114, $p > .05$.)

The saccadic amplitude was smaller for L2 than L1 (cf. Cop, Drieghe, & Duyck, 2015). The low-level L2 reading processes seem more demanding, so a smaller amount of information (i.e., fewer words or letters) can be processed simultaneously in L2 than in L1, resulting in a higher need to make fixations and smaller saccades. It might also be that readers are more familiar with the nature of their (written) mother tongue and are better at estimating how to split the text in manageable chunks, resulting in a more efficient saccade pattern.

The saccadic amplitude was larger for studying than reading informatively. The most likely reason for this is that readers make larger jumps while rereading parts of the text at a later stage of studying because they are already familiar with the material and primarily want to better organize their mental model of the text. To test this possibility, we ran an additional analysis of the saccades in the first passage versus later passages as an extra factor. This provided evidence in line with the hypothesis: In the first passage, there was no difference in saccadic amplitude between reading and studying, but there was a statistically significant difference on later passages (β = 0.146, SE = 0.054, t = 2.679, $p < .01$). So, participants at first proceed through the text similarly when reading or studying, but because studying involves going through the same text repeatedly, further repetitions of the text involve larger jumps.

As mentioned earlier, we found no effects of information centrality on any of the measures. We have three possible explanations to account for these null effects in comparison with Yeari et al.'s (2015) study. First, we applied a different operationalization of information units in comparison with Yeari et al.'s study: Whereas they included the main predicates and accompanying arguments, we only included the smallest possible piece of unique information. For example, one of their central information units was "Mount Vesuvius is a volcano located between the ancient Italian cities of Pompeii and Herculaneum" (p. 1093). According to our definition of a unit, this sentence consists of two units: "Mount Vesuvius is a volcano" and "located between the ancient Italian cities of Pompeii and Herculaneum." It could be that readers rather pay more attention to whole sentences, which they judge to contain central information, instead of just looking for the most important parts within the sentence. This could have concealed information centrality effects in our study.

Second, we applied a different approach in the rating of information centrality. Yeari et al. (2015) had three expert judges rate the centrality of each information unit, based on two criteria: the importance of this piece of information for the overall understanding of the text and whether the text understanding would be impaired without this piece of information. In our procedure, we combined the judgments of experts and our participant population and did not explicitly ask whether the understanding would be impaired if the unit was missing. Still, because we had a larger number of raters, our ratings ought to result in a more reliable estimation of what readers experience as important units. Finally, the difference of information centrality rating between our central and peripheral units was smaller than Yeari et al.'s: In the current study and in Yeari et al.'s, respectively, M = 4.12, SD = 0.26 versus M = 4.7, SD = 0.3 for high centrality, and M = 2.69, SD = 0.40 versus M = 1.6, SD = 0.4 for low centrality. It could be that our difference was too small to result in statistically significant differences on the eye movement measures.

In summary, differences between central and peripheral information units did not appear, but we found clear effects of reading goal as a contextual factor in various eye movement measures, which indicate qualitative differences in approaching a text for informational reading or studying purposes (cf. Yeari et al., 2015). This is in line with predictions of the RESOLV model of reading (e.g., Rouet et al., 2017), which states that the reading process is determined, among others, by the context model that readers construct prior to reading. We also found that interactions between contextual features have an impact on reading outcomes, as shown in the eye movement patterns. In L1, the first-pass time was shorter for studying than for informational reading, showing a less thorough shallow processing of the text, but the total reading time was longer, and in total, more fixations were made. This indicates that students first went through the texts they needed to study rather quickly, after which they began studying parts of the texts again. In regard to the text language, we expected a similar yet somewhat inflated pattern for L2 studying versus informational reading. L2 processing was indeed more effortful for studying than reading, and this difference was larger than in L1 processing. Unit length had a statistically significant impact on this language difference, as more complex sentences required more additional effort to study in L2 than in L1. This shows that especially the memorization of long, difficult texts, such as in higher education handbooks, has a large impact on the reading process while studying texts in L2, more than just informational reading in L2.

Memory for Texts

Our third goal was to examine whether memory for texts was affected by reading goal, language, or information centrality. All participants received 40 true/false statements on the content of information units (the language was congruent with the texts they received). As expected, participants whose goal was to memorize text content had a higher score than those who read them informatively (this result had a large effect size). This shows that when more time is spent on a text, a richer mental model (cf. the landscape model; van den Broek et al., 1999) of its content can be built, improving the retention and resulting in higher test accuracy. There was again no effect of information centrality and no language effect either: Test performance was equal in L1 and L2. This suggests that with a cued recognition procedure, memory traces of studied text facts are equally accessible in both languages. This is in line with the idea of a common underlying proficiency (Cummins, 1979), or the notion of (partly) language-independent knowledge representations underlying surface features. Cummins (1979) proposed that the knowledge and skills a bilingual learns in one language strengthen the common underlying proficiency and can be accessed through or transferred to the other language. In other words, it should not matter in which language propositions are memorized, as they are maintained through the common underlying proficiency, which benefits from the knowledge and proficiency in both languages. In the current context, this seems to be the case with the true/false judgments. As test achievement was equal for L1 and L2, we could argue that the storage and retrieval of information units (propositions) in common underlying proficiency is evenly efficient in L1 and L2.

How do we reconcile this with the poorer performance on a text recall task, as reported by Vander Beken and Brysbaert (2018)? The most likely explanation is related to the richness and structure of the developed mental landscape (van den Broek et al., 1999). When individual information units are probed by means of strong memory cues, they are equally available whether they were learned in L1 or in L2. However, when the relevant task is a free recall and participants learned the materials in L1 rather than L2, more associated concepts are co-activated during encoding in L1 than in L2, and these concepts form extra memory links between the various propositions, resulting in a better recall performance in L1.

These results are in line with those of Chen and Donin (1997), who reported longer study times for L2 but an equal performance on L1 and L2 tests. The results are also a confirmation of the findings of Vander Beken and Brysbaert (2018) and Vander Beken et al. (2018), who found no differences between L1 and L2 on yes/no recognition questions. For recognition tests, reading/studying time seems to be the essential component. When participants can study in L2 as long as they think is necessary, their performance is equal to L1.

How is it possible that studying takes longer in L2 but that this compensatory strategy results in similar test outcomes in both languages? There are two possible answers to this question. First, similar to real-life studying, we did not impose a narrow time limit, so sufficient time could be taken to compensate for L2-processing costs. When allowed by the context, students adapt their reading process: They take the necessary time and are good at guessing how much additional time is needed to reach a similar studying goal in L2 compared with L1.

The second explanation can be found in the levels-of-processing framework (Craik & Lockhart, 1972). This framework assumes that the initial stage of encoding is aimed at surface form (e.g., extracting and memorizing information), whereas higher order processing occurs in the later stages for deeper encoding (e.g., making inferences or links to other information units in the text and/or prior knowledge). Recognition tasks such as true/false statements benefit from the knowledge of details rather than the ability to see the bigger picture (for a discussion, see Vander Beken et al., 2018). Indeed, the levels-of-processing effect seems to be smaller in L2 compared with L1 (Francis & Gutiérrez, 2012): There is an L2 advantage for shallow processing tasks, but this decreases for tasks that require deeper understanding. This can also be stated the other way around: L1 benefits more from deeper processing tasks than L2. In the eye movement data, we found more and longer fixations in L2 and smaller saccades, which might indicate that a lot of attention is directed toward the lexical level for a correct identification of the words (resulting in sufficiently strong memory traces for individual words and sentences), but which leaves the L2 reader with less time to integrate information in the mental model of the text as well as in L1. This fits with results of studies on L1 and L2 text comprehension, which showed that L2 readers allocate more resources to low-level processing (cf. longer and more fixations as indications of hampered word identification) and in turn show reduced higher order processing of the text, as shown in reduced inference making (Horiba, 1996; Pérez et al., 2018), for example. It also relates to the resource hypothesis, which states that L2 processing is more demanding for the working memory (Sandoval et al., 2010), resulting in a smaller capacity for higher order processing.

The Relation Between Reading Time and Accuracy Score

Finally, we were interested in whether the quality of the memory trace improves by fixating the information unit longer. As none of the correlations between the timed measures and accuracy scores were statistically significant, we concur with Yeari et al. (2015) that "attention allocation is only a minor factor in determining the memory strength of textual ideas" (p. 1088). Apparently, memory for academic studying is determined by factors other than simple encoding.

To some extent, this is a surprising finding because, in the very same experiment, we see that students are able to improve their scores by reading longer when they know a test will follow. We also see that they spontaneously take more time to read/study in L2, so their performance is comparable to that in L1. So, it is not that students lack the metacognitive skills to predict/monitor their performance. Still, there is no correlation between reading time and memory accuracy for individual information units. Note that some models of eye movement control decouple attention and motor processes, so attention is not centered at the fixation point but could be directed toward words or objects in the peripheral view (e.g., Reichle, Pollatsek, & Rayner, 2006).

Limitations of the Current Study and Guidelines for Future Research

In the current study, we were limited by the specific text and test types that we used. Although the texts were representative of the expository and academic texts used in higher education, they were especially limited in length. Students often have to read multipage academic articles or entire book chapters. To increase the ecological validity of this type of research, a recommendation is to work with longer texts.

Regarding the reading goal, we only focused on one particular form of studying and related test type, assessing reading behavior for a specific test type but in depth. Our participants had to memorize the information in the texts and were, for example, not required to make inferences between (different parts of the) texts. In naturalistic settings, many test options other than the cued recognition test that we used exist. This could influence the similarity or difference between L1 and L2 performance. In higher education, some examples are cued recall, free recall, inferential multiple-choice questions that combine information from multiple parts of a text, essay writing, and so forth. Furthermore, as the specific test type was known to the participants, the eye movement patterns were also bound to this specific test, as the participants specifically prepared for the true/false statements in the studying condition. It would thus also be interesting in further eye-tracking studies to compare L1 and L2 reading patterns for other types of reading goals.

Although the eye-tracking methodology has many strengths and resembles daily real-life reading, there is one obvious disadvantage: Participants had to study a text on a screen while their head rested on the chin- and head-rest. One way to estimate the impact of this restriction is to have a parallel group doing the same experiment but without eye tracking. What we should find is that the pattern of behavioral measures (total reading/studying time and test scores) are the same.

In future research, attention could also be given to the subjective assessment of reading strategies by the participants. We were able to make some inferences based on the eye movement data, but it would be interesting to see whether this matches with the reading strategies that participants report having adopted during reading or studying.

Regarding language, although we tested bilingual participants, they read each text and solved each accompanying test in the same language. In an English as a Medium of Instruction setting, it also occurs that students have English handbooks for an L1 course, which requires them at some point to translate this material and which goes beyond a strict unilingual setting. A recognition or recall procedure involving a language switch between studying and test taking would be an ecologically valid addition to research on English as a Medium of Instruction. Furthermore, a design with a free recall test, crossing the text and test language, could also shed light on the issue of the involvement of encoding versus (re)production in the L2 recall cost: If a L2 cost is present in both L2 encoding conditions, this indicates that the encoding stage is involved in the cost. If the cost is only found in the L2 text—L2 test condition (but not L2 text—L1 test), the cost can be situated at the (re)production stage. There is also the possibility that the disadvantage of L2 studying is situated in both processes, which would result in recall costs in both L2 encoding conditions and in the L1 encoding—L2 test condition. A final recommendation for future research is a further investigation of slower L2 processing as a function of unit length. It would be interesting to check whether the L2 processing speed can be increased if the same content is presented in shorter, easier to process sentences.

Conclusion

We showed that participants need about 20% more time to study texts in L2 compared with L1. Then, the information units in the text are equally well retained, resulting in similar scores for L1 and L2 recognition tests. The investigation of various eye movement measures has provided valuable insights into the similarities, differences, and strategies applied when two important contextual features are varied: reading goal (informational reading and studying) and text language (L1 and L2). When studying in L1, students seem to quickly scan the text in the first passage, which does not seem to occur in L2 studying. The many interactions of language with unit length indicate that especially complex sentences are difficult to study in L2. This is highly relevant for English as a Medium of Instruction in higher education settings with nonnative English bilinguals, where complex English academic texts are an important part of study material.

NOTES

This research was supported by the Concerted Research Action (BOF13/GOA/032) and a grant (BOF.PDO.2018.0010.01), both awarded by the Special Research Fund of Ghent University.

[1]We are aware that this is a narrow definition of studying. Indeed, studying is a multifaceted concept involving a broad range of learning activities, such as gaining new insights, in-depth analyzing, and inference making, in addition to fact memorization. The weights of the various components are likely to differ depending on the study goal (e.g., preparing to write an essay or answer multiple-choice questions). Fact memorization is only one facet of studying, but it was a good starting point for us to answer some of the research questions of the current study, by focusing on this one aspect through a well-controlled experimental design with detailed online (eye-tracking) measures. Research on text studying differs from research on text comprehension, because in studies on text comprehension, participants could be allowed to consult the text while answering the questions. We were interested in studying because that is what our students have to do to pass tests or exams.

[2]Vander Beken and Brysbaert (2018) did not report longer L2 study times, as their participants had a fixed interval of seven minutes to study the texts. However, it is not clear whether participants required the complete time interval for studying in L1, as this was not explicitly measured.

[3]This is a standard eye-tracking procedure, in which a dot is presented in the middle and at all four corners and sides of the screen. The participant needs to fixate these nine points, covering the entire screen, so the EyeLink software can infer the position of the eye when a certain location on the screen is fixated.

[4]As the full models for the dependent variables contained a considerable number of statistically nonsignificant interactions and main effects, we checked whether the pattern of results remained similar after a backfitting procedure of the nonsignificant terms. This did not result in any statistically significant changes from the result presented here.

[5]All p-values of the paired sample t-tests were Bonferroni corrected for multiple comparisons. Note that the last passage can differ a lot between information units and participants, as the fifth or 15th passage, for example, can be last.

REFERENCES

Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading*, *33*(3), 263–283. https://doi.org/10.1111/j.1467-9817.2009.01413.x

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Boston, M.F., Hale, J.T., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1–12. https://doi.org/10.16910/jemr.2.1.1

Britt, M.A., Rouet, J.-F., & Durik, A.M. (2018). *Literacy beyond text comprehension*. New York, NY: Routledge.

Britton, B. (1994). Understanding expository text. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). San Diego, CA: Academic.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1* (1), article 9. https://doi.org/10.5334/joc.10

Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *Modern Language Journal*, *81*(2), 209–227. https://doi.org/10.1111/j.1540-4781.1997.tb01176.x

Coiro, J. (2011). Predicting reading comprehension on the internet. *Journal of Literacy Research*, *43*(4), 352–392. https://doi.org/10.1177/1086296X11421979

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS One*, *10*(8), e0134008. https://doi.org/10.1371/journal.pone.0134008

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Dafouz, E., & Camacho-Miñano, M.M. (2016). Exploring the impact of English-medium instruction on university student academic achievement: The case of accounting. *English for Specific Purposes*, *44*, 57–67. https://doi.org/10.1016/j.esp.2016.06.001

De Bruyne, E., Aesaert, K., & Valcke, M. (2017). The development of knowledge tests when studying academic texts in higher education. In L. Gómez Chova, A. López Martínez & I. Candel Torres (Eds.), *EduLearn17 conference proceedings* (pp. 10494–10507). Valencia, Spain: IATED Academy.

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

Dijkstra, T., & van Heuven, W.J.B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Doiz, A., Lasagabaster, D., & Sierra, J. (2013). Globalisation, internationalisation, multilingualism and linguistic strains in higher education. *Studies in Higher Education*, *38*(9), 1407–1421. https://doi.org/10.1080/03075079.2011.642349

Donin, J., Graves, B., & Goyette, E. (2004). Second language text comprehension: Processing within a multilayered system. *Canadian Modern Language Review*, *61*(1), 53–77. https://doi.org/10.3138/cmlr.61.1.53

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology*, *16*(1/2), 79–103. https://doi.org/10.1080/09541440340000141

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*(1), 78–103. https://doi.org/10.1598/RRQ.38.1.4

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

Francis, W.S., & Gutiérrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory & Cognition*, *40*(3), 496–503. https://doi.org/10.3758/s13421-011-0163-3

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Gablasova, D. (2014). Learning and retaining specialized vocabulary from textbook reading: Comparison of learning outcomes through L1 and L2. *Modern Language Journal*, *98*(4), 976–991. https://doi.org/10.1111/modl.12150

Goldman, S.R., & Varnhagen, C.K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language*, *25*(4), 401–418. https://doi.org/10.1016/0749-596X(86)90034-3

Gollan, T.H., Fennema-Notestine, C., Montoya, R.I., & Jernigan, T.L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*(2). https://doi.org/10.1017/S1355617707070038

Gollan, T.H., Montoya, R.I., Cera, C., & Sandoval, T.C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Grosjean, F. (2008). *Studying bilinguals*. Oxford, UK: Oxford University Press.

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, *18*(4), 433–473. https://doi.org/10.1017/S0272263100015370

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, *73*(3), 259–280. https://doi.org/10.1016/0001-6918(90)90026-C

Joh, J. (2006). What happens when L2 readers recall? *Language Research*, *42*(1), 205–238.

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

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Kingston, N.M. (2008). Comparability of computer- and paper-administered multiple-choice tests for K–12 populations: A synthesis. *Applied Measurement in Education*, *22*(1), 22–37. https://doi.org/10.1080/08957340802558326

Kintsch, W. (1974). *The representation of meaning in memory*. New York, NY: Erlbaum.

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G., & Keenan, J.M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 196–214. https://doi.org/10.1016/S0022-5371(75)80065-X

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Liversedge, S.P., & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

Mayer, R.E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*(1/2), 49–63. https://doi.org/10.1023/A:1003088013286

Miller, A.C., & Keenan, J.M. (2009). How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia*, *59*(2), 99–113. https://doi.org/10.1007/s11881-009-0025-x

Noyes, J.M., & Garland, K.J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

Pérez, A., Hansen, L., & Bajo, T. (2018). The nature of first and second language processing: The role of cognitive control and L2 proficiency during text-level comprehension. *Bilingualism: Language and Cognition*. Advance online publication. https://doi.org/10.1017/s1366728918000846

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236. https://doi.org/10.1016/0022-0965(86)90037-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Chace, K.H., Slattery, T.J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. https://doi.org/10.1207/s1532799xssr1003_3

Rayner, K., Slattery, T.J., Drieghe, D., & Liversedge, S.P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528. https://doi.org/10.1037/a0020990

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reichle, E.D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4–22. https://doi.org/10.1016/j.cogsys.2005.07.002

Roediger, H.L., & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rouet, J.-F., Britt, M.A., & Durik, A.M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and*

Instruction, *52*, 69–79. https://doi.org/10.1016/j.learninstruc.2017.04.007
Sandoval, T.C., Gollan, T.H., Ferreira, V.S., & Salmon, D.P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514
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
Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods*, *38*(2), 280–286. https://doi.org/10.3758/BF03192779
van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of memory representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.
Vander Beken, H., & Brysbaert, M. (2018). Studying texts in a second language: The importance of test type. *Bilingualism: Language and Cognition*, *21*(5), 1062–1074. https://doi.org/10.1017/S1366728917000189
Vander Beken, H., Woumans, E., & Brysbaert, M. (2018). Studying texts in a second language: No disadvantage in long-term recognition memory. *Bilingualism: Language and Cognition*, *21*(4), 826–838. https://doi.org/10.1017/S1366728917000360
van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521
Vauras, M., Hyönä, J., & Niemi, P. (1992). Comprehending coherent and incoherent texts: Evidence from eye movement patterns and recall performance. *Journal of Research in Reading*, *15*(1), 39–54. https://doi.org/10.1111/j.1467-9817.1992.tb00020.x
Warren, T., White, S.J., & Reichle, E.D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, *111*(1), 132–137. https://doi.org/10.1016/j.cognition.2008.12.011
Westfall, J., Kenny, D.A., & Judd, C.M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014
Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. https://doi.org/10.3758/s13423-011-0179-5
Whitford, V., & Titone, D. (2019). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition*, *22*(1), 58–77. https://doi.org/10.1017/S1366728917000554
Yeari, M., Oudega, M., & van den Broek, P. (2017). The effect of highlighting on processing and memory of central and peripheral text information: Evidence from eye movements. *Journal of Research in Reading*, *40*(4), 365–383. https://doi.org/10.1111/1467-9817.12072
Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*(8), 1071–1097. https://doi.org/10.1007/s11145-015-9561-4
Zambarbieri, D., & Carniglia, E. (2012). Eye movement analysis of reading from computer displays, eReaders and printed books. *Ophthalmic and Physiological Optics*, *32*(5), 390–396. https://doi.org/10.1111/j.1475-1313.2012.00930.x

Submitted May 22, 2018
Final revision received July 24, 2019
Accepted July 24, 2019

NICOLAS DIRIX (corresponding author) is a postdoctoral researcher in the Department of Experimental Psychology at Ghent University, Belgium; email nicolas.dirix@ugent.be. His research interests are bilingual reading and studying and the use of eye movements in language research.

HELEEN VANDER BEKEN is currently employed at Artevelde University College, Ghent, Belgium; email heleen.vanderbeken@arteveldehs.be. She obtained her PhD from the Department of Experimental Psychology at Ghent University, Belgium, when this study was conducted. Her research interests are language, memory, and study orientation.

ELLEN DE BRUYNE is a doctoral candidate in the Department of Educational Studies at Ghent University, Belgium; email eidbruyn.debruyne@ugent.be. Her research interests are English as a medium of instruction and supporting reading in a second-language higher education setting.

MARC BRYSBAERT is a professor in the Department of Experimental Psychology at Ghent University, Belgium; email marc.brysbaert@ugent.be. His research interests are word recognition, word frequency measures, and the use of big data in language research.

WOUTER DUYCK is a professor in the Department of Experimental Psychology at Ghent University, Belgium; email wouter.duyck@ugent.be. His research interests are bilingual processing, language development, and orientation and selection tests for tertiary education.

APPENDIX

The following tables present the outcomes for the models of all dependent variables. For the dichotomous factors language, reading goal, and information centrality, the same factor level was set as the reference level for each analysis. For language, the reference level is L1; for reading goal, it is informational reading; and for information centrality, it is central information. We included the remaining level with the factor name in the tables.

TABLE A1
Estimates, Standard Errors (*SE*s), *t*-Values, *p*-Values, and Effect Sizes for the Fixed and Random Effects of the Linear Mixed-Effects Models for the First-Pass Time, Total Reading Time, Fixation Count, Regression Count, and Saccadic Amplitude

	First-pass duration					Total reading time					Fixation count					Regression count					Saccadic amplitude				
Fixed effect	β	*SE*	*t*	*p*	Cohen's *d*	β	*SE*	*t*	*p*	Cohen's *d*	β	*SE*	*t*	*p*	Cohen's *d*	β	*SE*	*t*	*p*	Cohen's *d*	β	*SE*	*t*	*p*	Cohen's *d*
Intercept	1.188	0.120	9.904	<.001		2.802	0.420	6.668	<.001		11.367	1.642	6.923	<.001		0.835	0.349	2.395	.018		2.050	0.089	23.117	<.001	
Language: L2	−0.004	0.146	−0.026	.979	0.003	0.464	0.341	1.359	.176	0.094	1.273	1.343	0.948	.345	0.068	0.127	0.447	0.284	.777	0.055	−0.267	0.065	−4.077	<.001	0.130
Reading goal: Studying	−0.331	0.130	−2.541	.012	0.256	6.765	0.736	9.190	<.001	1.371	26.990	2.819	9.574	<.001	1.432	3.270	0.290	11.277	<.001	1.413	0.307	0.120	2.554	.012	0.150
Information centrality: Peripheral	−0.053	0.141	−0.374	.710	0.041	−0.179	0.299	−0.598	.550	0.036	−0.792	1.105	−0.716	.474	0.042	0.074	0.439	0.169	.866	0.032	0.021	0.066	0.318	.751	0.010
Average word frequency	−0.239	0.183	−1.304	.197	0.185	−0.306	0.390	−0.784	.433	0.062	−0.644	1.447	−0.445	.657	0.034	0.173	0.572	0.303	.762	0.075	0.158	0.086	1.843	.068	0.077
Average word length	−0.024	0.072	−0.327	.745	0.019	0.338	0.155	2.174	.030	0.069	1.496	0.577	2.594	.010	0.079	0.199	0.224	0.887	.377	0.086	0.053	0.032	1.632	.106	0.026
Unit length	0.068	0.020	3.396	.001	0.053	0.230	0.043	5.355	<.001	0.047	0.946	0.158	5.970	<.001	0.050	0.055	0.062	0.876	.383	0.024	0.007	0.010	0.761	.448	0.003
L1 proficiency	0.007	0.015	0.456	.650	0.005	−0.003	0.072	−0.038	.970	0.001	0.014	0.284	0.050	.961	0.001	−0.005	0.034	−0.147	.884	0.002	0.013	0.015	0.905	.368	0.006
L2 proficiency	−0.002	0.008	−0.247	.806	0.002	−0.002	0.041	−0.043	.966	<0.001	−0.038	0.162	−0.237	.813	0.002	0.003	0.020	0.157	.875	0.001	0.020	0.008	2.320	.023	0.010
Language: L2 × Reading Goal: Studying	0.231	0.143	1.617	.108	0.179	1.638	0.778	2.105	.038	0.332	5.154	2.943	1.751	.083.	0.273	−0.012	0.308	−0.041	.968	0.005	−0.006	0.084	−0.073	.942	0.003
Language: L2 × Information Centrality: Peripheral	0.145	0.201	0.724	.471	0.112	0.154	0.429	0.360	.719	0.031	0.647	1.582	0.409	.683	0.034	0.062	0.625	0.100	.921	0.027	−0.095	0.093	−1.017	.312	0.046
Language: L2 × Average Word Frequency	0.342	0.240	1.426	.158	0.264	0.047	0.516	0.091	.927	0.010	0.040	1.900	0.021	.983	0.002	−0.061	0.747	−0.081	.935	0.026	−0.098	0.121	−0.808	.421	0.048
Language: L2 × Average Word Length	0.001	0.102	0.004	.997	0.001	0.047	0.219	0.213	.831	0.010	0.203	0.808	0.251	.802	0.011	−0.050	0.317	−0.157	.876	0.022	0.012	0.050	0.233	.816	0.006
Language: L2 × Unit Length	−0.030	0.027	−1.144	.257	0.023	0.075	0.057	1.315	.189	0.015	0.260	0.210	1.236	.217	0.014	0.025	0.083	0.306	.761	0.011	−0.006	0.013	−0.448	.655	0.003
Language: L2 × L1 Proficiency	−0.012	0.014	−0.811	.420	0.009	−0.002	0.050	−0.045	.964	<0.001	0.012	0.203	0.059	.953	0.001	0.011	0.033	0.342	.733	0.005	−0.002	0.008	−0.263	.793	0.001

(continued)

TABLE A1
Estimates, Standard Errors (SEs), t-Values, p-Values, and Effect Sizes for the Fixed and Random Effects of the Linear Mixed-Effects Models for the First-Pass Time, Total Reading Time, Fixation Count, Regression Count, and Saccadic Amplitude (*continued*)

	First-pass duration					Total reading time					Fixation count					Regression count					Saccadic amplitude				
Fixed effect	ß	*SE*	*t*	*p*	Cohen's *d*	ß	*SE*	*t*	*p*	Cohen's *d*	ß	*SE*	*t*	*p*	Cohen's *d*	ß	*SE*	*t*	*p*	Cohen's *d*	ß	*SE*	*t*	*p*	Cohen's *d*
Language: L2 × L2 Proficiency	<0.001	0.008	0.074	.941	0.001	0.003	0.028	0.119	.905	0.001	0.017	0.115	0.152	.880	0.001	0.003	0.019	0.153	.879	0.001	<0.001	0.004	−0.035	.972	0.000
Reading Goal: Studying × Information Centrality: Peripheral	0.096	0.127	0.756	.450	0.074	−0.109	0.757	−0.144	.886	0.022	0.423	2.790	0.152	.880	0.022	0.228	0.255	0.892	.372	0.099	0.041	0.083	0.493	.624	0.020
Reading Goal: Studying × Average Word Frequency	0.172	0.166	1.037	.301	0.133	0.402	0.986	0.408	.685	0.081	0.994	3.637	0.273	.785	0.053	1.084	0.334	3.248	.001	0.468	0.220	0.108	2.033	.045	0.107
Reading Goal: Studying × Average Word Length	0.115	0.066	1.753	.081	0.089	1.336	0.388	3.441	.001	0.271	5.312	1.432	3.710	<.001	0.282	0.811	0.132	6.122	<.001	0.350	0.092	0.041	2.259	.027	0.045
Reading Goal: Studying × Unit Length	−0.052	0.018	−2.839	.005	0.040	0.439	0.108	4.062	<.001	0.089	1.898	0.398	4.771	<.001	0.101	0.234	0.037	6.354	<.001	0.101	−0.002	0.012	−0.167	.868	0.001
Reading Goal: Studying × L1 Proficiency	−0.021	0.021	−0.966	.337	0.016	0.106	0.105	1.014	.314	0.021	0.420	0.413	1.017	.313	0.022	0.081	0.050	1.635	.107	0.035	−0.007	0.021	−0.328	.744	0.003
Reading Goal: Studying × L2 Proficiency	0.003	0.012	0.280	.780	0.002	−0.037	0.057	−0.640	.524	0.008	−0.133	0.226	−0.585	.560	0.007	0.001	0.027	−0.018	.986	0.000	−0.019	0.011	−1.631	.107	0.009
Language: L2 × Reading Goal: Studying × Information Centrality: Peripheral	−0.137	0.182	−0.754	.452	0.106	−0.534	1.079	−0.495	.622	0.108	−2.357	3.976	−0.593	.555	0.125	0.140	0.366	0.383	.701	0.060	−0.102	0.118	−0.867	.388	0.050
Language: L2 × Reading Goal: Studying × Average Word Frequency	−0.342	0.219	−1.561	.120	0.264	−1.979	1.293	−1.531	.130	0.401	−7.233	4.762	−1.519	.133	0.384	−1.346	0.441	−3.053	.002	0.582	−0.138	0.153	−0.907	.367	0.067
Language: L2 × Reading Goal: Studying × Average Word Length	−0.097	0.093	−1.038	.300	0.075	−0.748	0.548	−1.364	.177	0.152	−3.195	2.018	−1.584	.118	0.169	−0.469	0.187	−2.501	.012	0.203	−0.036	0.063	−0.577	.565	0.018

(continued)

TABLE A1
Estimates, Standard Errors (SEs), t-Values, p-Values, and Effect Sizes for the Fixed and Random Effects of the Linear Mixed-Effects Models for the First-Pass Time, Total Reading Time, Fixation Count, Regression Count, and Saccadic Amplitude (*continued*)

	First-pass duration					Total reading time					Fixation count					Regression count					Saccadic amplitude				
Fixed effect	B	*SE*	*t*	*p*	Cohen's *d*	B	*SE*	*t*	*p*	Cohen's *d*	B	*SE*	*t*	*p*	Cohen's *d*	B	*SE*	*t*	*p*	Cohen's *d*	B	*SE*	*t*	*p*	Cohen's *d*
Language: L2 × Reading Goal: Studying × Unit Length	0.055	0.024	2.286	.023	0.043	0.395	0.143	2.757	.007	0.080	1.470	0.527	2.788	.007	0.078	0.046	0.049	0.944	.345	0.020	−0.003	0.016	−0.185	.854	0.001
Language: L2 × Reading Goal: Studying × L1 Proficiency	0.021	0.021	0.982	.329	0.016	0.119	0.072	1.652	.103	0.024	0.197	0.296	0.664	.509	0.010	−0.003	0.048	−0.053	.958	0.001	0.002	0.009	0.196	.845	0.001
Language: L2 × Reading Goal: Studying × L2 Proficiency	−0.007	0.011	−0.604	.548	0.005	−0.029	0.040	−0.729	.468	0.006	0.005	0.162	0.030	.976	<0.001	0.012	0.026	0.452	.653	0.005	0.013	0.005	2.448	.016	0.006

Random effect	First-pass duration Variance	First-pass duration Standard deviation	Total reading time Variance	Total reading time Standard deviation	Fixation count Variance	Fixation count Standard deviation	Regression count Variance	Regression count Standard deviation	Saccadic amplitude Variance	Saccadic amplitude Standard deviation
Unit										
• Intercept	0.111	0.333	0.029	0.169	0.324	0.569	1.454	1.206	0.014	0.118
• Reading goal	0.013	0.115	3.718	1.928	50.022	7.073	1	1	0.030	0.173
Subject										
• Intercept	0.168	0.410	4.830	2.198	76.261	8.733	0.967	0.984	0.208	0.456
• Language	0.089	0.299	1.120	1.059	23.173	4.814	0.596	0.772	0.015	0.122
Residual	1.293	1.137	14.638	3.826	205.587	14.338	5.725	2.393	3.937	1.984

Note. L1 = first language; L2 = second language.
[a]The random slope for reading goal could not be fitted for the regression count model.

TABLE A2
Estimates, Standard Errors (*SE*s), *z*-Values, *p*-Values, and Effect Sizes for the Fixed and Random Effects of the Generalized Linear Mixed-Effects Model for the Accuracy Scores

Fixed effect	ß	*SE*	*t*	*p*	Cohen's *d*
Intercept	1.029	0.559	1.843	.065	
Language: L2	0.328	0.326	1.003	.316	0.368
Reading goal: Studying	0.935	0.187	5.005	<.001	1.048
Information centrality: Peripheral	0.063	0.329	0.192	.848	0.071
Prior knowledge	−0.074	0.047	−1.577	.115	0.083
Text interest	−0.010	0.033	−0.313	.754	0.011
Content difficulty	−0.106	0.038	−2.809	.005	0.119
Reading motivation	0.087	0.044	2.010	.044	0.098
Perceived reading importance	−0.026	0.072	−0.364	.716	0.029
Reading self-efficacy	0.006	0.056	0.100	.920	0.007
L1 proficiency	0.011	0.010	1.049	.294	0.012
L2 proficiency	−0.001	0.006	−0.216	.829	0.001
Language: L2 × Reading Goal: Studying	−0.192	0.254	−0.756	.450	0.215
Language: L2 × Information Centrality: Peripheral	−0.205	0.466	−0.441	.659	0.230
Reading Goal: Studying × Information Centrality: Peripheral	−0.002	0.259	−0.009	.993	0.002
Language: L2 × Reading Goal: Studying × Information Centrality: Peripheral	−0.034	0.367	−0.092	.927	0.038

Random effect	Variance	Standard deviation
Unit intercept	0.761	0.872
Subject intercept	0.035	0.186

Note. L1 = first language; L2 = second language.