Self-regulated learning strategies in L1 and L2 reading

Marta Reyes[1] , M. Julia Morales[2] and M. Teresa Bajo[1]

[1]Department of Experimental Psychology–Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain and [2]Universidad Loyola Andalucía, Sevilla, Spain

Abstract

English as a second language (L2) has become the medium of instruction in numerous contexts even though many people may have difficulties to read and study in L2. According to the self-regulated framework, metacognitive strategies are essential to achieve successful learning, but they are resource-consuming and their use might be compromised in demanding contexts such as learning in L2. In Experiment 1, nonbalanced bilinguals read high- and low-cohesion texts in L1 and L2 and self-rated their learning using a *judgment of learning* (JOL). Then, they answered open-ended questions and responded a customized questionnaire regarding their strategies. In Experiment 2, we introduced two bilingual groups varying in L2 proficiency. Overall, participants could adjust their JOLs and detect the difficulty of the texts correctly in L1 and L2. However, results evidenced some nuances in learning strategies related to L2 proficiency. We discuss these findings within the context of the self-regulated learning.

Highlights

- Self-regulated learning appears not to be hindered by L2 processing.
- Bilingual participants monitored the to-be-studied material correctly.
- Low-proficiency bilinguals could not compensate for the perceived language difficulty.
- L2 level may impact the resources available for metacognitive processing.

When we read, many cognitive and metacognitive processes operate so that we understand what is in a text. These processes include word recognition, information updating, inferencing, integration with previous knowledge, monitoring, and control of cognitive resources (Castles et al., 2018). For expert readers, most of these processes take place rapidly, automatically, and with minimal effort in their native tongue (L1). However, one might expect these processes to be more effortful when reading in a second language (L2), especially when L2 proficiency is not at native-like levels. This is relevant since in many countries, English as an L2 is increasingly used as the medium of instruction (Byun et al., 2011; Macaro, 2018; Pessoa et al., 2014). Despite the clear benefits of L2 education in favoring transnational networking and communication in the long term (Dafouz & Camacho-Miñano, 2016; Doiz et al., 2013), many students with lower proficiency may struggle when they take courses in L2 and face the challenge of reading and studying text materials in English–L2. In this context, addressing how metacognitive and learning strategies unfold when studying in L2 is paramount. The focus of the present study is to investigate the consequences of studying texts in L2 in the interplay between monitoring and control strategies to achieve successful learning.

Previous studies have established that working in L2 is cognitively demanding and many resources might be directed to language control (e.g., Green & Abutalebi, 2013; Ma et al., 2014; Moreno et al., 2010; Soares et al., 2019). For example, within a bilingual brain, both languages are active during production or comprehension even in contexts when only one language is being used (Bialystok, 2017; Bialystok et al., 2012; Chen et al., 2017; Iniesta et al., 2021; Kroll et al., 2014). To manage the interference from one language to the other, and to select the appropriate language for the context, language control processes need to be engaged (Beatty-Martínez et al., 2020; Kroll et al., 2015; Macizo et al., 2010; Soares et al., 2019).

This might be even more so for unbalanced bilinguals whose asymmetrical language proficiency (Luk & Kroll, 2019) leads them to have weaker L2 semantic representations (Kroll & Stewart, 1994), more interference from their L1 (e.g., Meuter & Allport, 1999), and slower and less accurate L2 processing and word recognition (Dirix et al., 2020). All this suggests that, for unbalanced bilinguals, L2 processing is more challenging and therefore might engage more cognitive resources than L1 processing (Hessel & Schroeder, 2020, 2022; Pérez et al., 2018; see Adesope et al., 2010 for a review).

Research on second language acquisition suggests that the ability of L2 readers to construct the necessary inferences for forming situation models might be constrained by their proficiency in L2

reading and vocabulary (Joh & Plakans, 2017; Nassaji, 2011; Sidek & Rahim, 2015). Thus, whereas high-proficient readers are likely to effectively employ their prior knowledge to enhance their comprehension of the text, low-proficiency readers often struggle to construct accurate situation models, impeding their capacity to make inferences and to acquire the causal relations presented in the text (Hosoda, 2017). High-proficiency readers have more knowledge of academic vocabulary, while low-proficiency readers have less extensive vocabulary and less low-frequency vocabulary, which can make it difficult for them to generate inferences and comprehend the text (Silva & Otwinowska, 2019). In this way, when L2 reading skills are limited, readers are compelled to prioritize fundamental reading processes (i.e., word decoding and syntactic parsing) over inferential processing, allocating their cognitive resources accordingly (Horiba, 1996; Hosoda, 2014). This, in turn, may hinder their ability to internalize the causal relationships presented in the texts and to construct comprehensive situation models of the texts. Moreover, low-proficiency readers may not use their prior knowledge because they do not have the necessary vocabulary available (e.g., Sidek & Rahim, 2015; Silva & Otwinowska, 2019), further limiting their ability to understand and integrate the information presented. Given the number of students that usually acquire contents in their L2, understanding the possible consequences that reading and studying in an L2 might have for learning strategies and more specifically, whether learning processes are impaired when studying in L2 has remarkable importance.

Metacognitive strategies are conceived as a feeling-of-knowing state that serves a self-regulatory purpose whereby one can observe the ongoing processing, assess one's comprehension and/or learning, detect errors, and decide what strategies need to be employed to enhance the process. According to the classical model proposed by Nelson and Narens (1990), metacognitive processes include two general functions: *monitoring* and *control.* Metacognitive monitoring refers to the online supervision and assessment of the effectiveness of cognitive resources while metacognitive control refers to the management and regulation of cognitive resources.

More recent theories on metacognitive regulation propose a close association between monitoring accuracy and control effectiveness, as monitoring facilitates control. Self-regulated learning theories, in particular, suggest that individuals rely on continuous monitoring to determine the best course of action to achieve their learning goals (e.g., Dunlosky & Ariel, 2011; Koriat & Goldsmith, 1996; Metcalfe, 2009; Metcalfe & Finn, 2008; Pieger et al., 2016; for a review see Panadero, 2017). For instance, identifying the difficult parts of a text correctly leads to appropriate effort regulation and strategy selection, which, consequently, results in greater comprehension and better memory (Follmer & Sperling, 2018). Thus, from a learning perspective, metacognitive strategies and self-regulation have been linked to academic achievement (Pintrich & Zusho, 2007; Zimmerman, 2008; Zusho, 2017), as they are critical for comprehending and memorizing information (e.g., Collins et al., 1996; Fukaya, 2013; Huff & Nietfeld, 2009; Krebs & Roebers, 2012; Thiede et al., 2003).

Thus, Deekens et al. (2018) investigated the relationship between the frequency of metacognitive monitoring and the utilization of surface and deep-level strategies. Surface-level strategies usually imply investing minimal time and effort to meet the requirements (e.g., rote learning or memorizing key concepts, Cano, 2007), whereas deep-level strategies involve paying attention to the meaning, relating ideas, and integrating them with previous knowledge, to maximize understanding. Deep-level strategies are

regarded as more effective strategies for producing longer-lasting learning (Deekens et al., 2018; Lonka et al., 2004; Vermunt & Vermetten, 2004). Deekens et al. (2018) found that students who enacted more frequent learning monitoring also engaged in deep strategies more frequently than low-monitoring students, and this resulted in better performance on academic evaluations. This pattern suggests that the combination of metacognitive monitoring and deep-level learning strategies is intrinsically linked to successful academic achievement.

The interaction between monitoring, strategies used, and learning, is significantly relevant when the to-be-learned materials vary in difficulty. Previous research has proven that *judgments of learning* (JOLs), a measure of the monitoring process, are sensitive to different cues and item-based features such as font type, concreteness, and relatedness (e.g., Magreehan et al., 2016; Matvey et al., 2006; Undorf et al., 2018). According to the cue utilization approach (Koriat, 1997), learners base their JOLs on different sources of information, namely intrinsic, extrinsic, and mnemonic cues. Intrinsic cues refer to features of the material that indicate how easy or difficult it will be to learn (e.g., word frequency, associative strength, text cohesion). Extrinsic cues concern the study environment (e.g., the use of interactive imagery, time constraints, and repeated study trials). Mnemonic cues are internal states that provide information about how well an item has been learned (e.g., the subjective experience of processing an item fluently, past experiences in similar situations or beliefs).

The effects of some variables on JOLs have been extensively investigated at the word-unit level (e.g., Halamish, 2018; Hourihan et al., 2017; Hu et al., 2016; Li et al., 2021; Undorf et al., 2017, 2018; Undorf & Bröder, 2020; Undorf & Erdfelder, 2015; Witherby & Tauber, 2017) but also when learning larger chunks of information such as lists, paragraphs, and texts (Ackerman & Goldsmith, 2011; Ariel et al., 2020; Lefèvre & Lories, 2004; Nguyen & McDaniel, 2016; Pieger et al., 2016; see Prinz et al., 2020 for a meta-analysis). For example, text cohesion has been shown to influence JOL magnitude. Text cohesion refers to linguistic cues that help readers to make connections between the presented ideas. Examples of cohesion cues include the overlap of words and concepts between sentences and the presence of discourse markers such as *because*, *therefore*, and *consequently* (Crossley et al., 2016; Halliday & Hasan, 1976). Poor cohesion texts impose higher demands on readers who would need to produce inferences to create a meaningful representation of the information in the text (Best et al., 2005). These texts are associated with poorer comprehension (Crossley et al., 2014, 2016; Hall et al., 2016).

Importantly, some studies show that participants can monitor text cohesion and adjust their JOLs accordingly (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). For example, Lefèvre and Lories (2004) manipulated cohesion by introducing or omitting a repetition of the antecedent to vary ambiguity in anaphoric processing. That is, they modified the complexity of resolving references to previously mentioned entities in the text. They observed that participants provided lower JOLs for low than for high-cohesion paragraphs. In addition, they found significant correlations between JOLs and comprehension scores. These results suggest that metacognitive monitoring is sensitive to the cohesion features of a text, as participants reported that they were poorly learning the low-cohesion texts which, indeed, were comprehended worse than high-cohesion texts. Similarly, Rawson and Dunlosky (2002) varied coherence by manipulating causal relatedness across sentence pairs and by altering the structure of sentences within paragraphs. They also found that both predictions and

memory performance were significantly lower for low-coherence pairs than for moderate-to-high coherence pairs. Finally, in Carroll and Korukina's (1999) experiment, sentence order was manipulated in narrative texts to create different coherence versions. They found a significant main effect of text coherence on both judgments and memory, as the ratings and the proportion of items that were immediately recalled were significantly greater for ordered texts than for nonordered texts.

Importantly, our study builds upon prior research by manipulating text cohesion within longer passages, contrasting with previous studies that primarily focused on cohesion within individual sentences. Overall, although JOLs have been proven to be sensitive to variations in the difficulty of the materials, people may in fact be less accurate in predicting their learning under circumstances with high cognitive load (Seufert, 2018; Wirth et al., 2020). Likewise, the language in which the learning is taking place might be a factor that mediates the cognitive and metacognitive resources that are devoted to the text (Reyes et al., 2023).

1. Present study

In two experiments, we aimed to investigate whether studying in an L2 context has an impact on the self-regulation of learning and achievement. For this, we manipulated the cohesion of texts and examined to what extent unbalanced bilinguals monitored and controlled their learning both in L1 and L2 and whether they adjusted their judgments and strategies according to the characteristics of the materials.

Previous research on bilingualism provides reasons to hypothesize that learners might find L2 materials more difficult to process in comparison with the same content presented in L1 (Reyes et al., 2023). At the same time, for the self-regulation of learning to occur, some processes must become automatic so that people can activate effective strategies (Winne, 2011; Zimmerman & Kitsantas, 2005). Hence, the nonautomatic processing and the potential extra cognitive demands imposed by L2 processing may have an impact on cognitive and metacognitive processes.

In a previous study, Reyes et al. (2023) reported that participants with intermediate L2 proficiency levels were able to correctly monitor their learning when studying in L2, adjusting their JOLs according to the difficulty of the to-be-studied material. That is, concrete words and semantically related word lists received higher JOLs, and were better remembered than abstract words and semantically unrelated word lists, respectively, both in L1 and L2. However, in this study monitoring and learning were assessed only for word lists and not for more complex materials such as expository texts, which align more with the materials used in academic and professional settings.

In the present study, we intended to extend previous results by studying the interaction among monitoring, strategies, and learning when the to-be-learned material was texts presented in participants' L1 or L2. With this aim, we varied the difficulty of the text and assessed participants' monitoring of the difficulty of the presented texts (i.e., JOL), their actual learning (i.e., open-ended questions), and the learning strategies used (i.e., self-report questionnaire). In Experiment 1, the tasks were presented to a sample of university students with an intermediate proficiency level of English–L2. In Experiment 2, higher and lower-proficiency groups were included.

Our main hypothesis was that studying texts in L2 may compromise the correct functioning of the processes implicated in self-regulated learning. On the one hand, participants might adjust their overall perception of learning according to the language context. If

participants used language as a diagnostic cue (Koriat, 1997), we would expect them to provide higher JOLs in the L1 than in the L2 contexts, assessing learning in L1 as easier and more successful compared to learning in L2. Second, we expected to observe a less accurate assessment of other cues that influence the material's difficulty in L2 compared to L1. Consequently, when participants studied in L2, they might not detect text cohesion as a useful cue for assessing their learning.

2. Experiment 1

2.1. Methods

2.2. Participants

We conducted a power analysis using G*Power (Faul et al., 2007) to determine the sample size. We calculated it considering a mixed-factor analysis of variance (ANOVA) with language and cohesion as repeated measure variables, and order of the language block as a between-participant variable. We estimated 30 participants, assuming a small to moderate effect size (partial eta-squared of 0.07) to observe significant ($\alpha = 0.05$) effects at 0.8 power. Due to an error in the text counterbalancing procedure, we had to recruit more participants to ensure a representative sample in each of the counterbalance lists. Participants had normal or corrected-to-normal vision and reported no neurological damage or other health problems. Participants gave informed consent before performing the experiment that was carried out following the Declaration of Helsinki (World Medical Association 2013). The protocol was approved by the institutional Ethical Committee of the University of Granada (857/CEIH/2019) and the Universidad Loyola Andalucía (201222 CE20371).

Sixty-eight psychology students from Universidad Loyola Andalucía (51.5%) and the University of Granada (48.5%) participated in this experiment. We removed from all analyses (1) a participant who did not vary the percentage given as a JOL in any of the texts and left it at the default value, and (2) two participants who gave answers in Spanish for both L1 and L2 block. We therefore had a total sample of 65 (18–28 years old, $M = 19.92$, $SD = 1.76$). Participants were tested remotely and individually in a two-session experiment and received course credit as compensation.

Participants were nonbalanced Spanish–English bilinguals although they started acquiring English as their L2 during childhood ($M = 6.75$, $SD = 2.75$). Subjective (Language Background Questionnaire, LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) and objective (MELICET Adapted Test, Michigan English Language Institute College Entrance Test) language measures indicated that the sample had an intermediate proficiency level in English ($M = 28.79$). See Table 1 for descriptive statistics.

2.3. Materials and Procedure

The experiment consisted of two online sessions that lasted 120 and 90 minutes respectively. We programmed tasks, presented stimuli, and collected data with Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants accessed the experiment remotely and individually. To ensure that participants did not open other windows in the computer while doing the tasks, they were forced to full-screen presentations. Recent research supports the validity and precision of experiments run online (Anwyl-Irvine et al., 2020, 2021; Gagné & Franzen, 2023).

The main task in both sessions was a *learn-judge-remember* task with a study phase and a recognition test. It simulated a learning task in a classroom environment in which students needed to learn

Table 1. Participants' information for demographic and language measures.

	Mean (SD)	
Age	19.72 (1.76)	
O-Span	0.62 (0.22)	
Self-reported measures	L1	L2
Daily exposure (%)	83.03 (15.02)	37.55 (21.00)
Age of acquisition (in years)	2.88 (1.13)	6.75 (2.76)
Age of becoming fluent (in years)	5.71 (2.29)	11.60 (3.43)
Speaking self-competence (0–10)	9.69 (0.68)	7.44 (1.46)
Reading self-competence (0–10)	9.51 (0.97)	8.16 (1.22)
Exposure to reading (0–10)	9.00 (1.29)	6.23 (2.54)
Learning by reading (0–10)	9.45 (1.06)	7.73 (2.03)
Language proficiency		
MELICET (0–50 points)		28.79 (8.47)

Note: T-tests for paired sample showed significant differences between languages in all the measures (all *p*-values < .001).

and remember information from texts either in Spanish–L1 or in English–L2, depending on the session. Additionally, we administered different tasks and questionnaires at the end of each session. Participants completed the MELICET Adapted Test (Michigan English Language Institute College Entrance Test) as an objective L2 proficiency measure.

At the end of the second session they fulfilled a customized metacognitive questionnaire regarding the strategies used when studying the texts in both languages, a language background and sociodemographic questionnaire (LEAP-Q, Marian et al., 2007), and the Spanish version of the Operational Digit Span task (O-Span) to assess that all participants ranged within normal standardized values of working memory capacity (Turner & Engle, 1989). We used a shortened version adapted from Oswald et al. (2014) in which participants were presented with a series of math problems followed by a to-be-remembered target letter. We calculated a working memory index by multiplying the mean proportion of successfully recalled letters and the mean proportion of correctly solved arithmetic equations (Conway et al., 2005).

In the *learn-judge-remember* task, participants were instructed to give a *JOL* and to answer some questions about the text they read. We manipulated the language (Spanish–L1 vs. English–L2) and the cohesion of the texts (high vs. low cohesion) as within-subjects factors. Language was a blocked variable and the assignment of L1 or L2 to the first or second session was counterbalanced across participants. Both high- and low-cohesion texts appeared along the study phase for each language block so that half of the texts within a block were of high cohesion and the other half was low cohesion.

In each session, participants were informed to read comprehensively 10 short texts for a later learning assessment test. Texts of high and low cohesion were presented in a pseudo-random order one at a time in the middle of the computer screen and remained for 3 minutes for self-reading. Immediately after the presentation of each text, participants gave a JOL to predict the likelihood of remembering the information they have just read on a 0–100 scale (0: *not all likely*, 100: *very likely*) by moving a handle slider to the desired number. This screen advanced when participants pressed ENTER.

After studying and judging each text, participants answered three open-ended questions as an objective measure of their

learning. Previous research exploring the consequences of studying in L1 vs. L2 on memory found different effects depending on the type of test. For example, Vander Beken et al. (2020) and Vander Beken and Brysbaert (2018) found that essay questions hindered performance in L2 presumably due to difficulties in writing production while no differences between L1 and L2 performance were found with open-ended questions and true/false recognition items. This suggests that language proficiency and background would make the writing process more complex and challenging in L2 than in L1. To avoid confounding effects with writing complexity, we discarded the essay and chose an open-ended format to better discriminate and prevent a possible ceiling effect that may appear with true/false recognition items. Questions covered a range of information from general ideas to examples or brief descriptions. Participants could respond with a single word, a noun phrase, or a concise sentence (e.g., "What type of teeth do meat-eating mammals have?"; see Appendix 1 for a detailed example). Open-ended questions were corrected automatically using a Python script (available at https://osf.io/dw4y7/?view_only=4eeb04437db14d69 b2269a8d19392df5) that matched a rubric criterion developed a priori. This script has been checked against a manual revision leading to a higher reliability in the final score/mark. We gave 1 point for fully correct answers in the language required and 0 points for incomplete or incorrect answers. We provided a full score if the key concepts in the rubric were included in the answer, accepting grammatical and spelling mistakes in both languages. We calculated the mean proportion of correct recall for each participant and condition.

We selected 22 English texts related to academic topics, which we retrieved from different books and previous studies: two of them were used as examples and the rest were testing materials (see Appendix 2 for references and find them available at https://osf.io/dw4y7/?view_only=4eeb04437db14d69b2269a8d19392df5). All texts included intermediate vocabulary, taught in secondary school. In this manner, we ensured that our university students-participants were familiar with it, given that understanding academic vocabulary is crucial for comprehending academic texts (Silva & Otwinowska, 2019). A Spanish native speaker translated texts into Spanish and manipulated their cohesion, both in English and Spanish texts, following norms from previous studies (see Table 2). Efforts were made to make the minimal

Table 2. Norms for manipulating text cohesion.

High-cohesion texts	References
Including connectors.	Gasparinatou and Grigoriadou (2013); Ozuru et al. (2012)
Increasing noun overlap.	Hall et al. (2016)
Shortening sentences.	Soemer and Schiefele (2019)
Low-cohesion texts	
Lengthening sentences.	Soemer and Schiefele (2019)
Including subordinate phrases.	
Replacing nouns with pronouns to create ambiguity.	Hall et al. (2016); Soemer and Schiefele (2019)
Using low-frequency synonyms for key concepts.	Hall et al. (2016); Soemer and Schiefele (2019)
Using the passive voice.	
Having abrupt gaps between sentences.	Soemer and Schiefele (2019)

changes possible and to ensure that the texts were comparable in terms of their difficulty (e.g., word frequency). Two judges, both native Spanish speakers, double-checked the translations and the cohesion manipulation. This resulted in four different versions for every text, one per language-cohesion condition: L1 high cohesion and L1 low cohesion, and L2 high cohesion and L2 low cohesion.

We created four counterbalanced text lists with each list containing five texts per language-cohesion condition. Therefore, each participant was presented with 20 different texts: 10 in the L1 session and 10 in the L2 session, of which five were low- and five high-cohesion texts. Repeated measures ANOVA (cohesion and language) showed that the texts were matched in length (number of words) between conditions, as no main effects or interaction were significant [all p-values > .05; L1: high ($M = 142.6$, $SD = 22.8$) and low cohesion ($M = 141.0$, $SD = 28.4$); L2: high ($M = 141.0$, $SD = 26.4$) and low cohesion ($M = 140.3$, $SE = 29.6$)]. See Appendix 1 for an example of a high- and low-cohesion version of a text in L2 and its open-ended questions.

For the *learn-judge-remember task*, we analyzed JOL responses in the study phase and the proportion of correct answers for the open-ended questions, grouped by condition (language and text cohesion).

In each language block, after the *learn-judge-remember* task, participants answered a customized metacognitive self-report questionnaire. We combined items from two different inventories into a single set of 8 questions and translated into Spanish by a Spanish native speaker, with subsequent double-checking by a second and third judge to ensure accuracy and consistency. We selected items from the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) and the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991). Both are validated questionnaires designed to be modular and can be used to fit the needs of the researcher. Although we could not test the reliability of the Spanish translation of this *ad hoc* questionnaire (as it was beyond the purpose of our study) it could serve to understand the metacognitive processes engaged in our participants' performance. Participants rated themselves on a seven-point Likert scale from "not at all true of me" to "very true of me."

Thus, we assessed cognitive and metacognitive learning strategies, effort regulation, mental demand, and self-perceived performance. Originally, a few items were intended for a general learning context, so we modified some expressions for the specific task. We also translated the items into Spanish since we administered the questionnaire in the language the session was taking place. The item referring to metacognitive self-regulation was reversed and thus we inverted their punctuation. For this questionnaire, we compared the score for items in L1 and L2. See Appendix 3 to check the set of questions included into the questionnaire.

3. Results

We performed a 2 × 2 × 2 (language × text cohesion × block order) mixed-factor ANOVAs for JOLs in the study phase and for learning assessment test. We included language order in the analyses since previous research (Reyes et al., 2023) suggests that it might influence monitoring and memory performance. Language (L1 vs. L2) and text cohesion (high vs. low cohesion) were within-subject factors and block order (L1-first vs. L2-first) was a between-subject factor. For all analyses, the alpha level was set to 0.05 and we corrected by Bonferroni for multiple comparisons. All effect sizes are reported in terms of partial-eta-squared (η_p^2) for ANOVAs and Cohen's d for t-tests.

Study phase (JOLs). To evaluate the effect of language and text cohesion on the magnitude of JOLs, we computed the mean across participants' JOLs for each condition (see Table 3 for partial means).

None of the interactions were significant (all p-values > .05). We did find significant main effects of language, $F(1, 63) = 20.13$, $p < .001$, $\eta_p^2 = .24$, and cohesion, $F(1, 63) = 15.35$, $p < .001$, $\eta_p^2 = .20$. Texts in L1 ($M = 68.8$, $SE = 1.58$) received higher JOLs than texts in L2 ($M = 61.6$, $SE = 1.95$). Similarly, high-cohesion texts received higher JOLs ($M = 66.5$, $SE = 1.60$) than low-cohesion texts ($M = 63.8$, $SE = 1.64$). The main effect of block order was not significant, $F(1, 63) = 0.06$, $p = .808$, $\eta_p^2 = .001$. Texts received comparable JOLs regardless of the language order (L1 first: $M = 64.8$, $SE = 2.22$; L2 first: $M = 65.5$, $SE = 2.26$).

Learning assessment test (accuracy). To evaluate learning performance, that is, how much participants remembered from the texts, we computed the proportion of correct responses in the open-ended questions across participants (see Table 3).

None of the interactions were significant (all p-values > .05). Yet, the analysis showed a significant main effect of language, $F(1, 63) = 6.41$, $p = .014$, $\eta_p^2 = .10$, and cohesion, $F(1, 63) = 24.66$, $p < .001$, $\eta_p^2 = .28$. That is, participants remembered information from texts in L1 ($M = 0.71$, $SE = 0.02$) better than from texts in L2 ($M = 0.65$, $SE = 0.02$); and from high-cohesion texts ($M = 0.71$, $SE = 0.02$) better than from low-cohesion texts ($M = 0.65$, $SE = 0.02$). The main effect of block order did not reach significance, $F(1, 63) = 1.62$, $p = .21$, $\eta_p^2 = .03$. Participants' accuracy in the learning assessment tests did not depend on which language block they performed on the first or second place (L1 first: $M = 0.65$, $SE = 0.03$; L2 first: $M = 0.70$, $SE = 0.03$).

Language metamemory accuracy—resolution. To examine participants' metamemory accuracy—resolution—across languages (i.e., to check whether participants' JOLs discriminate between the information recall of one text relative to another), we used a Goodman–Kruskal (GK) gamma correlation (Nelson, 1984) and a language-accuracy index correlation.

Table 3. Mean (and standard deviations) for JOL scores (1–100 scale) and learning performance (proportion of correct responses) across language, cohesion, and block order conditions.

Language	Block order	High-cohesion		Low-cohesion	
		JOL	Learning	JOL	Learning
L1	L1-first	69.4 (2.27)	0.74 (0.03)	66.2 (2.32)	0.65 (0.03)
	L2-first	70.9 (2.31)	0.75 (0.03)	68.6 (2.35)	0.70 (0.03)
L2	L1-first	63.0 (2.77)	0.65 (0.04)	60.5 (2.96)	0.58 (0.03)
	L2-first	62.8 (2.81)	0.71 (0.04)	59.9 (3.01)	0.65 (0.03)

GK gamma correlation is a nonparametric measure of the association between JOLs and subsequent recall. We calculated one gamma correlation for each participant in each of the language conditions. We then ran a *t*-test to examine whether the GK Gamma correlations differed across languages. No significant effects were found, $t(64) = -0.76$, $p = 0.45$, $d = -0.09$.

We also performed a language-accuracy index correlation as an additional measure of metamemory accuracy, which allows us to further explore participant's overall resolution in L1 and L2. To do so, we first calculated a language index for JOLs and for learning accuracy, by subtracting the mean scores in L2 from the mean scores of L1 of JOLs and learning accuracy respectively, and then performed correlation analyses of the two indexes. Interestingly, the JOLs index correlated with the accuracy index ($r = 0.6$), suggesting that participants' predictions during the study phase about what they would remember later agreed with what they actually recalled in the learning assessment test.

Customized metacognitive self-report questionnaire. We analyzed the questionnaire regarding the participants' learning strategies in the study phase (see Table 1S in Supplementary Materials for partial means in the questionnaire). We conducted *t*-tests to compare the frequency of each strategy in L1 and L2. Overall, we found that participants employed some strategies more frequently in L1 than in L2, namely, elaboration ("When reading the texts, I tried to relate the material to what I already knew"), $t(64) = 2.17$, $p = .033$, $d = 0.27$, metacognitive self-regulation ("When studying the materials in the texts, I often missed important points because I was thinking of other things"), $t(64) = 2.71$, $p = .008$, $d = 0.34$, and effort regulation ("I worked hard to do well even if I didn't like what I was studying in the texts"), $t(64) = 2.06$, $p = .044$, $d = 0.26$. No differences were found in terms of critical thinking strategy, $t(64) = -0.60$, $p = .55$, $d = -0.07$, or rehearsal, $t(64) = -0.59$, $p = .56$, $d = -0.07$. No strategy was more frequently used in L2 than in L1 either. As expected, and consistent with JOLs, participants reported that they experienced significantly higher mental demand in L2 than in L1, $t(64) = -8.81$, $p < .001$, $d = -1.09$. Similarly, participants felt their performance had been better in L1 than in L2, $t(64) = 2.17$, $p = .034$, $d = 0.27$.

3.1. Discussion

In this first experiment, we focused on investigating two potential outcomes. Initially, we wanted to observe if a cue such as text cohesion yielded variations in JOLs and learning. We also sought to ascertain whether these differences depended on the language environment (L1 or L2) within which the task was executed. Additionally, we aimed to evaluate whether that linguistic context exerted an influence on the global perception of task complexity.

Regarding text cohesion, we found a cohesion effect both in JOLs and in the learning assessment test. Participants predicted better performance for high-cohesion texts and, correspondingly, learning rates were higher for them compared to those with low cohesion. Thus, we replicated what had previously been reported in the monolingual text monitoring and comprehension literature (Carroll & Korukina, 1999; Crossley et al., 2014, 2016; Hall et al., 2016; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). More importantly, we expanded these findings to a bilingual sample, suggesting that the processes operating in L2 are similar to those in L1 concerning text monitoring and comprehension.

As to the language effect, participants judged L1 materials as easier to learn—giving higher JOLs—than materials in L2. As

shown by the learning assessment test, participants encountered significantly more difficulty in remembering information from texts in L2. Previous research exploring the consequences of studying in L1 versus L2 on memory observed different effects depending on the test type used. For example, Vander Beken et al. (2020) and Vander Beken and Brysbaert (2018) found that essay questions hindered performance in L2 presumably due to difficulties in writing production, while no differences between L1 and L2 performance were found with open-ended questions and true/false recognition items. Nevertheless, our participants did show an L2 recall cost despite the fact that we chose an open-ended format to avoid confounding effects with writing complexity and that our rubric accepted grammatical, syntactic, or orthographic errors (note that language mistakes are not punished—as long as they do not obscure meaning—in international reading comprehension assessments like the Programme for International Student Assessment, PISA). Hence other factors related to the type of processing or strategies used during L1 and L2 may have produced differences in L1 and L2 memory performance.

In sum, the results of Experiment 1 indicate that participants with an intermediate level of English–L2 were able to use intrinsic cues such as text cohesion and language simultaneously, to monitor their learning both in L1 and L2. However, in Experiment 1 we did not manipulate participants' L2 proficiency, and it was plausible that differences in language proficiency and exposure could have influenced the monitoring behavior of the bilingual individuals. To explore this possibility, we conducted a second experiment to investigate the effects of L2 proficiency levels on monitoring and control processes of self-regulated learning. For this, we intentionally recruited participants with higher and lower L2 proficiency.

4. Experiment 2

In the second experiment, our aim was to investigate the influence of L2 proficiency on the dynamic relationship between monitoring and control during text-based learning. To achieve this, we recruited a sample that included both lower and higher English–L2 proficiency levels. We hypothesized that individuals in the lower-proficiency group might encounter challenges in effectively monitoring their learning due to allocating a greater share of cognitive resources to language control compared to their higher-proficiency counterparts (Francis & Gutiérrez, 2012; Sandoval et al., 2010). As a result, the cohesion effect in JOLs, which manifests as higher values with well-cohesive texts, could potentially diminish within the lower-proficiency group. This attenuation might arise from the substantial cognitive load imposed by learning in a demanding L2 context, potentially overshadowing the sensitivity to nuanced differences in text cohesion (see Magreehan et al. 2016, who did not find the font type effect on JOLs when other cues were available). In essence, we posit that texts in L2 may present inherent challenges for individuals with lower proficiency, regardless of their cohesion status.

In addition, we introduced two further modifications: (1) as block order was not significant in Experiment 1, we eliminated this variable from the procedure, and high- and low-cohesion texts in L1 and L2 appeared along the study phase in a pseudorandom order; (2) participants attended an in-person session at the laboratory for the second part of the experiment. The remaining conditions were held constant, mirroring the setup employed in Experiment 1.

4.1. Methods

4.1.1. Participants

Instructions for recruitment indicated that participants needed to have some English knowledge, although we did not specify the threshold for participation. Fifty-seven psychology students from the University of Granada (63.17%) and Universidad Loyola Andalucía (36.84%) enrolled in the experiment. Participants were tested individually in two sessions (a remote and an in-person session) and received course credit as compensation. We divided our sample into two independent groups by scores in MELICET. Based on previous studies (Kaan et al., 2020; López-Rojas et al., 2022), we established scores of 30 or above as the criteria to be included in the higher-proficiency group (upper-intermediate level, $n = 23$, 18–23 years old, $M = 19.05$, $SD = 1.36$), and scores of 25 or below were classified into the lower-proficiency group (pre-intermediate level, $n = 24$, 18–49 years old, $M = 21$, $SD = 6.81$). Participants with in-between scores were not included in the analyses. In addition, we removed two participants who did not vary the percentage given as a JOL in any of the texts and left it at the default value, so this resulted in a total sample of 49. No differences were found in the O-Span index (following the same calculation described in Experiment 1) between groups $t(45) = 1.49$, $p = .14$, $d = 0.43$ (higher-proficiency group: $M = 0.60$, $SE = 0.04$; lower-proficiency group: $M = 0.51$, $SE = 0.05$) suggesting that any possible difference between groups were not due to differences in working memory capacity. Comparisons between languages for all self-reported linguistic measures within groups showed that participants in both groups were unbalanced and significantly more fluent in L1 than in L2. All p-values were below .05. See Table 4 for further details.

4.1.2. Materials and procedure

The experiment consisted of two sessions for which participants received course credit as compensation. We programmed and administered the tasks with the same experiment builder as in Experiment 1 (Anwyl-Irvine et al., 2020). The first session lasted 30 minutes and was administered remotely. Participants completed some questionnaires, the LEAP-Q (Marian et al., 2007) and an objective L2 proficiency measure (MELICET). Then participants came to the laboratory to complete the second session in person, which lasted 120 minutes. The procedure was similar to that of Experiment 1, with a *learn-judge-remember* task, a customized metacognitive questionnaire regarding the strategies used when studying the texts, and the standard Operational Digit Span task (O-Span) to assess that all participants ranged within normal standardized values and that groups did not differ in working memory (Oswald et al., 2014).

4.2. Results

We report a 2 × 2 × 2 (language × text cohesion × proficiency group) mixed-factor ANOVAs for JOLs in the study phase and for accuracy in the learning assessment test. Language (Spanish–L1 vs. English–L2) and text cohesion (high vs. low) were within-subject factors, and the proficiency group (higher vs. lower) was a between-subject factor. As in Experiment 1, the alpha level was set to 0.05 and we corrected it by Bonferroni for multiple comparisons, in all analyses. All effect sizes are reported in terms of partial-eta-squared (η_p^2) for ANOVAs and Cohen's d for t-tests. Table 5 shows partial means for JOLs and accuracy in the learning assessment test.

Study phase (JOLs). Analysis on the JOLs showed a significant interaction between language and proficiency group, $F(1, 45) = 23.48$, $p < .001$, $\eta_p^2 = .34$. Post-hoc comparisons showed that participants in the higher-proficiency group did not differ in their JOLs between languages, $t(45) = -0.77$, $p = 1.00$ (L1: $M = 62.0$, $SE = 3.08$; L2: $M = 61.5$, $SE = 3.39$) whereas participants in the lower-proficiency group gave significantly higher JOLs for texts in L1 ($M = 65.3$, $SE = 3.02$) than for texts in L2 ($M = 49.4$, $SE = 3.32$), $t(45) = 7.13$, $p < .001$. Overall, this trend suggests that proficiency was modulating JOL values when studying in L2.

The interaction between language and cohesion was also significant, $F(1, 45) = 66.4$, $p < .001$, $\eta_p^2 = .60$. Post-hoc comparisons showed that JOLs for low-cohesion texts differed between languages, $t(45) = 8.31$, $p < .001$, while no significant difference between languages was found for JOLs in high-cohesion texts, $t(45) = 0.17$, $p = 1.00$. That is, participants gave higher JOLs in L1 for low-cohesion texts ($M = 66.1$, $SE = 2.2$) than in L2 ($M = 50.1$, $SE = 2.63$). This difference was not significant for high-cohesion

Table 4. Participants' information for demographic and language measures divided by proficiency group (higher- and lower-proficiency).

	Higher-proficiency		Lower-proficiency	
Age	19.05 (1.36)		21.64 (6.80)	
O-Span index	0.60 (0.20)		0.50 (0.22)	
Self-reported measures	L1	L2	L1	L2
Daily exposure (%)	70.3 (16.50)	32.6 (17.00)	90.2 (9.52)	22.80 (22.70)
Age of acquisition (in years)	2.84 (1.45)	5.98 (2.18)	3.14 (1.92)	7.41 (2.54)
Age of becoming fluent (in years)	5.41 (2.10)	10.80 (3.89)	5.92 (2.91)	13.90 (7.56)
Speaking self-competence (0–10)	9.74 (0.70)	8.09 (1.16)	9.62 (1.42)	6.35 (1.77)
Reading self-competence (0–10)	9.70 (0.47)	8.41 (1.40)	9.58 (0.81)	7.23 (1.42)
Exposure to reading (0–10)	8.57 (1.47)	6.52 (2.78)	9.19 (1.41)	4.84 (2.62)
Learning by reading (0–10)	9.35 (1.03)	8.22 (2.35)	8.96 (1.51)	6.81 (2.32)
Second language proficiency				
MELICET (0–50 points)	–	38.70 (5.26)	–	17.04 (5.15)

Note: Higher-proficiency group ($n = 23$) scored 30 or more in MELICET ($M = 38.7$, $SE = 1.1$) while lower-proficiency group ($n = 24$) scored 25 or less ($M = 17.7$, $SE = 0.85$) and significant differences between groups in this measure was found $t(45) = 15.2$, $p < 0.01$, $d = 4.44$. (*) Significant differences between groups ($p < .05$).

Table 5. Mean (and standard deviations) for JOLs scores (1–100 scale) and learning performance (proportion of correct responses) across language, cohesion, and proficiency group conditions.

Language	Proficiency group	High-cohesion		Low-cohesion	
		JOL	Learning	JOL	Learning
L1	Higher-proficiency	59.5 (3.24)	0.70 (0.03)	64.5 (3.14)	0.62 (0.04)
	Lower-proficiency	62.8 (3.17)	0.65 (0.03)	67.8 (3.08)	0.60 (0.04)
L2	Higher-proficiency	66.3 (3.37)	0.74 (0.05)	56.7 (3.76)	0.59 (0.04)
	Lower-proficiency	55.3 (3.30)	0.39 (0.05)	43.5 (3.68)	0.31 (0.04)

texts, that received comparable JOLs in both languages (L1: $M = 61.1$, $SE = 2.26$; L2: $M = 60.8$, $SE = 2.36$). This suggests that the cohesion effect is significantly more salient in L2 and texts with low cohesion in L2 were the most difficult condition among all four.

The main effects of language, $F(1, 45) = 26.41$, $p < .001$, $\eta_p^2 = .37$, and cohesion, $F(1, 45) = 8.47$, $p = .006$, $\eta_p^2 = .16$ were significant. As expected, texts in L1 received higher JOLs ($M = 63.6$, $SE = 2.16$) than texts in L2 ($M = 55.4$, $SE = 2.37$). Similarly, texts with high cohesion received higher JOLs ($M = 61.0$, $SE = 2.13$) than texts with low cohesion ($M = 58.1$, $SE = 2.22$). The main effect of the group did not reach significance, $F(1, 45) = 1.07$, $p = .31$, $\eta_p^2 = .02$. Overall, the higher-proficiency group ($M = 61.7$, $SE = 3.03$) gave similar JOLs values to the lower-proficiency group ($M = 57.3$, $SE = 2.97$).

Learning assessment test (accuracy). As for accuracy, the analysis showed a similar pattern. We found a significant interaction between language and proficiency group, $F(1, 45) = 24.65$, $p < .001$, $\eta_p^2 = .35$. Post-hoc comparisons showed that the higher-proficiency group had comparable accuracy in both languages, $t(45) = -0.14$, $p = 1.00$, so their learning was similar in L1 ($M = 0.66$, $SE = 0.03$) and L2 ($M = 0.67$, $SE = 0.04$). In contrast, accuracy for participants with lower proficiency did differ between languages, $t(45) = 6.96$, $p < .001$, and they achieved significantly better learning for texts in L1 ($M = 0.63$, $SE = 0.03$) than for texts in L2 ($M = 0.35$, $SE = 0.04$). It seems that proficiency plays a role when learning in L2 and lower-proficiency level might hinder learning.

The interaction between language and cohesion was marginally significant, $F(1, 45) = 2.88$, $p = .096$, $\eta_p^2 = .06$. Post hoc comparisons showed that accuracy for texts in L2 differed between cohesion condition, $t(45) = 4.50$, $p < .001$. Participants were more accurate in high-cohesion texts ($M = 0.56$, $SE = 0.03$) than in low-cohesion texts ($M = 0.45$, $SE = 0.03$). Nevertheless, no significant difference between cohesion conditions was found for texts in L1, $t(45) = 2.36$, $p = .14$ (high cohesion: $M = 0.67$, $SE = 0.02$; low cohesion: $M = 0.61$, $SE = 0.03$). Again, it seems that the cohesion effect is significantly more salient in L2 not only in JOLs but also in learning.

We found three significant main effects: language, $F(1, 45) = 22.77$, $p < .001$, $\eta_p^2 = .34$, with texts in L1 ($M = 0.64$, $SE = 0.02$) receiving higher scores in the learning assessment test than texts in L2 ($M = 0.51$, $SE = 0.03$); cohesion, $F(1, 45) = 18.82$, $p < .001$, $\eta_p^2 = .30$, with high-cohesion texts receiving higher scores ($M = 0.62$, $SE = 0.02$) than low-cohesion texts ($M = 0.53$, $SE = 0.02$); and proficiency group, $F(1, 45) = 17.9$, $p < .001$, $\eta_p^2 = .29$, with participants in the higher-proficiency group ($M = 0.66$, $SE = 0.03$) achieving overall better learning than participants in the lower-proficiency group ($M = 0.49$, $SE = 0.03$).

Language metamemory accuracy—resolution. We run a mixed-factor ANOVA and we found no significant main effect of Goodman–Kruskal Gamma correlations $F(1, 44) = 2.20$, $p = .15$, $\eta_p^2 = .05$, nor proficiency group, $F(1, 44) = 0.00$, $p = .96$, $\eta_p^2 = .00$.

However, the interaction between both factors was significant, $F(1, 44) = 4.93$, $p = .03$, $\eta_p^2 = .10$. Post-hoc comparisons showed a marginal tendency for lower-proficiency group to have better resolution in L2 ($M = 0.32$, $SD = 0.1$) than in L1 ($M = -0.00$, $SD = 0.09$), $t(44) = -2.62$, $p = 0.07$, while no difference was found for the high-proficiency group (L1: $M = 0.18$, $SD = 0.09$; L2: $M = 0.12$, $SD = 0.1$, $t(44) = 0.52$, $p = 1.00$).

We also ran JOL and accuracy language index correlations with the two proficiency groups independently, and we observed that a significant positive correlation appeared for participants with higher proficiency ($r = 0.42$, $p = .05$) while participants with lower proficiency showed a significant negative correlation between JOL and accuracy index ($r = -0.43$, $p = .04$).

4.3. Additional analyses collapsing across both experiments

To further explore proficiency effects, we performed statistical analyses by collapsing across both experiments. Collapsing data from the two experiments allowed us to examine the effect of L2 proficiency as a continuous variable, increasing sample size and L2 variability. Note, however, that the two experiments differed in the form of presenting L1 and L2 (blocked or mixed) and not only in the proficiency level of our participants. Therefore, we included Experiment as a variable for these analyses. We used linear mixed-effect models, as implemented in the lme4 package (version 1.1–27.1; Bates et al., 2014) in R using participants and items (texts) as crossed random effects. We selected JOLs and accuracy as the dependent variables of each model. We included as fixed effects language (L1, L2), cohesion condition (high, low), L2 proficiency (scores in MELICET), experiment (Exp1: language blocking, Exp 2: language mixing), and the interactions between them. We transformed the continuous variables (L2 proficiency) to normalize the distribution by scaling the scores. We selected sum contrast for language (L1 = −1; L2 = 1), cohesion (high = −1; low = 1) and experiment (Exp1 = −1; Exp2 = 1). We fitted the maximal model first (Barr et al., 2013), and in case of nonconvergence or singularities, we simplified it following recommendations outlined in Bates et al. (2021). We considered significant any fixed effect with a t-statistic higher than 2.

In JOLs, the results showed an interaction between language and cohesion. Low cohesion texts in L2 ($M = 54.5$; $SE = 2.11$) received significantly lower JOLs than low cohesion texts in L1 ($M = 67.2$; $SE = 1.95$), $t(50.9) = 12.72$, $p < .0001$, while no differences were found between languages for texts with high cohesion (L1: $M = 66$; $SE = 1.95$; L2: $M = 62$; $SE = 2.10$), $t(48.8) = 4.03$, $p = .11$. The interaction between language and L2 proficiency was also significant. The higher the L2 proficiency, the higher the JOL values stemming from higher JOLs in L2 (see Figure 1). In other words, the difference in JOLs between L1 and L2 was significant for
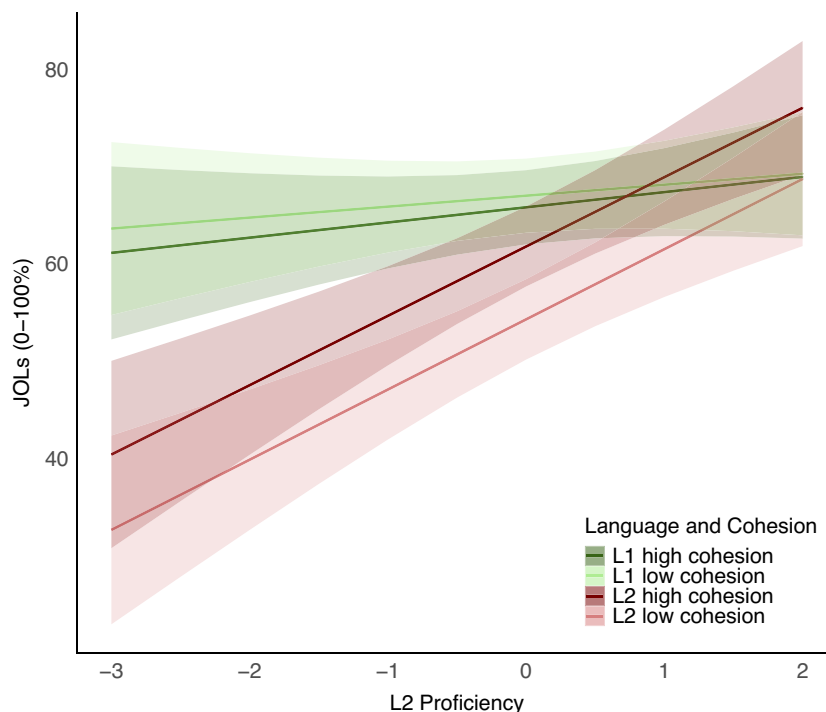
Figure 1. Reported JOLs along L2 proficiency divided by language and cohesion, with data collapsed across the two experiments. Note that higher L2 proficiency leads to higher JOL values in L2. There are significant differences between L1 and L2 for participants with low and medium L2 proficiency, but no language differences for those with high L2 proficiency.

people with low, $t(71.2)= 25.86$, $p < .0001$, and medium L2 proficiency, $t(45.7)= 11.29$, $p < .0001$, whereas people with higher L2 proficiency showed no differences between languages, $t(73.5)= -3.28$, $p = .29$.

The triple interaction between language, cohesion, and experiment was also significant (see Figure 2). In Experiment 1, where language was blocked, the difference between languages was significant in both cohesion conditions (high L1: $M = 68.7$; $SE = 2.25$,
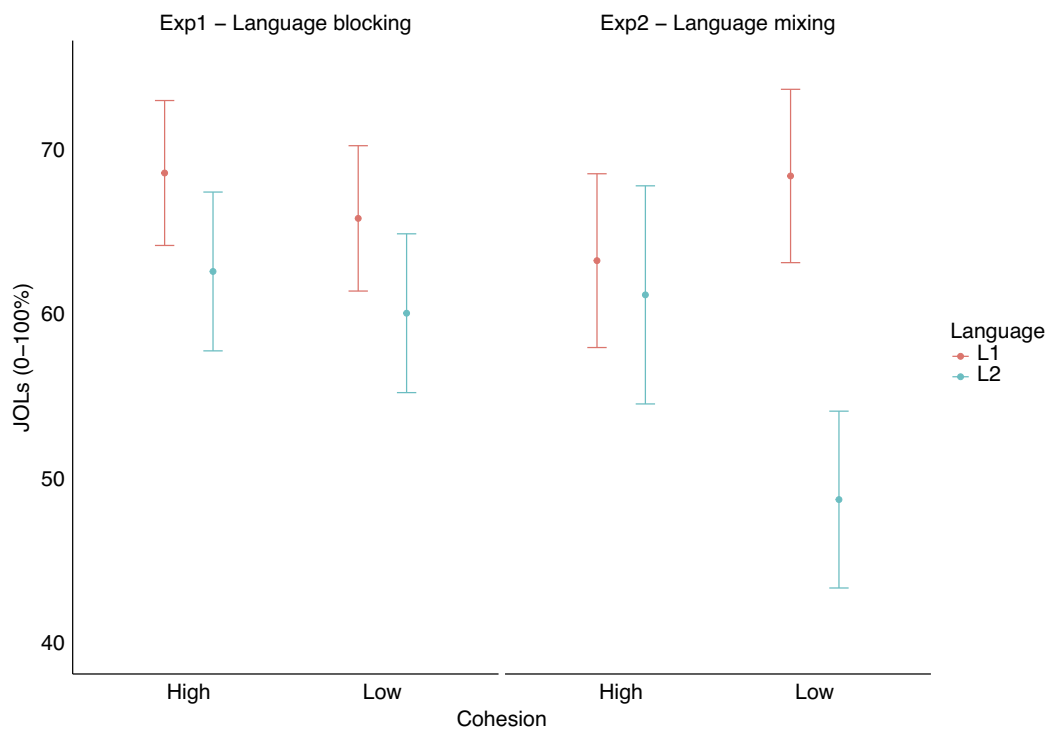

Figure 2. Reported JOLs divided by language, cohesion, and L2 proficiency by experiment. Note: Significant triple interaction between language, cohesion, and experiment. In Experiment 1, L1 texts received significantly higher JOLs than L2 texts, regardless of cohesion conditions. In Experiment 2, this difference only appeared for low-cohesion texts, with no significant language differences for high-cohesion texts.

high L2: $M = -62.7$; $SE = 2.47$, $t(69) = 5.99$, $p = .03$; low L1: $M = 65.9$; $SE = 2.26$; low L2 $M = 60.2$; $SE = 2.47$, $t(69.2) = 5.76$, $p = .03$). However, in Experiment 2, where language was mixed in the study phase, the difference in JOLs between languages (L1 vs. L2) was only significant for low cohesion texts, where L1 received significantly higher JOLs values ($M = 68.5$; $SE = 2.71$) than L2 ($M = 48.9$; $SE = 2.75$), $t(100) = 19.67$, $p < .0001$, with no significant differences in high cohesion texts (L1: $M = 63.4$; $SE = 2.71$; L2: $M = 61.3$; $SE = 3.39$, $t(53.6) = 2.08$, $p = .58$). Hence, analysis of the collapsed JOLs data did not change the pattern of results observed when the analyses were performed separately for each experiment.

For the learning assessment test, the results showed a significant interaction between language and L2 proficiency. Again, as shown in Figure 3, the difference in accuracy between L1 and L2 was significant for participants with low, (L1: $M = 0.49$; $SE = 0.29$; L2: $M = -0.99$; $SE = 0.28$, $t(\text{Inf})$ 1.48, $p < .0001$), and medium L2 proficiency, (L1: $M = 0.80$; $SE = 0.17$; L2: $M = 0.21$; $SE = 0.17$, $t(\text{Inf}) = 0.60$, $p = .01$), whereas people with higher L2 proficiency showed no differences between languages (L1: $M = 1.11$; $SE = 0.29$; L2: $M = 1.41$; $SE = 0.23$, $t(\text{Inf}) = -0.29$, $p = .31$). See supplementary materials for a summary of the JOLs mixed-effect model and the accuracy mixed-effect model.

Customized metacognitive self-report questionnaire. As in Experiment 1, we also explored whether the strategies differed between proficiency groups (see Table 2S in Supplementary Materials for partial means in the questionnaire). We run repeated measures ANOVAs for each of the items in the questionnaire with language as a within-subject factor and proficiency group as a between-subject factor.

We found a significant interaction between the **elaboration** strategy and proficiency group, $F(1, 45) = 4.37$, $p = .04$, $\eta_p^2 = .09$. However, post-hoc analysis showed no significant difference in any of the comparisons (all $p > .05$).

We also found a significant difference in the use of **metacognitive self-regulation** between languages $F(1, 45) = 4.12$, $p = .05$, $\eta_p^2 = .08$, with such strategy being more prevalent in L2 ($M = 4.11$, $SD = 0.27$) than in L1 ($M = 3.73$, $SD = 0.25$) regardless of the proficiency group.

Regarding the use of **effort regulation**, we found a marginal significant main effect of proficiency group, $F(1, 45) = 3.85$, $p = .06$, $\eta_p^2 = .08$, that was mediated by a significant interaction, $F(1, 45) = 11.41$, $p = .002$, $\eta_p^2 = .20$. Post-hoc comparison showed a significant difference between proficiency groups in the use of effort regulation in L2 (high-proficiency group: $M = 2.43$, $SE = 0.39$; low-proficiency group: $M = 4.13$, $SE = 0.39$), $t(45) = -3.07$, $p = .02$.

The main effect of **mental demand** was also significant, $F(1, 45) = 49.1$, $p < .001$, $\eta_p^2 = .52$, and was mediated by a significant interaction with proficiency group $F(1, 45) = 10.5$, $p = .002$, $\eta_p^2 = .19$. Participants in the lower group reported higher mental demand in L2 ($M = 5.63$, $SE = 0.24$) than in L1 ($M = 3.38$, $SE = 0.3$), $t(45) = -7.33$, $p < .001$. Such difference was only marginally significant for the higher-proficiency group (L1: $M = 3.91$, $SE = 0.32$; L2: $M = 4.74$, $SE = 0.25$, $t(45) = -2.63$, $p = .07$).

We found a significant main effect of **performance**, $F(1, 45) = 7.05$, $p = .003$, $\eta_p^2 = .18$, with participants reporting better self-perceived performance for texts in L1 ($M = 5.02$, $SE = 0.17$) than in L2 ($M = 4.47$, $SE = 0.19$), regardless of their proficiency level.

Similarly, we found a significant main effect of **effort**, $F(1, 45) = 4.47$, $p = .04$, $\eta_p^2 = .09$, with participants reporting having made
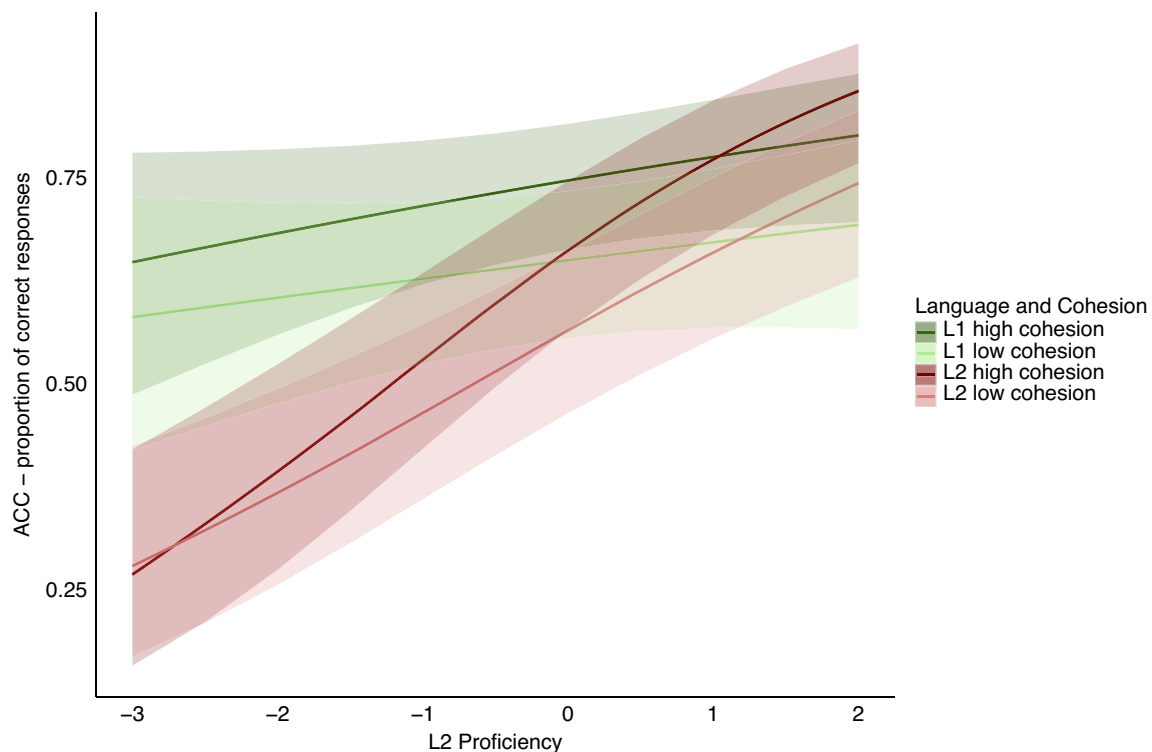

Figure 3. The proportion of correct responses along L2 proficiency divided by language and cohesion, with data, collapsed across the two experiments. Note that significant interaction between language and L2 proficiency in the learning assessment test. The accuracy difference between L1 and L2 was significant for participants with low and medium L2 proficiency, while no differences were observed for those with high L2 proficiency.

higher effort for texts in L2 (M = 4.95, SE = 0.19) than in L1 (M = 4.55, SE = 0.21), regardless of their proficiency level.

4.4. Discussion

In Experiment 2 we wanted to explore whether the effects encountered in the first experiment varied as a function of L2 proficiency level. Overall, we observed a similar pattern of results in both studies wherein we replicated the effects of language and cohesion in both JOLs and memory. Specifically, individuals were assigned lower JOLs when studying texts in L2 (as opposed to texts in L1) and for low-cohesion texts (in contrast to high-cohesion texts). However, different patterns emerged between the higher- and lower-proficiency groups in both the monitoring measure (JOLs) during the study phase and in the subsequent learning assessment test.

First, the cohesion effect interacted with the language effect so that JOLs for texts with high cohesion did not differ between languages, whereas JOLs for texts with low cohesion were significantly higher in L1. The observed interaction suggests that high-cohesion texts may create a perception of learnability regardless of the language in which they are presented. However, the significant difference in JOLs between languages for low-cohesion texts indicates that the impact of cohesion on perceived learning difficulty might vary considerably between L1 and L2 and people found the L2 low-cohesion texts as the most difficult condition among all. For learning outcomes, this interaction revealed that people achieved similar learning within cohesion conditions for texts in L1, but high-cohesion texts were favored, compared to low-cohesion texts, in L2. Results from previous studies seem to contradict this pattern. For example, Jung (2018) investigated whether cognitive task complexity affects L2 reading comprehension and found that task complexity did not affect reading comprehension scores, although participants perceived the complex tasks significantly more demanding. Nevertheless, bilingual participants in Jung's (2018) study reported staying in English-speaking countries, for at least 6 months. In our experiment, even people in the high-proficiency group were moderate in English–L2 and reported significantly less frequent exposure and use of English–L2 compared to Spanish–L1. This makes the results of the two studies difficult to compare.

More interestingly, the language effect in JOLs was modulated by the proficiency level. The differences encountered in JOLs between languages were only evident for the lower-proficiency group, who predicted greater difficulties in L2 compared to L1, as opposed to the higher-proficiency group, who predicted similar performance in both languages. This was exactly what the learning assessment test revealed. Participants with higher proficiency learned information from texts in L1 and L2 equally. However, participants with lower proficiency levels showed an L2 cost for learning. It seems that the proficiency level plays a crucial role in L2 self-regulated learning. These results go in line with previous research that had already reported that less proficient L2-English speakers needed longer time for reading, particularly when encountering sentences that conflicted with the previously established expectations in the text. This indicates a decreased efficacy in the high-level cognitive processes involved in L2 processing (Pérez et al., 2023). On the contrary, in their study, higher-proficient participants showed better text comprehension, and better ability to generate predictive inferences. Thus, they concluded that linguistic proficiency makes a difference in high-ordered processes such as inferential evaluation, revision, and text comprehension (Pérez et al., 2023). It seems that when studying in L2, lower-

proficiency learners might encounter greater challenges, especially when the to-be-study material is ambiguous or incongruent.

Note that when merging data from both experiments, we found a three-way interaction involving language, proficiency, and type of study (language blocking vs. language mixing) in JOLs. The language difference in low cohesion texts was significant only for low-proficient participants in the language mixing condition (Experiment 2). Presenting texts with mixed languages likely increased the prominence of language as a cue for participants with lower L2 proficiency due to the need for language control from switching.

In sum, it seems that participants could correctly monitor their learning regardless of their proficiency level. Participants were able to detect the difficult parts of the material and adjust their judgments accordingly. Moreover, the learning assessment test was consistent with their predictions. This is true even when they are not highly proficient, as shown by the accuracy resolution. The interaction between the Goodman–Kruskal correlations and the proficiency group showed that the lower proficiency group was more accurate in L2 texts than in L1, suggesting that their performance in L2 was consistent with what they had previously predicted (JOLs) and reported in the questionnaire (higher mental demand L2 condition).

Apparently, participants could devote sufficient cognitive resources so as to unfold metacognitive strategies and correctly monitor their learning. Results from the qualitative questionnaire suggested that participants engaged different learning strategies depending on the language (L1 vs. L2) and their L2 proficiency level. For example, metacognitive self-regulation was more frequently used in L2 than in L1 in both proficiency groups. More interestingly, lower-proficiency participants used effort regulation more frequently in L2. This may suggest that people had enough cognitive resources available and could use them to select efficient learning strategies even when studying in L2. Studies have highlighted the flexibility of cognitive processes, indicating that individuals can flexibly allocate cognitive resources depending on task demands and situational factors (Broekkamp & Van Hout-Wolters, 2007; Panadero et al., 2019). This adaptability allows individuals to optimize their learning strategies, even in challenging L2 learning contexts. Nevertheless, although participants could devote sufficient cognitive resources so as to unfold metacognitive strategies and correctly monitor their learning, L2 proficiency seems to play a critical role in learning outcomes, as lower proficient participants showed an L2 learning cost despite unfolding monitoring processes correctly.

5. General discussion

The goal of the present study was to explore the consequences of studying texts in L2 on the cognitive and metacognitive processes involved in successful learning, and whether they varied as a function of L2 proficiency. In two experiments, university students were asked to study L1 and L2 texts that differed in difficulty (high and low cohesion). After comprehensively reading each text, they were asked to judge their learning and to answer some questions regarding what they had just studied. Results indicated that studying texts in L2 did not compromise the monitoring of learning since participants—regardless of their proficiency level—were able to use language and cohesion as cues to indicate the difficulty of the texts and judge their learning accordingly. Moreover, data from the learning assessment test validated the pattern observed in the study phase with the JOLs.

Overall, participants judged texts in L2 as more difficult than texts in L1. This language effect in JOLs is consistent with the results of our previous studies involving single words and lists in L1 and L2. Thus, Reyes et al. (2023) also reported that participants were sensitive to linguistic features of the to-be-studied material and predicted better learning in L1 than in L2 when they studied concrete vs. abstract words and lists of words grouped into semantic categories vs. lists of unrelated words. Hence, our new results extend these findings with a more complex set of materials and two different proficiency groups.

Similarly, high-cohesion texts were judged as easier to learn than low-cohesion texts. The effect of cohesion goes in line with previous studies on monolingual text comprehension and learning assessment (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). But the most remarkable pattern here is that even though participants judged texts in L2 as more difficult to learn, this difficulty did not preclude the use of monitoring processes that allowed them to detect difficult material (low-cohesion texts) and to accurately judge it as more challenging to learn. Hence, monitoring processes were not impaired as a consequence of L2 context.

In Experiment 2, the main effects in JOLs were modulated by the interaction between factors. Thus, cohesion interacted with language such that there was no difference in L1 and L2 for high-cohesion texts, yet for low-cohesion texts, the L2 condition received significantly lower JOLs. Results after collapsing across experiments validated this interaction. This suggests that participants found low-cohesion texts in L2 as the most difficult condition among all four. This pattern did not support our initial hypothesis that, under the L2 condition monitoring and regulation processes might be compromised. In fact, the contrast between high and low cohesion in L1 was less pronounced than that observed in L2. Although the result was unexpected, it is conceivable that the presence of a salient cue such as language (L1 vs. L2) could diminish the salience of text cue (cohesion) as a cue, which, in turn, may reduce the perceived learning difficulties (JOLs) within the easier L1 condition. In this line, Magreehan et al. (2016) reported data indicating a reduction of the font type effect on JOLs when other cues were available and thus, we might have expected an attenuation of the cohesion effect in our study when L1-L2 language cues were evident.

Note that a difference between Experiment 1 and 2 was that language was blocked in Experiment 1, whereas it was semirandom in Experiment 2. In fact, the triple interaction between language, proficiency, and type of study (language blocking vs. language mixing) was significant when merging both experiments. The difference between languages in low cohesion texts was only significant when participants studied in a language mixing condition (Experiment 2). Mixing languages in text presentation may have heightened the salience of language as a cue due to the need for language control derived from the switching, at the expense of within-text cues like cohesion. More cognitive resources were devoted to the task because of the necessary language control. Consequently, the cohesion cue was more prominent under the more demanding language condition (L2), which made the interaction between the two factors (cohesion x language) appear. Interestingly, the subsequent learning test also showed a tendency for the cohesion effect to be modulated by language. Thus, the learning outcome was significantly worse in low-cohesion texts than in high-cohesion texts, but only in L2. Hence, it is also possible that the semirandom mix of the languages across the texts may have increased the overall need for regulation, and participants may have engaged in control processes leading them to learn and perceive low-cohesion texts equally easily than high cohesion in L1. In contrast, when learning was performed in the more demanding L2 condition, the engagement of control processes may have not reduced the learning difficulty of the texts which were in turn also perceived as more difficult. Because these explanations are ad hoc, they should be more directly tested. Further research should directly address the consequences of language blocking and mixing in JOLs and memory performance. Overall, regarding our main question of whether L2 texts compromised metacognitive monitoring, results seem to suggest otherwise.

Furthermore, some nuances were found in Experiment 2, between proficiency groups. Interestingly, participants with a higher proficiency level judged L1 and L2 texts equally easy to learn and indeed did not show any sign of L2 cost in the learning assessment test. The lower-proficiency group, though, considered L2 texts as more difficult during the study and actually showed a disadvantage in L2 learning. This pattern of results remains consistent when considering both experiments together. The most remarkable pattern is that language did not impede monitoring under text cohesion manipulations. Low and medium-proficiency participants can still metacognitively monitor the difficulty of the texts, although they cannot compensate for such difficulty. Low and medium proficiency may impact learning, although participants seem to still use metacognitive processes to detect difficulties and regulate their learning. The poorer learning outcomes in L2 for low and medium-proficiency relative to high-proficiency participants go in line with previous research that reported a recall cost in L2 when participants were tested with essay-type questions, presumably due to a lack of writing skills (Vander Beken et al., 2018). We intended to overcome this issue by including open-ended questions that did not require much elaboration, which might have prevented the production deficit. Furthermore, our rubric accepted answers if they included the keywords, regardless of grammatical, syntactic, or orthographic errors. One might claim that both proficiency groups differed in their reading and writing skills, that the lower-proficiency group is composed of unskilled or inexperienced readers in general. However, proficiency groups did not differ in their accuracy for texts in L1, which proves that the differences in the learning assessment test cannot be explained by participants in the higher-proficiency group being better comprehenders. Hence, several factors could explain our lower-proficiency group recall cost in L2: impaired encoding, difficulty in integrating the information, or simply retrieving it from memory.

Note, however, that none of the experiments showed any evidence for compensation. Thus, participants in Experiment 1 did not compensate for the difficulty detected in the study phase to achieve comparable learning in L2 (in contrast to L1) and in low-cohesion (in contrast to high-cohesion) texts. Similarly, in Experiment 2, lower-proficiency individuals did not show compensation effects despite they also detected the difficulty of the low-cohesion texts. The only condition where learning for low-cohesion and high-cohesion texts was comparable was when higher-proficiency individuals studied in L1. However, the effect of cohesion was also absent in JOLs for these individuals, suggesting that for higher-proficiency individuals, the lack of coherence effects in learning outcomes was not due to compensation but to the fact that low- and high-cohesion texts produced similar levels of difficulty for them. As mentioned, this might be due to the greater engagement of control processes when L1 and L2 are presented in a mixed format which may induce better learning in L1. Koriat et al. (2006) proposed that the relationship between monitoring and control processes arises from the fact that metacognitive judgments are based

on feedback from the outcome of control operations. Monitoring does not occur prior to the controlled action, but rather, it takes place afterward. According to this hypothesis, the difficulty of an item is monitored ad hoc: learners allocate the appropriate resources to an item based on its demands, and they recognize that a specific item will be challenging to remember when they realize that it requires a relatively higher level of effort to commit to memory. Thus, although the initial assessment of a situation provides valuable information for executing control actions, the feedback obtained from these actions can subsequently be used as a foundation for monitoring. This monitoring process, in turn, can guide future control operations, creating a cyclical relationship between monitoring and control. In other words, subjective experience informs the initiation and self-regulation of control operation that may in turn change subjective experience.

The lack of evidence for compensation agrees with previous studies indicating that participants do not fully compensate for item difficulty effects through self-regulation. Although one would expect that if learners detect difficulties in the learning materials, they would compensate for this by allocating more time or by selecting a different strategy to better learn this information, there is strong evidence that neither self-paced study (how people allocate their study time), item selection for re-study, or the use of strategies (e.g., distributed practice, retrieval practice) completely compensate for difficulty (Cull & Zechmeister, 1994; Koriat, 2008; Koriat et al., 2006; Koriat & Ma'ayan, 2005; Le Ny et al., 1972; Mazzoni et al., 1990; Mazzoni & Cornoldi, 1993; Nelson & Leonesio, 1988; Pelegrina et al., 2000; see Tekin, 2022 for a review).

The fact that the cohesion effects in learning were evident in most conditions of Experiments 1 and 2 does not discard the possibility that learning is influenced by the language of study. First, medium-low-proficiency participants showed poorer levels of learning in their L2 than in their L1. Second, results from the qualitative questionnaire suggested that participants engaged in different learning strategies when studying in L1 and L2. This might support the idea that they confront L2 learning with different learning strategies. Overall, participants relied on deep-level strategies more in L1 than in L2. However, metacognitive self-regulation was more frequently used in L2 than in L1 in both proficiency groups in Experiment 2. This may suggest that participants had enough cognitive resources available and could use them to select efficient learning strategies even when studying in L2. On the other hand, the selection of some strategies might require extra study time, which was not possible under the time constraint of our experiment. Thus, participants might have used deep-level strategies more frequently in L1 because they might have needed longer study time allocation to use them in L2 learning (Stoff & Eagle, 1971). Nevertheless, the qualitative language difference in the use of metacognitive strategies is an aspect to further explore in the future.

In sum, self-regulated learning is a complex process that appears not to be hindered by L2 processing. So far, we have explored the monitoring process and the short-term consequences in memory and learning derived from studying in L2. As self-regulated learning unfolds, monitoring and control processes arise to evaluate and adjust one's attention, understanding, and behavior. Participants estimate task difficulty, make learning judgments, and allocate resources accordingly, adjusting the pace of learning or regulating strategies use, for instance (Panadero, 2017). Apparently, our sample monitored the to-be-studied material correctly. However, medium-low participants were not able to compensate for the self-perceived difficulty in achieving successful learning in L2, as evidenced by the accuracy of the learning assessment test. This suggests that while L2 learners can monitor effectively, their control processes might be compromised, particularly among those with lower proficiency. Apart from being cognitively demanding, reading and learning in L2 involves a range of factors intertwined with the learner's L2 proficiency, such as the knowledge of morphosyntax and vocabulary, especially academic vocabulary (Silva & Otwinowska, 2019). Thus, our results raise a note of caution for L2 learning in academic settings. Proficiency level seems to introduce some nuances in the learning processes, which may impact the resources available for metacognitive processing while performing school-related activities. Future research should identify the direct and indirect effects of language proficiency and self-regulated learning strategies on academic outcomes in bilingual education.

In sum, self-regulated learning is a complex process that appears not to be hindered by L2 processing. So far, we have explored the monitoring process and the short-term consequences in memory and learning derived from studying in L2. As self-regulated learning unfolds, monitoring and control processes arise so as to evaluate and adjust one's attention, understanding, and behavior. Participants estimate task difficulty, make learning judgments, and allocate resources in accordance, adjusting the pace of learning or regulating strategies use, for instance (Panadero, 2017). Apparently, our sample monitored the to-be-studied material correctly. However, medium-low participants were not able to compensate for the self-perceived difficulty in achieving successful learning in L2 as proven by the accuracy of the learning assessment test. Thus, our results raise a note of caution for L2 learning in academic settings. Proficiency level seem to introduce some nuances in the learning processes which may impact the resources available for metacognitive processing while performing school-related activities. Future research should identify the direct and indirect effects of language proficiency and self-regulated learning strategies on academic outcomes in bilingual education.

Moreover, there are other crucial factors that might also play a role in the self-regulation of learning (Lonka et al., 2004). Text difficulty and language proficiency might not be the only aspects that guide self-regulated learning: executive functions, motivation, self-efficacy, and other task or situational demands can also influence the learning process. A framework that incorporates all these metacognitive and motivational components of self-regulation into account would be more suitable to explain the effects of learning in L2 on monitoring and retention. Overall, our study underscores the importance of considering factors such as L2 proficiency that affect self-regulated learning. It also highlights the need to accommodate these complex dynamics between monitoring and control strategies to enhance academic success in bilingual settings.

References

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

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, **80**(2), 207–245. https://doi.org/10.3102/0034654310368803

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, **53**(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, **52**(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2020). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09556-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv, arXiv:1406.*

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(6), 1022–1047. https://doi.org/10.1037/xlm0000770

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, **25**(1), 65–83.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, **143**, 233–262. https://doi.org/10.1037/bul0000099

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, **16**(4), 240–249. https://doi.org/10.1016/j.tics.2012.03.001

Broekkamp, H., & Van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, **19**(4), 401–428. https://doi.org/10.1007/s10648-006-9025-0

Brookshear, J. G. (2005). *Computer science. An overview*. (9th ed.). Pearson Internation Edition.

Byun, K., Chu, H., Kim, M., Park, I., Kim, S., & Jung, J. (2011). English-medium teaching in Korean higher education: Policy debates and reality. *Higher Education*, **62**(4), 431–449. https://doi.org/10.1007/s10734-010-9397-4

Cano, F. (2007). Approaches to learning and study orchestrations in high school students. *European Journal of Psychology of Education*, **22**(2), 131–151.

Carroll, M., & Korukina, S. (1999). The effect of text coherence and modality on metamemory judgements. *Memory*, **7**(3), 309–322. https://doi.org/10.1080/096582199387940

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, **19**(1), 5–51. https://doi.org/10.1177/1529100618772271

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, **1665**, 50–64. https://doi.org/10.1016/j.brainres.2017.03.023.Neural

Collins, V. L., Dickson, S. V., Simmons, D. C., & Kameenui, E. J. (1996). Metacognition and Its Relation to Reading Comprehension: A Synthesis of the Research. (Tech. Rep. No. 23).

Conway, A. R. A., Kane, M. J., & Al, C. E. T. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**(5), 769–786. https://doi.org/10.1021/cen-v020n021.p1399

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, **32**, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language*, **26**(1), 92–113.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition*, **22**(2), 249–257.

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

Deekens, V. M., Greene, J. A., & Lobczowski, N. G. (2018). Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology*, **88**(1), 63–79. https://doi.org/10.1111/bjep.12174

Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, **55**(0), 371–397. https://doi.org/10.1002/rrq.277

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

Dufort, B., Erickson, S., Hamilton, M., Soderquist, D., & Zigray, S. (2018). *World geography*. (version 1.). Michigan Open Book Project. https://doi.org/10.4135/9781412972024.n1099

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed.), *Psychology of learning and motivation* (pp. 103–140). Elsevier.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**(2), 175–191. https://doi.org/10.3758/BF03193146

Follmer, D. J., & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences*, **67**(September 2017), 177–187. https://doi.org/10.1016/j.lindif.2018.08.005

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

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, **8**(1), 1–18. https://doi.org/10.1007/s11409-012-9093-0

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best Practice suggestions for cognitive psychology and neuroscience. *Swiss Psychology Open*, **3**(1), 1. https://doi.org/10.5334/spo.34

Gasparinatou, A., & Grigoriadou, M. (2013). Exploring the effect of background knowledge and text cohesion on learning from texts in computer science. *Educational Psychology*, **33**(6), 645–670. https://doi.org/10.1080/01443410.2013.790309

Grant, R., Parker, P., Coward, F., Cussans, T., Regan, S., Levy, J., & Wilkinson, P. (2016). *The history book. Big ideas simply explained*. A. Beeden & S. Kennedy (Eds.). DK Publishing.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Halamish, V. (2018). Can very small font size enhance memory? *Memory and Cognition*, **46**(6), 979–993. https://doi.org/10.3758/s13421-018-0816-6

Hall, S. S., Maltby, J., Filik, R., & Paterson, K. B. (2016). Key skills for science learning: The importance of text cohesion and reading ability. *Educational Psychology*, **36**(2), 191–215. https://doi.org/10.1080/01443410.2014.926313

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Elvesier Science Publisher B. V. https://doi.org/10.1007/s10749-010-0111-6

Hessel, A. K., & Schroeder, S. (2020). Interactions between lower- and higher-level processing when reading in a second language: An eye-tracking study. *Discourse Processes*, **57**(10), 940–964. https://doi.org/10.1080/0163853X.2020.1833673

Hessel, A. K., & Schroeder, S. (2022). Word processing difficulty and executive control interactively shape comprehension monitoring in a second language: An eye-tracking study. *Reading and Writing* 35, 2287–2312 (2022). https://doi.org/10.1007/s11145-022-10269-3

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

Hosoda, M. (2014). Generating causal bridging inferences in EFL expository reading: Combining on-line and off-line processing. *ARELE (Annual Review of English Language Education in Japan)*, 25, 239–254.

Hosoda, M. (2017). Learning from Expository Text in L2 reading: Memory for causal relations and L2 reading proficiency. *Reading in a Foreign Language*, 29(2), 245–263.

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory and Cognition*, 45(1), 121–136. https://doi.org/10.3758/s13421-016-0646-3

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory*, 24(3), 383–398. https://doi.org/10.1080/09658211.2015.1009470

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, 4(2), 161–176. https://doi.org/10.1007/s11409-009-9042-8

Iniesta, A., Paolieri, D., Serrano, F., & Bajo, M. T. (2021). Bilingual writing coactivation: Lexical and sublexical processing in a word dictation task. *Bilingualism: Language and Cognition*, 24, 1–16. https://doi.org/10.1017/S1366728921000274

Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System*, 70, 107–120. https://doi.org/10.1016/j.system.2017.07.007

Jung, J. (2018). Effects of task complexity and working memory capacity on L2 reading comprehension. *System*, 74, 21–37. https://doi.org/10.1016/j.system.2018.02.005

Kaan, E., Kheder, S., Kreidler, A., Tomíc, A., & Valdés Kroff, J. R. (2020). Processing code-switches in the presence of others: An ERP study. *Frontiers in Psychology*, 11, 1288. https://doi.org/10.3389/fpsyg.2020.01288

Koriat, A. (1997). Monitoring One's own knowledge during Study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, 36(2), 416–428.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52(4), 478–492.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Krebs, S. S., & Roebers, C. M. (2012). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition and Learning*, 7(2), 75–90. https://doi.org/10.1007/s11409-011-9079-3

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a Tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, 23(3), 159–163. https://doi.org/10.1177/0963721414528511

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1(1), 377–394. https://doi.org/10.1016/j.physbeh.2017.03.040

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections betwwen bilingual memory representations. *Journal of Memory and Language*, 33, 149–174. https://doi.org/10.1121/1.2934955

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarty in self-paced learning conditions. *Acta Psychologica*, 36(4), 280–289.

Lefèvre, N., & Lories, G. (2004). Text cohesion and metacomprehension: Immediate and delayed judgments. *Memory and Cognition*, 32(8), 1238–1254. https://doi.org/10.3758/BF03206315

Li, X., Chen, G., & Yang, C. (2021). How cognitive conflict affects judgments of learning: Evaluating the contributions of processing fluency and metamemory beliefs. *Memory and Cognition*, 912–922. https://doi.org/10.3758/s13421-021-01143-8

Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and prospects of measuring studying and learning in higher education. *Educational Psychology Review*, 16(4), 301–323. https://doi.org/10.1007/s10648-004-0002-1

López-Rojas, C., Rossi, E., Marful, A., & Bajo, M. T. (2022). Prospective memory in bilinguals and monolinguals: ERP and behavioural correlates of prospective processing in bilinguals. *Brain and Language*, 225. https://doi.org/10.1016/j.bandl.2021.105059

Luk, G., & Kroll, J. F. (2019). Bilingualism and education. Bridging cognitive science research to language learning. In J. Dunlosky, & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 292–319). Cambridge University Press. https://doi.org/10.1017/9781108235631.013

Ma, H., Hu, J., Xi, J., Shen, W., Ge, J., Geng, F., & Yao, D. (2014). Bilingual cognitive control in language switching: An fMRI study of English-Chinese late bilinguals. *PloS One*, 9(9), e106468.

Macaro, E. (2018). *English medium instruction*. Oxford University Press.

Macizo, P., Bajo, M. T., & Cruz Martín, M. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish-English interlexical homographs. *Journal of Memory and Language*, 63(2), 232–244. https://doi.org/10.1016/j.jml.2010.04.002

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, 11(1), 35–56. https://doi.org/10.1007/s11409-015-9147-1

Marcouse, I., Anderson, P., Black, A., Machin, D., & Watson, N. (2014). The business book. In S. Atkinson (Ed.), *Big ideas simply explained*. New York: DK publishing.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50(4), 940. https://doi.org/10.1044/1092-4388(2007/067)

Matvey, G., Dunlosky, J., & Schwartz, B. L. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, 14(2), 253–261. https://doi.org/10.1080/09658210500216844

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study item allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, 122, 47–60.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, 18(2), 196–204.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18, 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174–179.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, 23(6), 564–579. https://doi.org/10.1016/j.jneuroling.2010.05.002.Conflict

Nassaji, H. (2011). Issues in second-language reading: Implications for acquisition and instruction. *Reading Research Quarterly*, 46(2), 173–184. https://doi.org/10.1598/RRQ.46.2.5

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109–133. https://doi.org/10.1037/0033-2909.95.1.109.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**(4), 676–686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation*. (pp. 125–141). Academic Press.

Nguyen, K., & McDaniel, M. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, **22**(1), 59–71. https://doi.org/10.1037/xap0000066

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2014). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, **47**(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Ozuru, Y., Kurby, C. A., & McNamara, D. S. (2012). The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy. *Metacognition and Learning*, **7**(2), 113–131. https://doi.org/10.1007/s11409-012-9087-y

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, **8**(422), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: Examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*, **44**(3), 379–397. https://doi.org/10.1080/02602938.2018.1512553

Pelegrina, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory*, **8**(6), 377–392. https://doi.org/10.1080/09658210050156831

Pérez, A. I., Hansen, L., & Bajo, M. 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, 1–19. https://doi.org/10.1017/S1366728918000846

Pérez, A. I., Schmidt, E., & Tsimpli, I. (2023). *Inferential evaluation and revision in L1 and L2 text comprehension: An eye movement study*. Bilingualism Language and Cognition. 1–14.

Pessoa, S., Miller, R. T., & Kaufer, D. (2014). Students' challenges and development in the transition to academic writing at an English-medium university in Qatar. *IRAL, International Review of Applied Linguistics in Language Teaching*, **52**(2), 127–156. https://doi.org/10.1515/iral-2014-0006

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency - Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, **44**, 31–40. https://doi.org/10.1016/j.learninstruc.2016.01.012

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A Manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Office of Educational Research and Improvement. National Center for Research to Improve Postsecondary Teaching and Learning. https://doi.org/10.5901/mjss.2015.v6n1p156

Pintrich, P. R., & Zusho, A. (2007). Motivation and self-regulated learning in the college classroom. In R. Perry, & J. Smart (Eds.), *Handbook on teaching and learning in higher education*. Springer Publishers.

Prinz, A., Golke, S., & Wittwer, J. (2020). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review*, **31**(August). https://doi.org/10.1016/j.edurev.2020.100358

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, **28**(1), 69–80.

Reyes, M., Morales, M. J., & Bajo, M. T. (2023). Judgments of learning in bilinguals: Does studying in a L2 hinder learning monitoring? *PLOS ONE*, **18**(12), e0286516. https://doi.org/10.1371/journal.pone.0286516

Sandoval, T., Gollan, T., Ferreira, V., & Salmon, D. (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

Schroeder, J., Harvey, D., Coyne, C., Jackson, T., Farndon, J., Singer, A., & Harris, T. (2019). *The ecology book. Big ideas simply explained*. H. Fewster & C. Hallinan (Eds.). DK Publishing. https://doi.org/10.1080/00988157.1974.9977066

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review*, **24**(March), 116–129. https://doi.org/10.1016/j.edurev.2018.03.004

Sidek, H. M., & Rahim, H. A. (2015). The role of vocabulary knowledge in reading comprehension: A cross-linguistic study. *Procedia—Social and Behavioral Sciences*, **197**, 50–56. https://doi.org/10.1016/j.sbspro.2015.07.046

Silva, B. B., & Otwinowska, A. (2019). VST as a reliable academic placement tool despite cognate inflation effects. *English for Specific Purposes*, **54**, 35–49. https://doi.org/10.1016/j.esp.2018.12.001

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., MacEdo, A. F., Ferré, P., Acuña-Fariña, C., Hernández-Cabrera, J., & Fraga, I. (2019). Lexico-syntactic interactions during the processing of temporally ambiguous L2 relative clauses: An eye-tracking study with intermediate and advanced Portuguese-English bilinguals. *PLoS ONE*, **14**(5), 1–27. https://doi.org/10.1371/journal.pone.0216779

Soemer, A., & Schiefele, U. (2019). Text difficulty, topic interest, and mind wandering during reading. *Learning and Instruction*, **61**(December 2018), 12–22. https://doi.org/10.1016/j.learninstruc.2018.12.006

Stoff, D. M., & Eagle, M. N. (1971). The relationship among reported strategies, presentation rate, and verbal ability and their effects on free recall learning. *Journal of Experimental Psychology*, **87**(3), 423–428.

Tekin, E. (2022). Can learners allocate their study time effectively? It is complicated. *Educational Psychology Review*, **34**(2). https://doi.org/10.1007/s10648-021-09645-2

Thiede, K. W., Anderson, M. C. M., & Herriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, **95**, 66–73.

Thorpe, C., Todd, M., Yuill, C., Tomley, S., Hobbs, M., & Weeks, M. (2015). *The Sociology Book. Big ideas simply explained*. S. Atkinson (Ed.). DK Publishing.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, **28**, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, **73**(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory and Cognition*, **43**(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory and Cognition*, **46**(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, **92**, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

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., De Bruyne, E., & Brysbaert, M. (2020). Studying texts in a non-native language: A further investigation of factors involved in the L2 recall cost. *Quarterly Journal of Experimental Psychology*, **73**(6), 891–907. https://doi.org/10.1177/1747021820910694

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

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, **16**(4), 359–384. https://doi.org/10.1007/s10648-004-0005-y

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–32). Routledge.

Wirth, J., Stebner, F., Trypke, M., Schuster, C., & Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load. *Educational Psychology Review*, **32**(4), 1127–1149. https://doi.org/10.1007/s10648-020-09568-4

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory and Cognition*, **45**(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, **45**(1), 166–183. https://doi.org/10.3102/0002831207312909

Zimmerman, B. J., & Kitsantas, A. (2005). The hidden dimension of personal competence: Self-regulated learning and practice. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation.* (pp. 509–526). Guilford Press.

Zusho, A. (2017). Toward an integrated model of student learning in the college classroom. *Educational Psychology Review*, **29**, 301–324. https://doi.org/10.1007/s10648-017-9408-4

Appendix 1. Example of a high-cohesion and a low-cohesion English version

Title: Traits of mammals
High-cohesion version
Over many years[1], mammals have developed different kinds of specialized teeth, which has enabled them to be successful. These different teeth[2] allow mammals[2] to eat many different kinds of food. This trait also helps them to live in different kinds of environments. Basically[1], there are four types of teeth in mammals[2]: incisors, canines, premolars, and molars. The number and shape of each of these types of teeth[1] are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. Mammals such as we, humans, have many different kinds of teeth, which help us eat the many different kinds of food in their diets.
Low-cohesion version
Mammals have very specialized teeth that have made[3] them successful and allowed them to eat many different kinds of food in different kinds of environments. For the four types of teeth, there are, the number and shape of each of them are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth that are used for tearing. Their[4] incisors are chisel-shaped and are used for cutting. Those[4] that are plant-eating, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. We humans have many different kinds of teeth as we eat many different kinds of food in the diet.
Open-ended questions
What type of teeth do meat-eating mammals have? Long and pointed canine teeth.   What type of teeth do plant-eating mammals have? Large and flat premolars and molars   What does having different kinds of specialized teeth allow mammals? These different teeth allow mammals to eat many different kinds of food and to live in different kinds of environments.

Note: 1 = noun repetition, 2 = use of connectors, 3 = use of passive voice, 4 = pronoun replacement to create ambiguity.

Appendix 2. Texts and references they were extracted from

	Text	References
1	Networking and the Internet	Extracted from Brookshear (2005), a book used by Gasparinatou and Grigoriadou (2013).
2	Security – Forms of attack	
3	Artificial Intelligence	
4	Geologic Processes	Adapted from Ariel, Karpicke, Witherby, and Tauber (2020).
5	Inorganic Substances	
6	Crystalline Solids	
7	Elements	
8	Compounds	
9	Traits of Mammals	Adapted from McNamara, Kintsch, Songer, and Kintsch (1996).
10	Heart Disease	
11	Heat Distribution in Animals	Adapted from Ozuru, Kurby, & McNamara (2012).
12	Africa	Extracted from the interactive book *World Geography. Michigan Open Project* (Dufort et al., 2018).
13	Apartheid	
14	Antarctica	
15	The global financial crisis	Extracted from *The History Book. Big ideas simply explained* (Grant et al., 2016)

(Continued)

(*Continued*)

	Text	References
16	Global issues, local perspectives	Extracted from *The Sociology Book. Big ideas simply explained* (Thorpe et al., 2015)
17	Climate change	
18	Environmental pollution	Extracted from *The Ecology Book. Big ideas simply explained* (Schroeder et al., 2019).
19	Effects of pollution on health	
20	Emotional Intelligence	Extracted from *The Business Book. Big ideas are simply explained* (Marcouse et al., 2014).
E1	Supermarkets	Extracted from an FCE reading test.
E2	Charlie Chaplin	Extracted from the book *The History Book. Big ideas simply explained* (Grant et al., 2016)

Appendix 3. English version of the customized metacognitive questionnaire. Note that the learning strategy each item refers to appears in brackets

1. I memorized and repeated keywords to remind me of important concepts in the texts. (Rehearsal)
2. I tried to understand the material in the texts by making connections between the concepts. (Elaboration)
3. When studying the information in the texts, I found myself questioning things to decide if I find them convincing. (Critical thinking)
4. When studying the materials in the texts, I often missed important points because I was thinking of other things. (R) (Metacognitive self-regulation)
5. When the content was difficult, I gave up or only paid attention to the easy parts. (R) (Effort regulation)
6. How mentally demanding was the task? (Mental demand)
7. How successful were you in accomplishing what you were asked to do? (Performance)
8. How hard did you have to work to accomplish your level of performance? (Effort)