



# Word processing difficulty and executive control interactively shape comprehension monitoring in a second language: an eye-tracking study

Annina K. Hessel<sup>1,2</sup> · Sascha Schroeder<sup>1</sup>

Accepted: 4 February 2022 / Published online: 4 March 2022  
© The Author(s) 2022

## Abstract

Successful reading comprehension—especially in a second language (L2)—relies on the ability to monitor one’s comprehension, that is, to notice comprehension breaks and make repairs. Comprehension monitoring may be limited given effortful word processing but may also be supported through active reading. The current study addresses to what extent word processing difficulty reduces adolescents’ ability to monitor their comprehension in their L2, and whether readers can compensate limitations given sufficient executive control. We conducted an eye-tracking experiment in which 34 adolescent L2 learners (aged 13–17 years) read short expository texts containing two within-subject manipulations. First, comprehension monitoring was tested through inconsistencies, for example, when the topic changed from *Spanish* to *Russian* vis-à-vis consistent controls. Second, word processing difficulty was altered by inserting either shorter and higher-frequency words such as *want*, or longer and lower-frequency words such as *prefer*. We additionally measured participants’ executive control. Outcome variables were reading times on the whole texts and the words manipulated for inconsistency and word processing difficulty. We found evidence of successful moment-to-moment monitoring, as visible in adolescents’ increased rereading of inconsistent compared to consistent information. We also found that adolescents adapted their monitoring differently to word processing difficulty, depending on their executive control: while adolescents with weaker control reduced their monitoring given higher word processing difficulty, adolescents with stronger control monitored their comprehension more (instead of less) on difficult texts. These findings provide insights into how L2 comprehension monitoring arises in the interplay of lower-level processing load and active reading processes.

**Keywords** Reading · Second language · Eye movements · Comprehension monitoring · Executive control · Inconsistency

---

✉ Annina K. Hessel  
a.hessel@posteo.de

Extended author information available on the last page of the article

## Introduction

When we read, a lot needs to happen for us to grasp what a text is saying: from understanding words, to drawing connections, to monitoring when comprehension goes astray (Castles et al., 2018). For most of us who are reading this specific manuscript, all these things will need to happen in our second language (L2), making this task even more remarkable. And yet, some lower-level processes—such as word reading—are so automatic that skilled readers can even not stop them in their L2, as famously demonstrated in the Stroop task (Braet et al., 2011; Jensen & Rohwer, 1966). Even complex reading processes may arise without additional effort, such as monitoring one's comprehension—but only as long as high-quality word knowledge allows smooth word processing, as laid out by the Reading Systems Framework (Stafura & Perfetti, 2014). However, most theories agree that reading requires not only automatic, but also active and resource-demanding processes that may particularly matter when re-analysing coherence breaks (O'Brien & Cook, 2016; Singer et al., 1994; van den Broek & Helder, 2017). In contrast to the Reading Systems Framework, these latter theories would then predict increased active processing in the face of difficulties (see for instance van den Broek & Helder, 2017), rather than a reduction of reading processes under high load (as it follows from Stafura & Perfetti, 2014). Just as word knowledge lies at the heart of efficient word processing, strong executive functions lie at the heart of active processing (Kaakinen et al., 2003; van Moort et al., 2018). Additionally, strong executive control may also allow readers to compensate for less than smooth word processing (Hamilton et al., 2016). In the current study, we therefore ask whether executive control and word processing difficulty interact in their influence on adolescents' comprehension monitoring. This question is particularly relevant for adolescents' L2 learners for whom online monitoring is a strong predictor of their overall comprehension (Mulder et al., 2021). Additionally, word processing tends to be more difficult when reading in one's L2 vis-à-vis L1 (Cop et al., 2015), which impacts sentence and higher-level processing (Hopp, 2017; Lim & Christianson, 2014; Pérez & Bajo, 2019; van den Bosch et al., 2018; Whitford & Titone, 2012). In this context, strong executive control abilities have been found to be more influential, arguably because they allow to compensate difficulties (Prehn et al., 2018; Raudszus et al., 2017). It is thus pivotal to understand how adolescents' (here, aged 13–17 years) comprehension monitoring emerges in this interplay of word processing difficulty and executive control when reading in the L2.

Readers engage in comprehension monitoring whenever they notice and correct mismatches between what they currently read and what they have read before or know to be true. Comprehension monitoring has frequently been studied *offline*, that is, based on readers' explicit reflections on a text. Typically, this is done by inserting inconsistent information—be it a syntactic mismatch between a pronoun and a verb, or a semantic mismatch between sentences—and ask participants to verbalize or underline any inconsistencies they notice (Baker, 1984; Kim & Phillips, 2014; Oakhill et al., 2005). These offline studies have demonstrated a

close link between comprehension monitoring and general reading comprehension (Kim & Phillips, 2014): older children (e.g., 9 to 11-year-olds) and skilled comprehenders monitor their comprehension more skilfully (that is, they noticed and identified more and also harder-to-detect types of errors) compared to younger children (e.g., 5 to 7-year-olds) and less-skilled comprehenders (Baker, 1984; Oakhill et al., 2005; Zabrocky & Ratner, 1989).

Offline studies alone cannot, however, reveal how comprehension monitoring proceeds without prompting, nor are they free of secondary tasks that require metacognitive and linguistic efforts. To study comprehension monitoring directly while it happens, we need to look at the reading process itself using online methods. A prime method for this is eye-tracking, or the recording of readers' eye movements on the text. Eye-tracking captures reading behaviour while leaving readers free to roam the text with no secondary tasks. In the eye movement record, comprehension monitoring shows in increased rereading of inconsistent compared to consistent information, a sign that readers notice and re-analyse the inconsistent information (Connor et al., 2015; Hessel et al., 2021; Joseph et al., 2008; Rayner et al., 2004). This increased rereading is important for comprehension: more pronounced slow-downs on inconsistencies are associated with stronger reading comprehension (Mulder et al., 2021) and comprehension suffers when rereading is made impossible (Schotter et al., 2014). Thus, rereading of inconsistencies provides an excellent window into ongoing monitoring.

A key question is what readers need to successfully monitor their comprehension or indeed, what may be lacking when they fail to do so. A prominent theoretical position is provided by the Reading Systems Framework. It suggests that higher-level comprehension rests primarily on the efficiency of word processing, which in turn depends on high-quality word knowledge (Perfetti, 1985; Perfetti & Stafura, 2014). The prediction is that when word meanings are unclear or slow to retrieve, drawing higher-level connections becomes difficult to impossible. Support comes from eye movement studies that find that stronger vocabulary and comprehension skills go hand in hand with children's and adolescents' more pronounced monitoring, as visible in increased and targeted rereading of inconsistent compared to consistent information (Connor et al., 2015; Eilers et al., 2018; Hessel et al., 2021). Thus, eye movement evidence supports the idea that comprehension monitoring relies on efficient word processing.

However, individual difference studies can only indicate but never directly pinpoint the source of reading differences as more and less skilled readers always vary on multiple dimensions (Castles et al., 2018). For this reason, experimental manipulations of word processing efficiency have been used to pinpoint the source of differences between L1 and L2 readers (Hopp, 2016; McDonald, 2006) or more or less successful comprehension monitoring (Hessel & Schroeder, 2020). In these studies, readers are presented with texts where the phenomenon of interest—for example a certain syntactic structure or semantic inconsistency—is surrounded by either easy (that is, short and frequent) or difficult (that is, longer and less frequent) words. Both lowering a word's frequency and increasing its lengths are known to increase word and text processing difficulties (Inhoff & Rayner, 1986; Joseph et al., 2013; Tiffin-Richards & Schroeder, 2015). Of interest is how syntactic processing or monitoring

change in response to the variations in word processing difficulty. Using such an experimental design, adult L2 learners have been found to indeed reduce their monitoring under higher word processing difficulty (Hessel & Schroeder, 2020), supporting the idea that effortful word processing directly impacts higher-level monitoring. So far, however, this experimental evidence is limited to university students in the lab who differ from L2 learners at school in both age and language experience. The current study takes the same experimental paradigm into the field by investigating how word processing influences comprehension monitoring amongst younger, adolescent L2 readers.

Word processing difficulty is, however, not the only variable of interest when studying comprehension monitoring. Most reading theories agree that beyond the passive spread of activation from word knowledge to textual coherence (Perfetti & Stafura, 2014), reading at times also requires more active, reader-initiated processing (O'Brien & Cook, 2016; Singer et al., 1994; van den Broek & Helder, 2017). While passive processes such as word recognition are thought to be mostly automatic and effortless, reader-initiated processes require executive control in order to store information (Just & Carpenter, 1980), bind word meanings into a coherent whole (Hagoort, 2005), and focus attention on relevant information (Engle, 2002). This link is supported by studies that associations between stronger executive control and better text (but not word) comprehension (Sesma et al., 2009), more elaborative inference-making (Calvo, 2001), better metacognitive monitoring on difficult texts (Ikeda & Kitagami, 2013), and reading tailored to text structure (Hyönä et al., 2002) or reading purpose (Kaakinen et al., 2003; Linderholm & Van den Broek, 2002). In the context of the current study, we define executive control as domain-independent cognitive resource that allows people to not only store and rehearse information, but to also focus their attention on task-relevant information (Engle & Kane, 2003). In line with this definition, we chose the automated operation span task as our measure of executive control, as it is a complex task that does not involve reading comprehension (as opposed to the reading span task for example; Engle & Kane, 2003).

When it comes to online monitoring, however, the exact influence of executive control is hard to predict. While one study found that readers with stronger executive control adapted their reading *less* to inconsistencies compared to those with weaker control—which the authors interpreted as more efficient re-analysis (van Moort et al., 2018)—another study reports that given momentarily stronger control capacities (thanks to the absence of secondary tasks), readers adapted their reading *more* to inconsistencies compared to when taxed by secondary tasks—which the authors interpreted as more thorough re-analysis (de Bruïne et al., 2021). That is, stronger executive control was linked to either more or less pronounced monitoring across studies. Beyond differences in study design (individual differences vs. experimental manipulation), a reason for the differing results could be that active, reader-initiated processes may only kick in when comprehension falls below a reader's expectations (van den Broek & Helder, 2017). If this was true, stronger executive control would only lead to increased monitoring when a reader judges their comprehension to be too low. Support for this notion comes from studies that find that stronger executive control is particularly beneficial when comprehension is impeded due to weaker L2 abilities (Prehn et al., 2018), slower decoding (Hamilton et al., 2016), or higher

text difficulty (Ikeda & Kitagami, 2013). Note that these findings are in line with the Reading Framework which predicts a switch from passive to active processing in the face of difficulties (van den Broek & Helder, 2017). In the current study, we therefore ask how adolescents comprehension monitoring in the L2 emerges in the interplay of word processing difficulty and executive control.

## The present study

We wanted to know whether word processing difficulty and executive control would interactively influence adolescent readers' comprehension monitoring in their L2. In our study, German adolescents read short expository texts in English while their eye movements were monitored. Texts contained an inconsistency manipulation (consistent vs. inconsistent information) crossed with a manipulation of word processing difficulty (difficult vs. easy context words). The inconsistency manipulation served to tap comprehension monitoring (by observing the main effect of reading inconsistent words for longer than consistent words), while the word processing difficulty manipulation served to see whether word processing difficulty (as visible in longer reading of words and texts with higher word processing difficulty) would reduce such signs of comprehension monitoring. We made the following predictions: we expected that comprehension monitoring would show in longer and increased (re) reading of inconsistent over consistent information. We further expected adolescent readers to monitor their comprehension less strongly under higher word processing difficulty (which would show in reduced inconsistency rereading), in line with the idea that more effortful passive word processing limits higher-level monitoring (Perfetti & Stafura, 2014). Moving on to the exploratory part of our study, we also measured participants' individual differences in executive control. Given the complexity of previous evidence (de Bruine et al., 2021; van Moort et al., 2018), we took a more explorative approach to the role of executive control by merely asking if online comprehension monitoring was additionally influenced by the readers' executive control, which would indicate the involvement of active, reader-initiated processing (van den Broek & Helder, 2017).

## Methods

### Participants

Thirty-nine adolescent L2 learners were recruited from 9 and 10th grade at two German secondary schools.<sup>1</sup> We secured consent from their legal guardian and the participants themselves. Two participants were excluded because their German language skills were not sufficient to understand the German-language

---

<sup>1</sup> In the first of the two participating schools,  $n=20$  and  $n=8$  adolescents were recruited from the 9th and 10th grade, respectively ( $n=28$ ). In the second school, we recruited  $n=3$  adolescents each from the 9th and 10th grade ( $n=6$ ).

**Table 1** English language learning abilities in our sample

	Mean (SD)	Range	Age equivalent
Single word reading	121 (14)	85–149	12;08
Vocabulary (overall)	47 (11)	18–71	10;10 <sup>a</sup>
Synonyms sub-score	22 (6)	5–31	
Antonyms sub-score	25 (6)	13–34	

<sup>a</sup>The age equivalent for vocabulary was determined by doubling the scores from the two administered sub-tests

task instructions of our experiment. Another participant was excluded for having learned English as their family language, as confirmed through our language questionnaire. Three further participants were unable to finish the experiment, two due to technical problems and one due to a developmental reading difficulty. The final sample thus comprised 34 participants ( $n=20$  female, mean age of 14.5 years,  $SD=0.8$ ). All participants had normal or corrected-to-normal vision.

In order to describe our sample, we asked participants about their language learning history in a language background questionnaire (all materials are available on the Open Science Framework (OSF), [https://osf.io/7jdpu/?view\\_only=c7ca336f2bcb4661a3a07ad5ccc695fb](https://osf.io/7jdpu/?view_only=c7ca336f2bcb4661a3a07ad5ccc695fb)). All participants had grown up learning German and additionally, 11 children had grown up learning one or two non-German home languages (two each with Russian, Croatian, and Portuguese, and one each with Arabic, Indonesian, Rumanian, Polish, and Vietnamese) and had started learning German by the age of 2.0 years ( $SD=3.3$ ) on average. As to English, all but one participant had learned the language since primary school and all of them learned English at secondary school. Roughly half of the sample ( $n=17$ ) were receiving English-medium instruction in subjects such as biology or science and four had had additional English instruction through voluntary school clubs or language study trips. Participants also reported considerable exposure to English in their free-time: the three most common activities were playing English-language video games (with  $n=32$  playing daily and  $n=24$  mostly in English), engaging with English-language social media (with  $n=29$  using social media daily and  $n=14$  mostly in English), or watching English-language videos (with  $n=26$  watching daily and  $n=10$  mostly in English).

To be able to describe our sample, we also assessed our participants' English-language abilities using standardised tests of vocabulary and word reading (see [materials](#) section for a description of tasks). We chose vocabulary and word reading to tap our participants' L2 proficiency as both measures are widely used as such since they tap important facets of L2 comprehension (cf. de Cat, 2020; Lervåg & Aukrust, 2010), namely, knowledge of the written form and of word meanings (Torgesen et al., 2012; Wiig & Secord, 1992). On average, participants in our sample had word reading skills and vocabulary knowledge comparable to that of 13- or 10-year-old L1 learners, respectively (see Table 1).

**Table 2** Example story and question

In the US, people speak many different languages, which can make/render using hospitals difficult/challenging. In some states, many people speak only *Spanish/Russian*. These states have therefore written/prepared some information in hospitals in *Spanish* to help people communicate. But/Regrettably, some Americans would want/prefer to see only English in hospitals. Everyone would need to learn English or miss out.

Local comprehension question:

Some Americans don't like seeing and hearing Spanish in hospitals.

**TRUE**                      **FALSE**

Global comprehension question:

The text was about different opinions on languages in American hospitals.

**TRUE**                      **FALSE**

Word difficult targets are underlined, inconsistency pre-target and target words are in italics, correct answers to comprehension questions are in bold

## Reading task

### Materials

Thirty-six 3- to 5-sentence texts were adapted from a previous study (Hessel & Schroeder, 2020). They had been developed based on short texts from instructional websites (Word Generation, Using English, BBC Bitesize, or TOEFL test preparation materials) and checked by two native speakers of British English for acceptability (Hessel & Schroeder, 2020). To adapt the reading task for use with adolescent readers within a school lesson of 45 min, we shortened the texts (to  $M=69$ ,  $SD=8$  words per story) and reduced their number to 36. Furthermore, we had the texts checked by an experienced German English-language teacher who identified and replaced grammatical constructions and words beyond the language skills of our target population. Finally, we piloted the reading task with two adolescent L2 readers to confirm its suitability. The final texts cover a range of topics, including climate change, education, vegetarian diet, and pets. An example is shown in Table 2 (for a full list of stimuli, see the projects' online repository on the OSF).

Comprehension monitoring was assessed using an inconsistency manipulation which resulted from a combination of either consistent or inconsistent pre-target and target words. To reiterate, successful comprehension monitoring would then be visible in increased (re)reading of inconsistent vis-à-vis consistent target words. Pre-targets were created by replacing key words with a plausible alternative that did however not match the following target word (e.g., *Spanish* was replaced with *Russian*). As pre-targets were plausible, target words were the first inconsistent words in each text (e.g., when *Spanish* appeared in a text that previously talked about *Russian*). Pre-targets were placed at equal distance before targets (12 words and 1 sentence) and matched across conditions in lengths (<2 letters difference) and frequency ( $M=4.7$ ,  $SD=0.67$  Zipf frequency for consistent and  $M=4.8$ ,  $SD=0.6$  for inconsistent pre-targets in the SUBTLEX-UK). Target words were high-frequency nouns or verbs ( $M=4.7$ ,  $SD=0.6$  Zipf frequency in the SUBTLEX-UK; Heuven et al., 2014) of on average 6 ( $SD=1.4$ ) letters length. To avoid interference through

**Table 3** Matching of easy and difficult word processing difficulty manipulated words

	Easy	Difficult	<i>t</i> -test results
Frequency in SUB-TLEX-UK	5.5 (0.8) Range: 3.2–6.9	3.6 (0.7) Range: 1.3–5.0	$t(219.7) = 19.0, p < .0001$
Frequency in BNC	5.4 (0.7) Range: 3.0–6.8	4.0 (0.6) Range: 2.6–5.3	$t(417.5) = 21.6, p < .0001$
Letter lengths	5.1 (1.6)	7.8 (1.8)	$t(201.5) = -12.0, p < .0001$

Values are means with standard deviations in brackets, frequency is Zipf-Frequency (log-frequency + 3)

clause or sentence wrap-up, target words were followed by minimally three words. Eye movement signatures of comprehension monitoring were subsequently only analysed on the target words, allowing us to study monitoring on the same word (here, *Spanish*), beyond any confounds from word characteristics.

Word processing difficulty was manipulated by replacing five words in each text with a less frequent and longer alternative (e.g., by replacing *want* with *prefer*), thereby creating a difficult and an easy version of each text (see Table 3). As a reminder, increased word processing difficulty was expected to be visible in longer reading times on words and texts with higher word processing difficulty, and furthermore, to reduce signs of comprehension monitoring in the eye movement record. Words were spread across texts to create an overall difference in difficulty. Differences in word frequency were confirmed using two English-language corpora that tap main sources of our participants' English-language input, the SUBTLEX-UK (Heuven et al., 2014) that stems from films and series and the British National Corpus that is based on written texts (2007).

We split our texts into four lists following a Latin square design where each participant read all 36 texts once, 18 in the consistent and 18 in the inconsistent condition. Inconsistency was crossed with word processing difficulty by having participants read 9 consistent texts with easy words and 9 consistent texts with difficult words, and vice versa for the inconsistent condition. To encourage reading for meaning, each text was followed by one local and one global comprehension question that required students to rate statement as either true or false. On average, participants answered about four in five questions correctly ( $M = 79\%$ ,  $SD = 3\%$ ), confirming that they had read for meaning.

## Equipment

The texts were presented via the Experiment Builder software (SR Research, 2009) and eye movements were recorded using a laptop-mounted Eyelink Duo eye tracker at a sampling rate of 1000 Hz. Texts were formatted in black, Consolas font, size 22, with double spacing, and on a white background. Participants read the texts silently from a Dell Latitude E5530 15-inch laptop monitor, set at a refresh rate of 60 Hz with 1920 × 1080 resolution, at a viewing distance of 55 cm. Participants' eye movements were tracked for the whole duration of the reading task. Although our participants read with both eyes, only the right eye was monitored, unless problems (e.g.,



reflections on glasses) made it necessary to switch to the left eye to gain better data quality. A chin and forehead rest was used to minimize head movements. The camera was calibrated on a five-point grid and each trial started with drift corrections to make sure calibration was sufficiently precise ( $<0.5^\circ$  of accuracy). Calibration was repeated after each break and as necessary.

## Measures

### Executive control test

Executive control was measured using the automated operation span task as it is implemented in the testing software Inquisit (Inquisit4, 2015) on the Millisecond Library (Engle & Kane, 2003). In the automated operation span task, participants are asked to remember letter sequences that are presented one-by-one. Each letter is followed by a simple mathematical statement which participants have to judge to be true or false. The final absolute operation span score is accuracy-based and records the number of letters recalled within correct sequences. To adapt the automated operation span task for use with adolescents, we simplified its instructions and tested only 3 trials for each sequence length, which ranged from 2–6 letters. Additionally, we used only single-step mathematical equations with numbers ranging from 1 to 5 (Gradisar et al., 2008). To check the task's reliability, we assessed its split-half reliability (adjusted using the Spearman-Brown prophecy formula). Cronbach's  $\alpha$  was 0.82, similar to the operation span's reliabilities with adult test-takers (Engle & Kane, 2003). On average, our participants recalled 38 letters in correct sequence ( $SD = 12$ ).

### Word reading test

Word reading was assessed using the Test of Word Reading Efficiency 2 (TOWRE-2; Torgesen et al., 2012) where participants are asked to read out loud as many words and non-words as possible within 45 s.

### Vocabulary test

Vocabulary was assessed using two subtests from the Test of Word Knowledge (TOWK; Wiig & Secord, 1992) where participants had to choose either the correct synonym (for the synonyms subtest) or antonym (for the word opposites subtest) out of three or four answer options. The synonym and opposite subtests of the TOWK proved highly reliable in our sample, with Cronbach's  $\alpha$  of 0.91 and 0.89, respectively.

## Procedure

Each participant took part in one individual and one group test session, each lasting one school lesson or 45 min. During individual sessions, participants completed the

reading task: after a short instruction phase where participants also read one trial text, they read at their own pace and moved from one text to the next by pressing a button on a gamepad. During one of their two breaks from reading, participants completed the TOWRE-II. During group sessions, all participants from one class (4–10 students each time) worked side-by-side but individually on one laptop each. They completed the vocabulary test, the executive control task, and the background questionnaires. All procedures and measures were approved by the departmental ethics committee.

### Eye movement pre-processing

The eye movement data was cleaned and analysed using *popEye* (Schroeder, 2019). During pre-processing, gaze drift was corrected semi-automatically and fixations shorter than 80 ms were first merged within 1 letter distance and then, fixations shorter than 40 ms were merged within 2 letters distance. Eye movements on inconsistency targets (e.g., *Spanish*) and word processing difficulty manipulated words (e.g., *want* or *prefer*) were distilled into early measures (gaze duration, or the times spent initially fixating a word, thought to tap word identification) and late measures (such as go-past time, or the time spent (re)fixating a word and its preceding context before continuing to the right; rereading time, or the time spent refixating a word; as well as regression probabilities, or the likelihood to refixate a word—all thought to reflect text-integration processes). To tap overall word processing difficulty, we additionally computed global measures that summarise reading behaviour on the entire text, both in terms of the time spent with initial reading (first-pass text reading time) and time spent rereading words (second-pass text reading time). Note that it is common in eye-tracking research to compute and analyse both early and late measures of reading as a means to provide a complete picture of the reading process—even if, as in our case, predictions for inconsistency target words focused on later measures, in particular rereading.

## Results

The data were analysed in R (R Core Team, 2021) using linear and binomial mixed effect models from the *lme4* package version 1.1.21 (Bates et al., 2015). All data and the analysis script are available on the projects' online repository on the OSF. Before running the models, the data were cleaned as follows. First, trials were deleted if data quality was too poor or if the target word had been skipped in the first run, which was the case for 3.3% and 22% of all trials, respectively. Then, extremely long and short reading times (gaze durations below 50 ms or above 800 ms; go-past time, rereading and total times above 4000 ms; whole text reading times above 30 s for first-pass and above 60 s for second-pass reading) were deleted, followed by model-based deletions based on a cut-off of 2.5 standard deviations from the person and item means for each reading time measure. This removed on average 2.2% ( $SD=3.4\%$ ) of the data. Reading times were log-transformed. We then ran models

with the fixed effects inconsistency (consistent vs. inconsistent), word processing difficulty (easy vs. difficult words), individual differences in executive control, and their interactions to predict reading times on both inconsistency targets (to tap comprehension monitoring) and words manipulated in their difficulty (to tap the effects of word processing difficulty). Note that these three fixed effects (inconsistency, word processing difficulty and executive control) were present in all computed models to allow us to test our predictions. Although executive control was not correlated with L2 vocabulary ( $r=0.12$ , *ns*), we further wanted to ensure that the impact of executive control was independent of that of English-language vocabulary. To this end, we reran all our models while controlling for L2 vocabulary, a measure widely used as a proxy of overall proficiency (cf. de Cat, 2020; Lervåg & Aukrust, 2010). These new models confirmed all previous results answering our research questions on inconsistency, executive control and word processing difficulty, the results of which we thus continue to report in this manuscript.

Crossed random intercepts and slopes by inconsistency and word processing difficulty were included for participants and items. Item random effects were estimated by stimuli texts (for models predicting reading on inconsistency targets) or by word processing difficulty targets (for models predicting reading on the words manipulated for word processing difficulty). This random structure was only reduced when high inter-correlation and singularity caused convergence problems (Barr et al., 2013).<sup>2</sup> Assumptions of linearity and normality of residuals as well as homoscedasticity of variance were confirmed through visual inspection of residual plots and histograms of residuals (Winter, 2013). Effect coding and Type II model comparisons were used to determine the significance of the fixed effects using the *Anova* function of the *car* package (Fox et al., 2013). Random effects were tested using likelihood ratio chi square comparisons as implemented in the *lmerTest* package for linear models (Kuznetsova et al., 2017) and the *anova* function of the *stats* package (R Core Team, 2021) for binomial models. Model fit was assessed with reference to marginal and conditional  $R^2$  (Nakagawa et al., 2017). Post-hoc tests were computed using cell means models and single degree of freedom contrasts implemented in the *glht* function of the *multcomp* package (Hothorn et al., 2008).

<sup>2</sup> The final random effects structure for most models predicting reading of inconsistency target words contained random intercepts only, except for models predicting gaze duration, go-past and total time that also retained word processing difficulty slopes for items. For models predicting reading on word processing difficulty targets, the final random effect structure was more complex, containing slopes in the following instances: models predicting gaze duration, go-past and rereading time and probability contained word processing difficulty slopes for both subjects and items, while models predicting regression in probability contained word processing difficulty slopes for items only. Additional inconsistency slopes for items were retained in the models predicting go-past and rereading time as well as regression in and rereading probability. Models predicting total time retained both inconsistency and word processing slopes for items and subjects. Finally, models predicting first-pass text reading retained inconsistency slopes for subjects and word processing difficulty slopes for items.

**Table 4** Model means for reading times of target words by inconsistency, word processing difficulty and executive control

	Reading of target words (e.g., <i>Spanish</i> )					
	Gaze duration	Go-past time	Regression in probability	Rereading probability	Rereading time	Total time
<i>Weaker executive control</i>						
Easy words						
Consistent	257 (14)	359 (29)	0.12 (0.04)	0.38 (0.08)	10 (5)	381 (40)
Inconsistent	291 (16)	368 (28)	0.21 (0.05)	0.54 (0.08)	22 (11)	473 (49)
Difficult words						
Consistent	299 (17)	354 (30)	0.15 (0.04)	0.39 (0.08)	11 (5)	407 (42)
Inconsistent	298 (17)	400 (33)	0.23 (0.05)	0.49 (0.08)	18 (9)	481 (49)
<i>Stronger executive control</i>						
Easy words						
Consistent	261 (15)	354 (29)	0.24 (0.05)	0.51 (0.08)	20 (10)	429 (46)
Inconsistent	258 (15)	356 (28)	0.21 (0.05)	0.43 (0.08)	14 (7)	409 (44)
Difficult words						
Consistent	294 (17)	370 (32)	0.10 (0.04)	0.32 (0.08)	8 (4)	398 (42)
Inconsistent	288 (16)	388 (32)	0.21 (0.05)	0.51 (0.08)	21 (11)	479 (50)

Back-transformed, rounded model means. Reading times in milliseconds. Standard errors in parentheses

### Inconsistency main effects

We begin by reporting L2 learners' reading of consistent and inconsistent target words as indication of their comprehension monitoring. The data are summarized in Tables 4 and 5 and model results are summarized in Tables 6 and 7. The models showed that inconsistent targets were reread for longer and more often than consistent targets, confirming that L2 learners had monitored their comprehension on inconsistent words during their reading (as evident in longer rereading and total times and higher rereading and regression in probabilities, see Fig. 1). Other than predicted, however, our adolescent readers in the current study did not read inconsistent targets longer in early reading times (inconsistency effect *ns* for gaze duration and go-past time).

### Word processing difficulty main effects

Word processing difficulty affected L2 learners' reading of manipulated words as well as their reading of the whole texts. As intended, difficult words and texts containing them were read for longer than easy words, both in early and late reading measures (as evident in longer gaze durations and longer go-past, rereading,

**Table 5** Model means for reading times of word processing difficulty manipulated words (e.g., written or prepared) and whole texts by inconsistency, word processing difficulty and executive control

		Word processing difficulty manipulated words					Whole text		
		Gaze duration	Go-past time	Regression in probability	Rereading probability	Rereading time	Total time	First-pass time	Rereading
<i>Weaker executive control</i>									
Easy words									
Consistent	273 (11)	386 (22)	0.16 (0.03)	0.43 (0.06)	12 (5)	415 (34)	14.5 (0.9)	6.8 (1.3)	
Inconsistent	278 (11)	390 (22)	0.16 (0.03)	0.46 (0.06)	15 (6)	440 (36)	14.9 (0.9)	7.4 (1.4)	
Difficult words									
Consistent	365 (16)	647 (51)	0.20 (0.03)	0.49 (0.08)	19 (8)	767 (78)	16.3 (1.0)	7.8 (1.5)	
Inconsistent	369 (16)	661 (52)	0.23 (0.03)	0.55 (0.08)	27 (11)	820 (83)	15.8 (0.9)	8.3 (1.6)	
<i>Stronger executive control</i>									
Easy words									
Consistent	271 (11)	374 (22)	0.17 (0.03)	0.42 (0.06)	12 (5)	417 (36)	13.1 (0.8)	6.5 (1.3)	
Inconsistent	267 (11)	367 (22)	0.16 (0.03)	0.43 (0.06)	13 (6)	411 (35)	13.1 (0.8)	6.1 (1.2)	
Difficult words									
Consistent	332 (15)	556 (45)	0.24 (0.03)	0.56 (0.08)	29 (13)	681 (73)	14.4 (0.9)	7.3 (1.5)	
Inconsistent	330 (15)	556 (45)	0.22 (0.03)	0.52 (0.08)	24 (11)	676 (72)	14.5 (0.9)	7.4 (1.5)	

Back-transformed, rounded model means. Reading times in ms for difficulty manipulated words and in s for whole texts. Standard errors in parentheses

**Table 6** Models predicting the reading of target words (e.g., *Spanisht*) from inconsistency, word processing difficulty, and executive control

Reading of target words (e.g., <i>Spanisht</i> )						
	Gaze duration	Go-past time	Regression in probability	Rereading probability	Rereading time	Total time
<i>Fixed effects</i>						
Word difficulty	0.05** (0.02)	0.02 (0.02)	-0.08 (0.09)	-0.08 (0.08)	-0.08 (0.09)	0.02 (0.02)
Inconsistency	0.01 (0.01)	0.02 (0.02)	0.23** (0.09)	0.19** (0.07)	0.24** (0.09)	0.07*** (0.02)
Executive control	-0.02 (0.03)	-0.00 (0.04)	0.03 (0.12)	-0.01 (0.19)	0.01 (0.25)	-0.01 (0.06)
Word difficulty* Inconsistency	-0.02 (0.01)	0.02 (0.02)	0.09 (0.09)	0.11 (0.07)	0.13 (0.09)	0.02 (0.02)
Word difficulty* Executive control	0.01 (0.01)	0.01 (0.02)	-0.18* (0.09)	-0.04 (0.07)	-0.05 (0.09)	0.00 (0.02)
Inconsistency* Executive control	-0.02 (0.01)	-0.01 (0.02)	-0.06 (0.09)	-0.07 (0.07)	-0.09 (0.09)	-0.03 (0.02)
Word difficulty* Inconsistency* Executive control	0.01 (0.01)	-0.01 (0.02)	0.14 (0.09)	0.18* (0.07)	0.20* (0.09)	0.03* (0.02)
<i>Random effects</i>						
Participants	41.6***	63.2***	9.3*	105.4***	141.3***	220.9***
Words	14.1**	62.6***	18.8***	2.0	3.9*	48.1***
Marginal R <sup>2</sup> /Conditional R <sup>2</sup>	0.019/0.160	0.005/0.221	0.031/0.181	0.021/0.264	0.015/0.240	0.018/0.360

Fixed effect estimates are b estimates with standard error in brackets. Random effect estimates are log-likelihood chi-square estimates for random slopes and intercepts

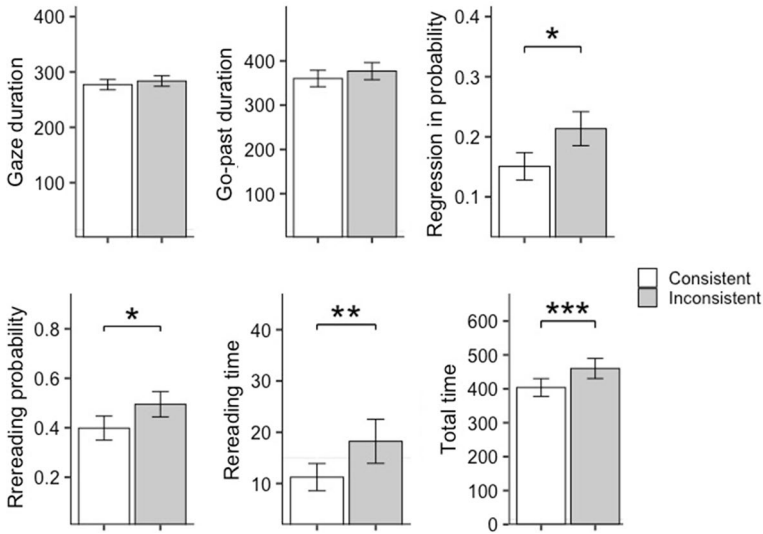
\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

**Table 7** Models predicting reading of word processing difficulty manipulated words (e.g., written or prepared) and whole texts from inconsistency, word processing difficulty and executive control

	Word processing difficulty manipulated words					Whole text		
	Gaze duration	Go-past time	Regression in probability	Rereading probability	Rereading time	Total time	First-pass time	Rereading
<i>Fixed effects</i>								
Word processing difficulty	0.12*** (0.01)	0.23*** (0.02)	0.19*** (0.04)	0.20*** (0.05)	0.33*** (0.06)	0.28*** (0.02)	0.05*** (0.00)	0.07*** (0.01)
Inconsistency	0.00 (0.01)	0.00 (0.01)	-0.02 (0.04)	0.03 (0.04)	0.05 (0.04)	0.01 (0.01)	-0.00 (0.00)	0.01 (0.01)
Executive control	-0.03 (0.02)	-0.05 (0.04)	0.04 (0.10)	0.00 (0.19)	0.01 (0.24)	-0.05 (0.06)	-0.06 (0.04)	-0.05 (0.12)
Word processing difficulty* Inconsistency	0.00 (0.01)	0.00 (0.01)	0.00 (0.04)	-0.01 (0.03)	-0.01 (0.04)	0.00 (0.01)	-0.01 (0.00)	0.00 (0.01)
Word processing difficulty* Executive control	-0.02 (0.01)	-0.03* (0.01)	0.01 (0.04)	0.04 (0.05)	0.04 (0.06)	-0.03* (0.01)	0.00 (0.00)	0.01 (0.01)
Inconsistency* Executive control	-0.01 (0.01)	-0.01 (0.01)	-0.05 (0.04)	-0.06 (0.03)	-0.09* (0.04)	-0.02 (0.01)	0.00 (0.00)	-0.02 (0.01)
Word processing difficulty* Inconsistency* Executive control	0.00 (0.01)	0.00 (0.01)	-0.03 (0.04)	-0.04 (0.03)	-0.04 (0.04)	0.00 (0.01)	0.01* (0.00)	0.01 (0.01)
<b>Random effects</b>								
Participants	255.5***	686.2***	126.9***	663.5***	806.1***	1428.5***	1400.9***	1096.7***
Words	153.4***	527.8***	106.9***	111.1***	155.1***	558.8***	519.8***	72.11***
Marginal R <sup>2</sup> /conditional R <sup>2</sup>	0.065/0.240	0.124/0.417	0.010/0.207	0.010/0.339	0.012/0.281	0.155/0.521	0.062/0.799	0.035/NA

Fixed effect estimates are b estimates with standard error in brackets. Random effect estimates are log-likelihood chi-square estimates for random slopes and intercepts  
 Model syntax was as follows for each model

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$



**Fig. 1** Main effect of inconsistency on the reading of target words. Reading times are model means in ms. *Note* Results stem from mixed effect models that predicted reading times from the fixed effects inconsistency, executive control, word processing difficulty, and their interactions

and total times as well as first-pass text reading). They also received more rereading and direct regressions in, as summarized in Fig. 2.

### Interactions between word processing difficulty and inconsistency

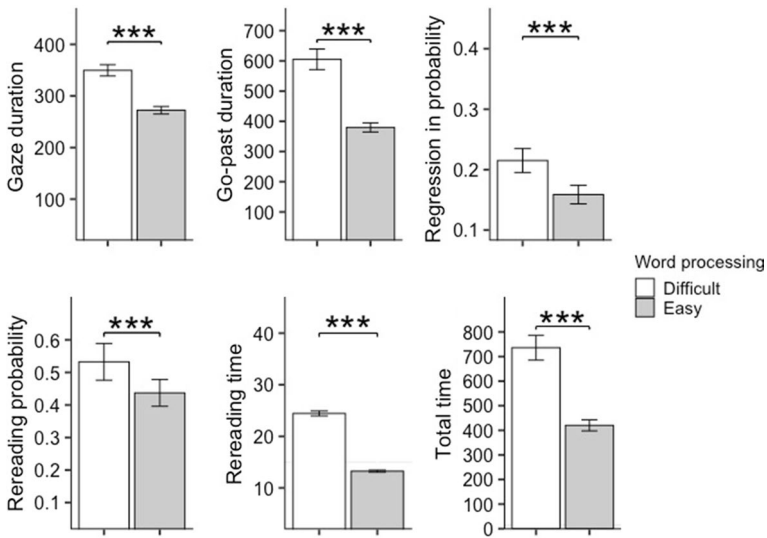
We also examined how word processing difficulty influenced L2 learners' monitoring on inconsistencies. Against our predictions, the two variables did not interact (interaction effects were not significant in go-past time, rereading time, and rereading and regression in probability), despite clear main effects of inconsistency and word processing difficulty in all these reading measures.

### Effects of individual differences in executive control

We also investigated the effect of individual differences in executive control. While we found no main effects, there were clear interactions with both inconsistency and word processing difficulty that indicated a relationship of comprehension monitoring with word processing difficulty and executive control. We explored these interactions by evaluating predicted reading times for adolescents at 1 *SD* above or below the sample mean of executive control.<sup>3</sup> This approach allowed us to study individual differences based on the entire variation of our continuous variable without having to assign participants to groups (Halekoh & Højsgaard, 2009), which would be

<sup>3</sup> The respective point estimates (these were 26 for -1SD and 50 for +1SD) were well within the executive control range of our sample.



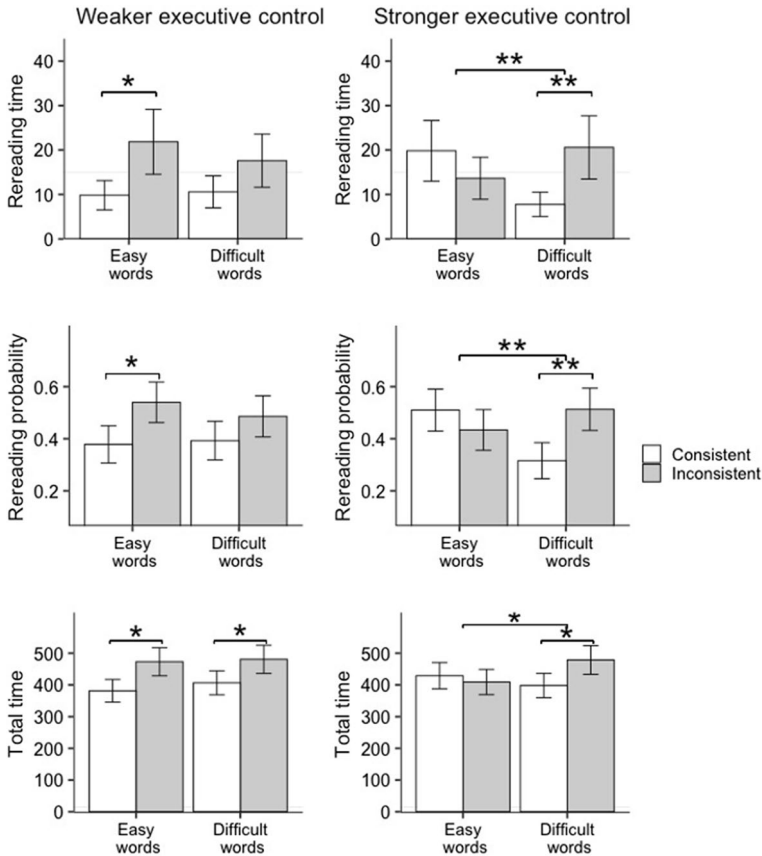


**Fig. 2** Word processing difficulty main effects on the reading of word processing difficulty manipulated target words. Reading times are model means in ms. *Note* Results stem from mixed effect models that predicted reading times from the fixed effects inconsistency, executive control, word processing difficulty, and their interactions

associated with a loss of data precision (Gelman & Hill, 2007). For ease of reading, we will however refer to these model-based estimates as the reading times of adolescents with stronger (+ 1 *SD*) or weaker executive control (− 1 *SD*), respectively.

We found several two-way interactions of executive control with inconsistency or word processing difficulty which showed that in general, adolescents with weaker executive control were more strongly affected by inconsistencies and difficult words. This was visible in them taking longer to read texts that contained both inconsistencies and difficult words ( $z=2.5$ ,  $p<0.05$  for first-pass text reading), to read difficult words ( $z=2.1$ ,  $p<0.05$  for go-past time and, marginally for total times,  $z=1.9$ ,  $p=0.058$ ) or difficulty manipulated words in inconsistent texts ( $z=2.1$ ,  $p<0.05$  for rereading time) vis-à-vis their peers with stronger executive control. They also took longer to read texts that contained both inconsistent and difficult words ( $z=-2.1$ ,  $p<0.05$  for first-pass text reading), as confirmed by a three-way interaction on overall text reading.

Further three-way interactions indicated that adolescents adapted their rereading of inconsistent information differently to easy or difficult texts, depending on their executive control (see Fig. 3): adolescents with weaker executive control took significantly longer to read inconsistent as opposed to consistent targets only in texts with easy words ( $z=2.4$ ,  $p<0.05$  for rereading time,  $z=2.3$ ,  $p<0.05$  for rereading probability, and  $z=3.2$ ,  $p<0.01$  for total time), but not in difficult texts (all effects *ns*). Adolescents with larger executive control, however, did the



**Fig. 3** Interaction between word processing difficulty, inconsistency and executive control on the (re)reading of target words. Reading times are model means in ms. *Note* Results stem from mixed effect models that predicted reading times from the fixed effects inconsistency, executive control, word processing difficulty, and their interactions

exact opposite: they regressed more directly into target words ( $z = 2.1, p < 0.05$ ) and reread inconsistencies for longer when texts were difficult, as visible both in significant interactions on rereading time ( $z = -2.7, p < 0.01$ ), rereading probability ( $z = -2.7, p < 0.01$ ), and total time ( $z = -2.3, p < 0.05$ ). This interaction effect seemed almost reversed on texts containing easy words, but this difference remained non-significant across all reading measures. In summary, while adolescents with weaker executive control showed the expected smaller adaptations to inconsistencies under increased word processing difficulty, adolescents with stronger executive control adapted their rereading more (instead of less) clearly to inconsistencies when reading difficult texts, indicating that comprehension monitoring under increased word processing difficulty varied by executive control abilities.

## Discussion

We had investigated how adolescents' comprehension monitoring in the L2 arises in the interplay of word processing difficulty (manipulated on the text-level) and individual executive control abilities. We had expected successful comprehension monitoring to show in increased rereading of inconsistent vis-à-vis consistent information, indicating re-analysis. We had further expected that higher word processing difficulty would reduce this re-analysis, in line with the bottleneck idea that effortful word processing limits higher-level processing (Hessel & Schroeder, 2020; Perfetti & Stafura, 2014). Finally, we were interested in seeing whether adolescents' comprehension monitoring on inconsistent words (and the way it decreased with difficult word processing) would also vary with adolescents' executive control, which would indicate an additional influence of active, reader-initiated processing (van den Broek & Helder, 2017). The overall emerging picture is that while increased word processing difficulty can reduce comprehension monitoring in the L2, sufficient executive control allows to compensate and even reverse this influence: as predicted, we found clear signs of adolescents' comprehension monitoring in slowing down on and rereading inconsistent information. We further saw clear signs of increased word processing difficulty on words manipulated in their word processing difficulty and in overall text reading times. Other than predicted, however, not all adolescents reduced their comprehension monitoring under higher word processing difficulty. Instead, we found that while adolescents with weaker executive control monitored their comprehension less clearly under high word processing difficulty, adolescent with stronger executive control instead monitored their comprehension *more* clearly under higher difficulty, indicating a switch to active, reader-initiated monitoring on demanding texts. In the following, we consider these key results in relation to theories and previous evidence.

We had measured comprehension monitoring online, by recording adolescents' eye movements on inconsistencies. We found clear signs that adolescents were able to monitor their comprehension when reading in the L2, as visible in increased rereading of inconsistent information. This finding is akin to other evidence with children and adults reading in their L1 (Connor et al., 2015; Hessel et al., 2021; Joseph et al., 2008; Rayner et al., 2004) and extends the evidence base on online monitoring in the L2 by adding detailed insights into adolescent readers (as previous L2 studies have either focused on younger children or adults; Hessel et al., 2021; Hessel & Schroeder, 2020; or been limited to less detailed reading time data from self-paced reading; Mulder et al., 2021). On a more general note, this finding could be considered encouraging news for all language practitioners as it indicates that adolescent learners can reliably process information from expository texts, even when reading in their L2.

We had also examined under which conditions students would be able to monitor their comprehension successfully. To find out, we had purposefully altered word processing difficulty across texts. As we had expected, texts designed to trigger higher word processing difficulty (that is, those that contained longer and

less frequent and thus more difficult words) slowed down adolescents' reading. These results are in line with previous studies on word length and frequency and show that our manipulation of word processing difficulty worked as intended (Gagl et al., 2015; Inhoff & Rayner, 1986; Joseph et al., 2009, 2013; Juhasz & Rayner, 2006; Tiffin-Richards & Schroeder, 2015). However, higher word processing difficulty did not influence comprehension monitoring in the predicted way. As a reminder, we had expected that all students would be *less* able to monitor their comprehension under higher difficulty due to a processing bottleneck, which would have been visible in reduced rereading of inconsistencies (Perfetti & Stafura, 2014). However, only a sub-group of students showed this predicted reading pattern, namely, those with relatively weaker executive control. For those adolescents, higher word processing difficulty limited their monitoring, as visible in them adapting their rereading less to inconsistent information when text contained difficult words. In line with the Reading Systems Framework (Perfetti & Stafura, 2014) and previous evidence with adults (Hessel & Schroeder, 2020), we can say that for adolescents with weaker executive control, less efficient word processing appears to have started a chain-reaction that ended in reduced higher-level coherence building.

For adolescents with stronger executive control, however, the relationship between word processing difficulty and monitoring was reversed: they engaged in *more* (instead of less) pronounced monitoring when word processing was difficult, as visible in increased rereading of inconsistencies. While going contrary to the mechanisms described in the Reading Systems Framework (Perfetti & Stafura, 2014), this reading pattern aligns with theories that emphasise active processing, such as the Reading Framework (van den Broek & Helder, 2017). If it is true that readers switch from passive to active processing in the face of insufficient comprehension (van den Broek & Helder, 2017), the observed pronounced monitoring on difficult texts could be adolescent's active response to increased comprehension demands. However, such active repair seems to only have been available to L2 readers with stronger executive control.

Note that numerically, adolescents with stronger executive control furthermore appeared to spend more time rereading consistent information when reading easy texts (rather than inconsistencies, as predicted). Importantly, this difference failed to reach significance and is thus hard to interpret without corroborating evidence. With that caveat in mind, we would like to suggest, however, that if this effect was found to be true, one explanation for it could be that high executive readers may have felt overly confident in their comprehension of easy texts. This could have led to shallow processing and overlooking coherence breaks, as previously found for metacognitive monitoring (Ikeda & Kitagami, 2013). However, this suggestion remains purely speculative without further supporting evidence.

Our results only partially replicate previous ones: in another study with comparable reading materials, more difficult word processing reduced readers' monitoring in a clear two-way interaction (Hessel & Schroeder, 2020). This was not the case in the current data. We would like to suggest two explanations for this incongruence—one based on reader, the other on task differences. Regarding reader differences, the adolescents in the current study were less skilled readers than the adults in the previous

study (as visible in fewer years of English language experience (about 4 years for adolescents as opposed to 13 years or more for adults) and lower word reading abilities (cf. the average score of 121 for adolescents as opposed to 136 for adults). For adolescents, reading in English was thus plausibly more difficult. On the one hand, more effort overall could have arguably increased (and not decreased) the impact of word processing difficulties on monitoring in the current study (Perfetti & Stafura, 2014), making our results seem even less intuitive. On the other hand, increased difficulties would arguably also increase the likelihood of active processing (van den Broek & Helder, 2017)—which could explain why those adolescents capable of it switched to an active processing style in the current study. Differences in reading tasks may have additionally supported such a switch, as task effects on reading and rereading in particular are well documented in adults (Kaakinen & Hyönä, 2005; Kaakinen et al., 2003; Weiss et al., 2018) and young readers (Kaakinen et al., 2015). While in the previous study, adults had only been asked to recall texts (Hessel & Schroeder, 2020), adolescents had now answered local and global comprehension questions. Recall has been found to tap shallower comprehension compared to multiple-choice questions (Cao & Kim, 2021). Plausibly, the reading task in the current study thus triggered higher comprehension standards. Together with higher comprehension difficulties, these increased standards could expectably inspire active processing in those readers who are capable of it (van den Broek & Helder, 2017), just as we had observed.

At this point, three aspects of our study must be acknowledged that limit the conclusions we can draw. First, as for any school-based study, our sample size was limited by response levels as well as the occasional technical and data quality issues that result from testing at schools. All our results and conclusions would benefit from replication in another, larger sample. Second, while we investigated the impact of higher word processing difficulty experimentally, we tapped executive control through individual differences. The latter thus remains subject to possible entanglements with related individual differences (Castles et al., 2018) until tested in a targeted experimental manipulation (such as in de Bruïne et al., 2021). Finally, we tested comprehension monitoring by observing adolescents' (re)reading of textual inconsistencies. While the inconsistency paradigm is a widely used approach in online studies of comprehension monitoring (Connor et al., 2015; Hessel et al., 2021; Joseph et al., 2008; Rayner et al., 2004) and while there is widely accepted evidence in favour of rereading as a critical support of comprehension (Mulder et al., 2021; Schotter et al., 2014), it is only fair to say that understanding the rereading of inconsistencies as online monitoring remains an interpretative step that relies on the evidence from these previous studies as its foundation.

More generally, our findings contribute to a more nuanced view of what adolescents need to engage in successful comprehension monitoring in their L2. Specifically, ease of word processing is often highlighted as the key driver of higher-level processing (Perfetti & Stafura, 2014) while complementary theories predict additional situation-specific support through active processing (van den Broek & Helder, 2017). Yet, we know little about the tipping point at which readers shift from passive to active processing during moment-to-moment monitoring. Our study illustrates that exact tipping point by showing how comprehension monitoring in the L2 may

be reduced under higher word processing load, while adolescents can compensate these effects through active reading given sufficient executive control. Future work could hone in further on this interplay of active and passive reading, for example by observing monitoring across reading tasks that yield varying comprehension standards, word difficulty, and executive control load.

For teaching, our results hold both good news and new directions. The good news is that overall, adolescents who read expository texts in their L2 are capable of monitoring their comprehension from moment to moment—an important message in times of the rise of content teaching in the L2 worldwide (Dearden, 2015). As for new directions, our findings indicate that comprehension monitoring in the L2 is carried by both fluent word processing and active monitoring. In principle, this gives teachers a greater toolkit that includes both vocabulary (Beck et al., 1982) and active monitoring exercises (Burton & Daneman, 2007). We believe that our understanding of higher-level reading comprehension will benefit from more insights into the complex interactions between active and passive processing, as we have provided them here for adolescents' comprehension monitoring when they read in the L2.

**Acknowledgements** We want to thank Lisa Schorn and Julian Pfeiffer for their important contribution to data collection.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by and under the supervision of AKH. The first draft of the manuscript was written by AKH and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data availability** Data, analysis script and all materials available via: [https://osf.io/7jdpu/?view\\_only=c7ca336f2bcb4661a3a07ad5ccc695fb](https://osf.io/7jdpu/?view_only=c7ca336f2bcb4661a3a07ad5ccc695fb).

## Declarations

**Conflict of interest** The authors declared that there were no conflict of interest with respect to the authorship or the publication of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Baker, L. (1984). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology*, 76(4), 588–597. <https://doi.org/10.1037/0022-0663.76.4.588>

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Beck, I., Perfetti, C. A., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521.
- Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. (2011). Increased stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology*, *64*(3), 596–607. <https://doi.org/10.1080/17470218.2010.513735>
- Burton, C., & Daneman, M. (2007). Compensating for a limited working memory capacity during reading: Evidence from eye movements. *Reading Psychology*, *28*(2), 163–186. <https://doi.org/10.1080/0270210601186407>
- Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, *9*(4–6), 365–381. <https://doi.org/10.1080/09658210143000083>
- Cao, Y., & Kim, Y. G. (2021). Is retell a valid measure of reading comprehension? *Educational Research Review*, *32*, 100375. <https://doi.org/10.1016/j.edurev.2020.100375>
- 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>
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & McDonald, C. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, *19*(2), 114–134. <https://doi.org/10.1080/10888438.2014.943905>
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE*, *10*(8), 1–38. <https://doi.org/10.1371/journal.pone.0134008>
- de Bruïne, A., Jolles, D., & van den Broek, P. (2021). Minding the load or loading the mind: The effect of manipulating working memory on coherence monitoring. *Journal of Memory and Language*, *118*, 104212. <https://doi.org/10.1016/j.jml.2020.104212>
- de Cat, C. (2020). Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition*, *42*(2), 279–325.
- Dearden, J. (2015). *English as a medium of instruction - a growing global phenomenon*. [http://www.teachingenglish.org.uk/sites/teacheng/files/pub\\_E484EMI-Coveroption\\_3FINAL\\_Web.pdf](http://www.teachingenglish.org.uk/sites/teacheng/files/pub_E484EMI-Coveroption_3FINAL_Web.pdf)
- Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018). Individual differences in children's pronoun processing during reading: Detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, *173*, 250–267. <https://doi.org/10.1016/j.jecp.2018.04.005>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, *46*, 145–199. [https://doi.org/10.1016/S0079-7421\(03\)44005-X](https://doi.org/10.1016/S0079-7421(03)44005-X)
- Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, *5*(1), 39–52.
- Gagl, B., Hawelka, S., & Wimmer, H. (2015). On sources of the word length effect in young readers. *Scientific Studies of Reading*, *19*(4), 289–306. <https://doi.org/10.1080/10888438.2015.1026969>
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gradisar, M., Terrill, G., Johnston, A., & Douglas, P. (2008). Adolescent sleep and working memory performance. *Sleep and Biological Rhythms*, *6*(3), 146–154. <https://doi.org/10.1111/j.1479-8425.2008.00353.x>
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. <https://doi.org/10.1016/j.tics.2005.07.004>
- Halekoh, U., & Højsgaard, S. (2009). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—The R package pbrtest. *Journal of Statistical Software*, *10*(2), 1–24. <https://doi.org/10.18637/jss.v069.i12>
- Hamilton, S., Freed, E., & Long, D. L. (2016). Word-decoding skill interacts with working memory capacity to influence inference generation during reading. *Reading Research Quarterly*, *51*(4), 391–402. <https://doi.org/10.1002/rrq.148>

- Hessel, A. K., Nation, K., & Murphy, V. A. (2021). Comprehension monitoring during reading: An eye-tracking study with children Learning English as an additional language. *Scientific Studies of Reading*, 25(2), 159–178. <https://doi.org/10.1080/10888438.2020.1740227>
- 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>
- Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, 37(5), 1253–1280. <https://doi.org/10.1017/S0142716415000569>
- Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, 7(1), 96–130. <https://doi.org/10.1075/lab.14027.hop>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1), 44–55. <https://doi.org/10.1037/0022-0663.94.1.44>
- Ikeda, K., & Kitagami, S. (2013). The interactive effect of working memory and text difficulty on meta-comprehension accuracy. *Journal of Cognitive Psychology*, 25(1), 94–106. <https://doi.org/10.1080/20445911.2012.748028>
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40(6), 431–439. <https://doi.org/10.3758/BF03208203>
- Inquisit 4*. (2015).
- Jensen, A. R., & Rohwer, W. D. (1966). The stroop color-word test: A review. *Acta Psychologica*, 25, 36–93. [https://doi.org/10.1016/0001-6918\(66\)90004-7](https://doi.org/10.1016/0001-6918(66)90004-7)
- Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 61(5), 708–723. <https://doi.org/10.1080/17470210701400657>
- Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, 49(16), 2078–2086. <https://doi.org/10.1016/j.visres.2009.05.015>
- Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, 42(2), 207–222. <https://doi.org/10.1080/02796015.2013.12087485>
- Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7–8), 846–863. <https://doi.org/10.1080/13506280544000075>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/h0021465>
- Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes*, 40(3), 239–257. <https://doi.org/10.1207/s15326950dp4003>
- Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 447–457. <https://doi.org/10.1037/0278-7393.29.3.447>
- Kaakinen, J. K., Lehtola, A., & Paattilampi, S. (2015). The influence of a reading task on children's eye movements during reading. *Journal of Cognitive Psychology*, 27(5), 640–656. <https://doi.org/10.1080/20445911.2015.1005623>
- Kim, Y. S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly*, 49(3), 269–281. <https://doi.org/10.1002/rrq.74>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 51(5), 612–620. <https://doi.org/10.1111/j.1469-7610.2009.02185.x>
- Lim, J. H., & Christianson, K. (2014). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, 36(6), 1283–1315. <https://doi.org/10.1017/S0142716414000290>



- Linderholm, T., & Van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*(4), 778–784. <https://doi.org/10.1037/0022-0663.94.4.778>
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*(3), 381–401. <https://doi.org/10.1016/j.jml.2006.06.006>
- Mulder, E., van de Ven, M., Segers, E., Krepel, A., de Bree, E. H., de Jong, P. F., & Verhoeven, L. (2021). Word-to-text integration in English as a second language reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 34*(4), 1049–1087. <https://doi.org/10.1007/s11145-020-10097-3>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R<sup>2</sup> and intraclass correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 944–946. <https://doi.org/10.1098/rsif.2017.0213>
- O'Brien, E. J., & Cook, A. E. (2016). Coherence threshold and the continuity of processing: The RI-Val model of comprehension. *Discourse Processes, 53*(5–6), 326–338. <https://doi.org/10.1080/016853X.2015.1123341>
- Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 18*(7–9), 657–686. <https://doi.org/10.1007/s11145-005-3355-z>
- Pérez, A. M., & Bajo, M. T. (2019). The nature of first and second language processing: The role of cognitive control and L2 proficiency during text-level comprehension. *Bilingualism: Language and Cognition, 22*(5), 930–948. <https://doi.org/10.1017/S1366728918000846>
- Perfetti, C. (1985). *Reading ability*. Oxford University Press.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Prehn, K., Taud, B., Reifegerste, J., Clahsen, H., & Flöel, A. (2018). Neural correlates of grammatical inflection in older native and second-language speakers. *Bilingualism: Language and Cognition, 21*(1), 1–12. <https://doi.org/10.1017/S1366728916001206>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/>
- Raudszus, H., Segers, E., & Verhoeven, L. (2017). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 31*(2), 1–20. <https://doi.org/10.1007/s11145-017-9791-8>
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1290–1301. <https://doi.org/10.1037/0278-7393.30.6.1290>
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science, 25*(6), 1218–1226. <https://doi.org/10.1177/0956797614531148>
- Schroeder, S. (2019). *popEye—An R package to analyse eye movement data from reading experiments*.
- Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15*(3), 232–246. <https://doi.org/10.1080/09297040802220029>
- Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language, 33*(4), 421–441. <https://doi.org/10.1006/jmla.1994.1020>
- SR Research. (2009). *Experiment builder*.
- Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia, 64*, 41–53. <https://doi.org/10.1016/j.neuropsychologia.2014.09.012>
- The British National Corpus. (2007). Bodleian Libraries, University of Oxford. <http://www.natcorp.ox.ac.uk/>
- Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research, 113*, 33–43. <https://doi.org/10.1016/j.visres.2015.05.008>
- Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE 2: Test of word reading efficiency*. Pro-Ed.
- van den Bosch, L. J., Segers, E., & Verhoeven, L. (2018). Online processing of causal relations in beginning first and second language readers. *Learning and Individual Differences, 61*, 59–67. <https://doi.org/10.1016/j.lindif.2017.11.007>

- van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*(5–6), 360–372. <https://doi.org/10.1080/0163853X.2017.1306677>
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2018). Validation: Knowledge- and text-based monitoring during reading. *Discourse Processes*, *55*(5–6), 480–496. <https://doi.org/10.1080/0163853X.2018.1426319>
- Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, *71*(1), 198–210. <https://doi.org/10.1080/17470218.2017.1307862>
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin and Review*, *19*(1), 73–80. <https://doi.org/10.3758/s13423-011-0179-5>
- Wiig, E. H., & Secord, W. (1992). *Test of word knowledge (TOWK): Complete kit*. Pearson.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications (pp. 1–42). ArXiv Preprint, <http://arxiv.org/abs/1308.5499>
- Zabucky, K., & Ratner, H. H. (1989). Effects of reading ability on children's comprehension evaluation and regulation. *Journal of Reading Behavior*, *21*(1), 69–83.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Annina K. Hessel<sup>1,2</sup>  · Sascha Schroeder<sup>1</sup>

<sup>1</sup> Department of Educational Psychology, University of Goettingen, Göttingen, Germany

<sup>2</sup> Department of Language and Learning, Mercator Institute for Literacy and Language Education, University of Cologne, Cologne, Germany