



## UvA-DARE (Digital Academic Repository)

### Word-to-text integration in English as a second language reading comprehension

Mulder, E.; van de Ven, M.; Segers, E.; Krepel, A.; de Bree, E.H.; de Jong, P.F.; Verhoeven, L.

**DOI**

[10.1007/s11145-020-10097-3](https://doi.org/10.1007/s11145-020-10097-3)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

Reading and Writing

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

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*, 34(4), 1049-1087. <https://doi.org/10.1007/s11145-020-10097-3>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.  
*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



# Word-to-text integration in English as a second language reading comprehension

Evelien Mulder<sup>1</sup> · Marco van de Ven<sup>1</sup> · Eliane Segers<sup>1</sup> · Alexander Krepel<sup>2</sup> · Elise H. de Bree<sup>2</sup> · Peter F. de Jong<sup>2</sup> · Ludo Verhoeven<sup>1</sup>

Accepted: 12 October 2020 / Published online: 7 November 2020  
© The Author(s) 2020

## Abstract

We assessed the relationship between word-to-text-integration (WTI) and reading comprehension in 7th grade students ( $n=441$ ) learning English as a second language (L2). The students performed a self-paced WTI reading task in Fall (T1) and Spring (T2), consisting of three text manipulation types (anaphora resolution, argument overlap, anomaly detection), divided in simple and complex passages. The passages contained proximate versus distant anaphora, explicit repetitions versus implicit inferences, and no anomalies versus anomalies. We first examined how WTI complexity was related to reading times on target, target plus one, and target plus two, controlling for word frequency, decoding fluency, gender, and age. Mixed-effects models showed shorter reading times on T2 than on T1 and for simple compared to complex passages, indicating improvement of L2 reading speed. Complexity affected WTI for our L2 learners, as was reflected by longer reading times on complex compared to simple argument overlap and anomaly detection passages. We then assessed whether reading comprehension could be predicted by WTI. Longer reading times on complex compared to simple argument overlap and anomaly detection passages predicted offline reading comprehension. These WTI-measures of complexity are thus indicators of WTI proficiency for novice L2 learners.

**Keywords** Reading comprehension · Second language learning · Word-to-text integration

---

✉ Evelien Mulder  
e.mulder@pwo.ru.nl

<sup>1</sup> Behavioural Science Institute, Radboud University, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

<sup>2</sup> Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

## Introduction

The ability to read text in a second language (L2) is highly important for, amongst other reasons, academic development (Collier, 1987). However, novice L2 learners often experience difficulties when reading a text in their L2, especially with respect to comprehension (Lesaux, Lipka, & Siegel, 2006). These difficulties may arise from word characteristics (e.g., word frequency; Clifton, Staub, & Rayner, 2007), characteristics of the text, specifically sentence complexity (van den Bosch, Segers, & Verhoeven, 2018), and individual differences, for example decoding fluency (Nahatme, 2018). Reading comprehension is partly driven by Word-to-Text Integration (WTI; Perfetti, Yang, & Schmalhofer, 2008). WTI has theoretically been defined as the ability to smoothly retrieve the meaning of a written word and integrate it into the meaning of the text (Perfetti et al., 2008). WTI can directly influence sentence recall (Rosenberg, 1987) and understanding at the discourse level (Gernsbacher, Varner, & Faust, 1990). WTI may therefore play an important part in unravelling the challenges posed by English as an L2 reading comprehension.

During the WTI process, the meanings of words are integrated into the mental model of the text. This takes place within clauses and sentences, but also across sentence boundaries (Chen, Fang, & Perfetti, 2017). A single sentence can be considered as text, for which a situation model needs to be built, just as well as for a passage consisting of multiple sentences. WTI is referred to as rapid text processing, as it is the result of prospective processes anticipating upcoming information and retrospective processes connecting what is currently being read to previously read text (Stafura et al., 2015). These processes, in turn, demand integration of different types of information, such as anaphoric referencing (Dussias, 2003), semantic binding as a result of argument overlap (Yang, Perfetti, & Schmalhofer, 2007), and anomaly detection (van Berkum, Hagoort, & Brown, 1999). In the present study, we were interested in item-based and subject-based variation in WTI processes, measured using word-by-word reading (Perfetti et al., 2008), and sampled three types of WTI processes, namely anaphora resolution, argument overlap, and anomaly detection, which have been studied separately so far. In the present study, we examined to what extent self-paced reading was related to a selection of different WTI text manipulations and aimed to uncover the relationship between WTI and reading comprehension in early L2 learners. An overview of the manipulations included in the present study is presented in Fig. 1.

## WTI and text manipulations

The first manipulation, anaphora resolution, targets syntactic integration. Understanding a sentence passage requires syntactic parsing, in order to move from understanding separate words to understanding word combinations with an underlying syntactic structure. Building a within-sentence structure is not only vital for WTI at a sentence level (Shapiro, Zurif, & Grimshaw, 1987), but also for comprehension at the text level (Helder et al., 2019). The ability to link words to

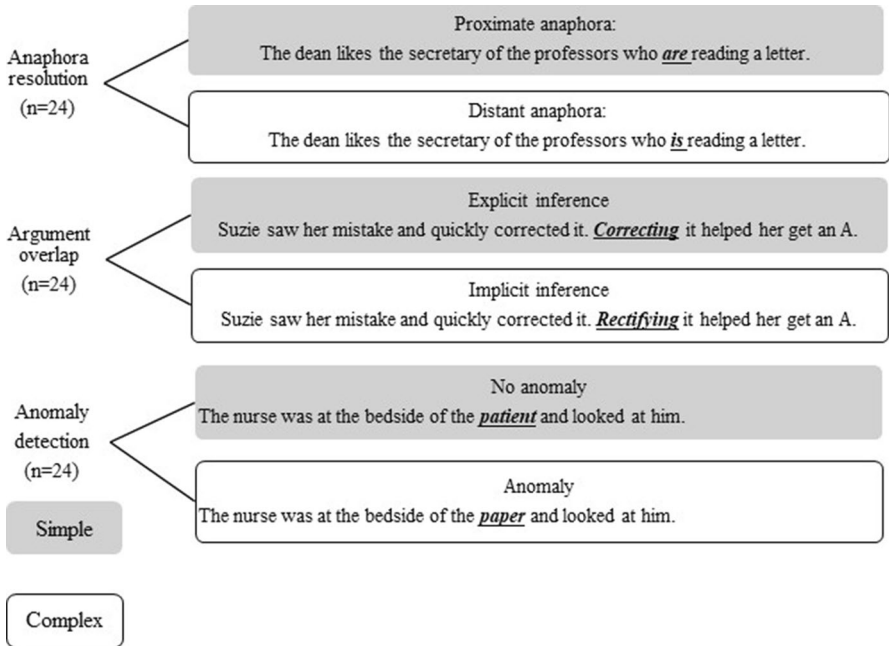


Fig. 1 Graphic overview of different WTI text manipulations and their corresponding complex and simple sentence passage

precedents through WTI may foster building a within-sentence structure. Whereas L1 learners of English will have been exposed to syntactic structures used in English, such as noun–verb correspondences, L2 learners may lack experience with these constructions (Ellis, 2013). Therefore, anaphora resolution may be particularly challenging. Several studies have looked at L1 and L2 anaphoric resolution skills (e.g., Dussias, 2003) with texts of different complexity levels. For example, in the passage ‘The dean likes the secretary of the professors who is/are reading a letter’, the relative clause may either refer to the noun ‘professors’, in the case of the plural verb form ‘are’ (proximate anaphora; Dussias, 2003) or to the noun ‘secretary’, in case of the singular verb form ‘is’ (distant anaphora; Dussias, 2003). Adult Spanish L2 English learners were found to show a preference for proximate over distant attachment during reading when they were asked to indicate their attachment preference on a questionnaire (Dussias, 2003). These results suggest that integration in the case of distant attachment may be more challenging and induces longer reading times compared to reading times during attachment of proximate words. Furthermore, there was a different pattern for syntactic ambiguity than for lexical ambiguity resolution. Whereas lexical ambiguity (‘We all should have known that some metal rings loudly and for a long time’ versus ‘We all should have known that some metal rings are very strong’) rendered longer gaze durations compared to a baseline sentence immediately after exposure to the ambiguity, longer fixations appeared later in the sentence for syntactic category ambiguity resolution (Rayner & Frazier, 1987). Therefore, ease of

syntactic integration may be dependent on anaphoric proximity, i.e., the syntactic complexity of a sentence.

Argument overlap, as a result of which inferences need to be made, is a second type of WTI text manipulation. In the passage ‘After being dropped from the plane, the bomb hit the ground and exploded. The explosion was quickly reported to the commander’, integration of the word ‘explosion’ at the beginning of the second sentence is fostered by the context and degree of inferencing required by the first (Yang, Perfetti, & Schmalhofer, 2007). Binding a word to a preceding referent is a central concept in integration processes (Perfetti et al., 2008). There is evidence that vocabulary knowledge interacts with binding words. On the one hand, inference ability has been associated with derivation of new vocabulary using context. On the other hand, the existence of associations between words, as a result of vocabulary knowledge, seems to support reading comprehension through inference making (Oakhill, Cain, & McCarthy, 2015). In L2 learners, L2 vocabulary is often less developed (Koda, 2007). Therefore, inference making could be more challenging for them than for L1 learners. Most studies concerning argument overlap have used ERP methodology. They show that the required degree of inferencing is related to the strength of the ERP elicited. More specifically, the N400 effect, a negative voltage shift between 300 and 500 ms after the onset of a word, has been related to semantic integration (Yang et al., 2007). The amplitude becomes larger as semantic integration difficulty increases. For example, compared to a baseline sentence that did not require inferencing, sentences with argument overlap in the form of explicit repetitions showed reduced N400 effects (Yang et al., 2007). This suggests that complexity, in terms of the degree of inferencing required as a result of argument overlap, is related to integration processes and may also be reflected in longer reading times.

A third and final type of WTI text manipulation that has often been examined requires semantic integration in the form of updating the mental representation of a text, but has not been addressed as WTI as such in previous studies. In order to understand the meaning of a passage, a reader needs to integrate the semantics of its individual words (van Berkum, Hagoort, & Brown, 1999). This has been operationalized by looking at sentences with semantic violations, or anomalies (van Berkum et al., 1999; Hagoort, 2003), such as: ‘He spread the warm bread with *socks*.’ Semantic challenges may also reflect the complexity of a text. For example, sentences that contain anomalies may be considered complex. Besides syntactic integration and argument overlap, the detection of anomalies is also an important aspect of integration. After all, when building up a coherent model of a text, the reader also needs to be able to update the semantic representations of separate words. Anomaly detection may be additionally challenging for L2 learners because, like simultaneous bilinguals, L2 learners have to inhibit their L1 while processing the anomaly (Bialystok, 2009; Hagoort, 2017).

Again, ERPs are often used to measure integration by means of anomaly detection, and have been shown to vary during semantic integration, i.e. connecting words, or updating. Thus, integration seems to be dependent on the challenges posed by the sentences read. For example, large N400 effects were elicited by semantic anomalies (Kutas & Federmeier, 2011). Both L1 and L2 adult learners show a delay in reading such semantic anomaly sentences as opposed to continuous sentences

(Ahn & Jiang, 2018). The appearance of an N400 (Chen et al., 2017; Helder et al., 2019) and P300 (Perfetti et al., 2008) effect have been proposed to reflect semantic integration, or WTI. If semantic integration fails or is highly complicated, as is the case while reading a semantic violation, large N400 responses have been observed (e.g., van Berkum et al., 1999). Also, in the case of easy semantic integration, N400 effects are still present, albeit reduced (Perfetti et al., 2008). Previous research, comparing both ERPs and reading times within self-paced reading in L1 adults, showed that reading anomalies or weakly constraining sentences resulted in both larger N400 effects and longer reading times than in continuous or highly constraining sentences (Ng, Payne, Steen, Stine-Morrow, & Federmeier, 2017). These results suggest that although good readers will probably fail to integrate an anomalous word into the text, they will attempt to do so. Less skilled readers will probably be less sensitive to anomalies and continue reading. Therefore, passages with anomalies were considered complex in the present study. Previous studies have used self-paced reading, looking at whole sentence reading times in relation to discourse updating (van der Schoot, Reijntjes, & van Lieshout, 2012). How challenges posed by the process of integration, required by the type of sentence and the complexity, are reflected in reading times in early L2 learners has barely been examined.

Besides the three different WTI text manipulations and level of complexity as were discussed above, other lexical factors, such as the frequency of occurrence of a word, have been related to integration processes (e.g., Clifton et al., 2007) and ought to be controlled for when examining integration. Looking at single-word reading, results from eye-tracking studies showed that both first fixations and gaze durations are shorter for high frequency words than for low frequency words (Clifton et al., 2007). However, if a word is encountered several times, these effects diminish dramatically for low frequency words and less so for high frequency words. Frequency effects in an L2 are often explained in the light of the lexical entrenchment paradigm (e.g., Diependaele, Lemhöfer, & Brysbaert, 2013; Whitford & Titone, 2017), which claims that repeated exposure to lexical items leads to fine-grained, integrated lexical representations.

### **Individual differences in L2 WTI and reading comprehension**

In addition to task-related characteristics of Word-to-Text Integration (WTI), participant-related differences could also affect WTI processes and could interact with the effects of the WTI manipulations. One source of individual differences could be decoding fluency. Smooth word decoding enables the availability of sufficient processing capacity to arrive at reading comprehension (Torgesen, 1986). In other words, problems with decoding may be reflected in poor processing on a sentence level, although contextual information could compensate for insufficient decoding skills. Indeed, previous studies have demonstrated that contextual information was predictive of oral reading rate in second grade children (Tortorelli, 2020) and that text higher text complexity was associated with reading errors in 9–15 years old children (Nguyen et al., 2020). The way language comprehension and text decoding are related differs between languages (Koda, 2007). However, previous studies

examining WTI have focused only on adults and L1 learners. As a result, it remains unknown how decoding is related to WTI in an L2, although it could be assumed that students who are slow decoders or less fluent readers also show poor performance on WTI (Torgesen, 1986).

Reading comprehension has been studied using an interactive model in which word identification and WTI processes play central roles (Verhoeven & Perfetti, 2008). According to this model, WTI is required for text comprehension. Words are connected to a text representation, which is continuously updated as words are being identified. Building on successful WTI, readers have to combine sentence meanings to prior knowledge, to comprehend text. Promising positive effects of a WTI intervention on reading comprehension were found in elementary school Dutch L1 learners (Swart et al., submitted). Furthermore, ERP results in adults suggest that weak comprehenders show less or later integration of what is read (Yang et al., 2005). With regard to the interaction between individual differences and WTI text manipulations, adolescents with stronger reading comprehension skills showed quicker knowledge-to-text integration in causal rather than temporal text passages. However, adolescents with weaker reading comprehension skills did not show a difference in the speed of knowledge-to-text integration between causal and temporal text passages (Barnes, Ahmed, Barth, & Francis, 2015). Although the Simple View of Reading seems to apply to L2 learners similarly as to L1 learners (Verhoeven & van Leeuwe, 2012), L2 learners may not be competent enough to benefit from supportive text-based factors, such as coherence marking (Degand & Sanders, 2002). Although some studies found that proficient L2 learners benefit more from contextual cues than do less proficient readers (e.g., Nahatme, 2018; Todaro, Millis, & Dandotkar, 2010).

## Present study

In summary, the different levels of linguistic representation involved in WTI have been measured using different WTI text manipulations and complexities, such as those that require anaphora resolution (Dussias, 2003), inference making as a result of argument overlap (Yang et al., 2007), and anomaly detection (van Berkum et al., 1999), all of which have mainly been investigated in L1 adults. A perspective on WTI in early L2 learners is thus lacking. In order to establish a multi-faceted measure of WTI, reading disruptions as a reflection of integration need to be examined.

Thus, in the present study, we used a computerized, self-paced reading task to measure WTI in novice Dutch students learning English as an L2 just after the beginning (T1) and near to the end (T2) of the 7th grade. The self-paced reading task consisted of 72 sentences passages, divided across three types of WTI text manipulations, with two levels of complexity, namely simple and complex: proximate versus distant anaphora, explicit repetition versus implicit inferences, and no anomalies versus anomalies. The single sentence passages we used in the present study were based on studies that also examined integration in single sentence passages, but did not address WTI as such.

We first explored whether reading times could be predicted by the three WTI text manipulations and their complexities, and by students' decoding fluency, after controlling for word frequency, gender, and age. We were specifically interested in reading times on the word positions target, target plus one, and target plus two (based on Bultena et al., 2015). The complexity effect was expected to be reflected as longer reading times on complex versus simple passages. Therefore, as a measure of WTI, for every participant we divided the reading times between complex and simple passages for each text manipulation and word position. With this index, we could examine the average additional reading time per participant, to read complex as compared to simple passages. We related these WTI measures to individual differences in reading comprehension.

To summarize, in the present study we addressed the following two questions:

1. How are the effects of WTI-complexity (simple versus complex) on self-paced reading times in different word positions (target, target plus one, target plus two; Bultena et al., 2015) reflected in different aspects of WTI (anaphora, argument overlap, and anomalies) over time, after controlling for word frequency and students' decoding fluency, gender, and age?
2. How does WTI, reflected by the average additional reading time required per participant to read complex as compared to simple passages for each text manipulation (anaphora, argument overlap, and anomalies) and word position (target, target plus one, target plus two), relate to reading comprehension?

Our hypotheses were as follows:

1. Self-paced reading times are longer at T1 than at T2, and for complex than for simple passages, and systematically varied across word positions, with different patterns for the three types of text manipulation: We expected an immediate effect of complexity on the target word for argument overlap and anomaly detection passages, but an effect after the target for anaphora passages. Further, we expected higher word frequency and stronger decoding skills to be related to shorter self-paced reading times, whereas lower word frequency and poorer decoding skills resulted in longer self-paced reading times, after controlling for multicollinearity.
2. Larger WTI-indices, i.e. longer average reading times on complex as compared to simple passages, are related to better reading comprehension.

## Methods

### Participants

The data were collected at seven schools in the Netherlands among 503 7th grade students. From the sample, data of the 441 students (238 boys and 203 girls) that completed all measures at T1 (November 2016) and T2 (April 2017) were included in the analyses. Students were between eleven and thirteen years old (mean = 12;3,

$SD=6$  months). The participants were part of the Dutch tracked school system, in which they were divided into the following tracks: lower and intermediate pre-vocational education of secondary education, intermediate education, or higher level of secondary education and pre-university education. All participants had also received English as a second language (L2) instruction in primary school, which focuses on communicative language teaching. Their formal English language instruction within secondary education, which combines communicative language teaching with elements of language awareness, had started three months prior to the onset of this study. Thus, at T1, participants had received three months of L2 English instruction in secondary education; at T2, this period had increased to eight months in total. Parents of all students were informed of the study and were at liberty to refuse their child's participation.

## Materials

### Word-to-text integration

Participants performed a computerized Word-to-text Integration (WTI) task, programmed in Inquisit 4 (2015), administered through silent self-paced reading. Figure 1 displays the design of the task. In the Figure, the target word is underlined in each passage and printed in bold and in italics. Three types of WTI text manipulation were included: anaphora resolution (syntactic integration), argument overlap, and anomaly detection (semantic integration). For each type of manipulation, we created simple and complex passages. It was proposed that WTI be reflected by the additional self-paced reading time needed to read complex compared to simple passages for each WTI text manipulation and word position. To examine this reading time effect, for every participant we calculated indices for each text manipulation, word position, and time dividing the reading time on complex passages per item by the mean reading time on all simple passages. This is explained in more detail in the following example:

Participant 1:

*The dean likes the secretary of the professors who is reading a letter (complex, anaphora, time 1)*

Reading time for Participant 1 reading the bold target word in the complex passage above: 500 ms.

Mean reading time for all target words in simple anaphora passages at Time 1: 300 ms.

WTI index:  $500/300=1.67$ .

The aforementioned calculation resulted in the following WTI indices: anaphora targets, anaphora targets plus one, anaphora targets plus two, argument overlap targets, argument overlap targets plus one, argument overlap targets plus two, anomalies targets, anomalies targets plus one, and anomalies targets plus two for Time 1 and Time 2. We calculated the average of the index on each word position per text

manipulation, after calculating the WTI-index for each text manipulation and word position separately. As a result, the index for each word position separately consisted of 12 scores, and the index for a text manipulation (the average across word positions) consisted of three indices per word position.

We used a within-subjects, between-items design, in which students were either provided with the simple or the complex passage at T1 and vice versa at T2. The passages were presented in a mixed order, and randomized across text types and complexities. We created an 'order' variable to control for effects of order in complexity. Furthermore, a *Complexity across Time* variable was created with four levels: T1\_simple, T1\_complex, T2\_simple, T2\_complex.

To verify whether students were actively reading, they answered comprehension questions after each passage; each WTI text manipulation had its own type of comprehension question. We specifically looked at reading times on critical words, i.e., target word (target), the word following that word (target plus one), and the word after that (target plus two), as in Bultena et al. (2015). Responses on the comprehension questions were coded correct or incorrect. Construct validity of the self-paced reading task was assumed, as texts were largely derived from previous studies.

### Sentence passage construction

For each type of WTI text manipulation, twelve simple and twelve complex passages were constructed. Pairs of simple and complex passages were constructed to always be identical, except for the target word, which was either simple or complex. In each manipulation, passages were constructed with the goal of invoking either simple or complex WTI-processes. In the analyses, we controlled for word length and word frequency of target words (target, target plus 1, target plus 2) as well as passage length. Word frequency and passage length were entered into the multilevel models as independent variables (with passage length not being a significant predictor of the outcome variable).

The anaphora resolution passages always consisted of one sentence and were derived from a study by Dussias (2003), which targeted Spanish learners of English as an L2. The target word was the single anaphor that the sentence contained, and hence the word that required anaphoric resolution to take place. Both simple and complex anaphora passages consisted of a noun phrase (for example: 'the dean'), followed by a verb phrase ('likes'), followed by another noun phrase ('the secretary of the professors who is reading a letter'). In the complex passages, the embedded sentence 'who is reading a letter' is attached to a distant anaphor ('the secretary of the professors') and in the simple version the embedded sentence is attached to a proximate anaphor ('the professors'). Simple passages contained short-distance anaphoric relations (proximate anaphora; for example: *The dean likes the secretary of the professors who are reading a letter*), whereas complex passages contained long-distance relations (distant anaphora; for example: *The dean likes the secretary of the professors who is reading a letter*). Each passage was eleven to sixteen words long. The target word was placed on the tenth or eleventh position of the passage ( $M = 10.17$ ,  $SD = 0.38$ ), following Dussias (2003). Reliability for anaphora reading times was  $\alpha = 0.72$ , which can be considered acceptable (Kline, 2013).

The argument overlap passages were adapted from a study by Yang et al. (2007) targeting English as an L1 adults, and always consisted of two sentences. The target was the word that required inferencing as a result of argument overlap. The passages consisted of two sentences and were twelve to nineteen words long. The second sentence always contained the target word at the beginning of the sentence. The syntactic structure of a pair of simple and complex passages was always identical. Furthermore, the first sentence of the passages was also always identical between the simple and complex version of a passage. In the simple passages, only familiar words were included, and these words were presented as explicit repetitions of the same words earlier in the same text (for example: *After being dropped from the plane, the bomb hit the ground and exploded. The explosion was quickly reported to the commander.*). Each complex passage included an unfamiliar target word, which was presumed to be unexpected based on low word frequency and understanding these words required implicit inferencing (for example: *After being dropped from the plane, the bomb hit the ground and exploded. The detonation was quickly reported to the commander.*). The target word was always placed in the second sentence in the text passage and was between the eighth and the seventeenth position ( $M=11.38$ ,  $SD=2.66$ ), following Yang and colleagues (2007). Reliability for argument overlap reading times was  $\alpha=0.79$ , which can be considered acceptable (Kline, 2013).

Passages that required anomaly detection were constructed for the purpose of this study and always consisted of one sentence. Syntactic structure was always identical between the simple and complex version of a passage. Most passages started with a noun phrase combined with a verb phrase. Some sentences started with a prepositional phrase (e.g., item 27 and item 37). The target word position was the position where a violation could be present or absent. Simple passages did not include an anomaly (for example: *The man with the umbrella walked through the rain alone.*), but complex passages did include an anomaly (for example: *The man with the umbrella walked through the lie alone.*). Passages were seven to fourteen words long. The target word was placed between the fourth and the tenth position in the passage ( $M=7.03$ ,  $SD=2.37$ ), trying to pursue placing words in the sentence-final position if possible (following e.g., Elgort, Perfetti, Rickles, & Stafura, 2015). Reliability for anomaly detection reading times was  $\alpha=0.72$ , which can be considered acceptable (Kline, 2013).

### Comprehension questions

Each passage was followed by a multiple-choice comprehension question. The comprehension questions differed across the types of WTI text manipulation: after the anaphora resolution passages, students had to choose out of four options to whom the verb in the passage referred, i.e. ‘who [verb phrase]?', following Dussias (2003). For example, after the passage: ‘The doctor contacts the nurses of the lawyer who are talking on the phone.’, the comprehension question was: ‘Who talks on the phone?’. Out of four options students had to choose the right answer. Reliability of the anaphora resolution comprehension questions was  $\alpha=0.69$ . The argument overlap passages were followed by a question that required participants to select the correct translation of the target word out of four options. For example, after the passage:

'The trapeze artist was very good, but tonight he fell. The plunge resulted in a broken leg', the question was: 'What does 'plunge' mean?'. Students had to choose the right answer out of four options. Reliability of the argument overlap comprehension questions was  $\alpha=0.69$ . After the anomaly detection passages, students were asked to judge the plausibility of the passage. For example, after the passage: 'On our way to the island we took the joke to the other side.', the question was: 'Is this passage plausible?'. Students could choose between plausible and implausible. Reliability of the anomaly detection comprehension questions was  $\alpha=0.67$ . Details concerning the stimulus passages can be found in "[Appendix 1](#)".

### Decoding fluency

Decoding fluency of the passages was derived from the word-by-word self-paced reading task. Decoding fluency was calculated by looking at the average reading time on each separate word preceding the target word, not on target words themselves. Hence, there was no overlap between decoding fluency and the reading times on the target words. All reading times were included, regardless of whether words were read correctly or incorrectly. Reliability of decoding fluency was  $\alpha=0.86$  for anaphora,  $\alpha=0.94$  for argument overlap, and  $\alpha=0.94$  for anomalies.

### Reading comprehension

English reading comprehension skills were measured using a nationally standardized reading comprehension test, normed on final-year students in pre-vocational education (College voor Toetsen en Examens—Board for Assessment and Exams, 2016). Students read three different texts. For each text, student had to answer multiple-choice questions with three to five options and/or open-ended questions, such as: 'How does the writer introduce the topic in paragraph 1?'. In total, the test consisted of thirteen items. All materials can be found in "[Appendix 2](#)". Reliability of the reading comprehension measure was  $\alpha=0.66$ .

### Procedure

Participants were selected based on a convenience sample in a larger longitudinal study. The data were collected around November 2016 and around April 2017. At both time points, students were tested in a 45-min individual session and two 50-min plenary classroom sessions. WTI was measured during the individual session and reading comprehension during the second classroom session. Both tasks took approximately fifteen minutes.

During the WTI-task, students were seated approximately 30 cm away from the computer screen and were presented with words in Consolas font; further, they were instructed to read carefully and silently through the passages, at a normal pace, without trying to memorize the passages. They were told that they would have to answer a comprehension question after each passage. After the instruction, students were presented with practice trials (one of each type of WTI text manipulation),

which resembled the experimental trials. After finishing the practice trials, participants were allowed to ask questions. After completing half of the passages, students received a one-minute break. The trials were built up as follows: Students were presented with a screen that had a dash to represent each word of the passage. Participants were presented with a passage one word at a time, and were instructed to press the space bar as soon as they had read the new word. As a result, this word would disappear and the next word would appear. After completing a trial, a comprehension question appeared, which students could answer by pressing 1–4.

## Analyses

To measure WTI, reading times for each word in the passage were recorded from the moment the word was presented. Responses to the comprehension questions were also registered. We looked at reading times on the critical words: target word (target), the word following that word (target plus one), and the word after that (target plus two; Bultena et al., 2015). Responses on the comprehension questions were coded correct or incorrect. After this, the data were analyzed using R, version 3.5.1 (R Core Team, 2018). Mixed effects models were fitted, using the logit link function (e.g., Breslow & Clayton, 1993; Jaeger, 2008) and lme4 (Bates, Maechler, Bolker, & Walker, 2015). Regression assumptions were checked: word frequency and decoding fluency were orthogonalized as a correction for multicollinearity (Wurm & Fisicaro, 2014). To do so, we created a linear model for frequency and a model for decoding fluency and saved the residuals from the linear model. The frequency model had log frequency as the dependent variable and Complexity across Time and Word Position as predictors, because frequency could vary between the levels of Complexity across Time and Word Position. The decoding fluency had decoding as the dependent variable and Complexity across Time as the predictor, because decoding fluency was significantly better at Time 2 than at Time 1. To control for outliers, reading times for which the standardized residuals were larger or smaller than 2.5 were cut after fitting the model (Baayen, 2008) following, for example, Viebahn, McQueen, Ernestus, Frauenfelder, & Bürki (2018). Furthermore, model residuals were normally distributed for anaphora and anomaly passages, but not for the argument overlap passages. Therefore, profile confidence intervals are reported for each of the three different manipulations, which were similar to non-bootstrapped confidence intervals (Bates et al., 2015). Finally, residuals for different random effects were all distributed normally.

To examine whether inclusion of a variable lead to a significantly better model fit, Chi-square tests were used. Additionally, we examined whether Akaike Information Criterion (AIC) values of these models were lower after inclusion of a variable. After the inclusion of the fixed effects, random intercepts (*Item* and *Participant*) and then random slopes (*Complexity across Time*) were added to the model and significance was tested using the same procedure as for the fixed effect (Baayen, 2008). Effect size was indicated by the size of the beta coefficient.

As we expected different progress times for the reading processes involved in processing the critical words in the anaphora resolution, argument overlap and anomaly detection passages, we created separate models for each type of WTI text manipulation. In all three models, we assessed to what extent task complexity and word frequency, affected reading times on the critical words.

The dependent variable in each of the three WTI text manipulation models was reading time on the three critical words. We included several control predictors, which were centered if numerical, and used contrast coding for factors. First, for *Word Position* (target, target plus one, target plus two) word position (target) served as the intercept level. Second, for *Frequency* word frequencies for target, target plus one, and target plus two were obtained from the Corpus of Contemporary American English (Davies, 2008). Finally, during the self-paced reading task, we recorded what *Trial* students were on. This way, we could examine whether students' reading times changed during the computerized WTI-task. *Educational Track* (lower and intermediate pre-vocational education, intermediate education, or higher-level of secondary education and pre-university education), *Word length* and *Passage Length* did not improve the model fit. Further, we included the factor *Complexity across Time* (T1 simple, T1 complex, T2 simple, T2 complex), with T1-simple as the intercept, and the student variable *Decoding Fluency* (average reading time for the words preceding the target). Finally, we added *Gender* and *Age* as control variables. In addition, we explored whether there were interactions between task and student characteristics. Given Occam's razor, which favors parsimonious models (Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987), we applied a backward stepwise regression procedure, in which predictors were removed if they were not significant at the 5% level.

To examine how WTI was related to reading comprehension, we calculated WTI process scores based on the raw reading times per text manipulation (*Anaphora Resolution*, *Argument Overlap*, *Anomaly Detection*) for each time (T1, T2) and word position (target, target plus 1, target plus 2) for every participant separately. As different reflections of WTI, for each text manipulation on each word position, we divided the average reading times on complex passages by the average reading times on simple passages. This resulted in reading time scores for every participant for Anaphora on Target, Anaphora on Target Plus 1, Anaphora on Target plus 2, Argument Overlap on Target, Argument Overlap on Target Plus 1, Argument Overlap on Target plus 2, Anomaly on Target, Anomaly on Target Plus 1, and Anomaly on Target plus 2. These indices indicate the additional time needed to process complex as compared to simple passages. We created a linear model with offline reading comprehension as the dependent variable, and the WTI scores as the independent variables.

**Table 1** Overview of means and (standard deviations) of raw self-paced reading times in milliseconds on critical words across the three WTI text manipulations, accuracy scores on comprehension questions and decoding fluency

	Target			Target plus one			Target plus two		
	Simple	Complex	Complex	Simple	Complex	Complex	Simple	Complex	Complex
<i>Anaphora</i>									
T1	586.04 (374.61)	571.3 (375.8)	648.39 (421.22)	664.99 (460.29)	593.51 (369.82)	598.07 (371.4)	593.51 (369.82)	664.99 (460.29)	598.07 (371.4)
T2	511.06 (242.16)	501.4 (263.7)	565.16 (304.44)	568.43 (358.24)	524.77 (262.46)	522.91 (294.37)	524.77 (262.46)	568.43 (358.24)	522.91 (294.37)
Decoding	665.69 (323.42)								
Comprehension	5.82 (2.41)								
<i>Argument overlap</i>									
T1	654.65 (433.74)	993.29 (827.31)	597.01 (336.56)	666.57 (350.85)	597.82 (318.44)	616.3 (313.63)	597.82 (318.44)	666.57 (350.85)	616.3 (313.63)
T2	551.14 (267.31)	853.83 (777.62)	527.84 (275.38)	593.83 (546.3)	530.61 (253.91)	548.68 (245.41)	530.61 (253.91)	593.83 (546.3)	548.68 (245.41)
Decoding	654.68 (303.98)								
Comprehension	8.51 (2.33)								
<i>Anomalies</i>									
T1	698.24 (480.08)	720.19 (453.34)	610.42 (364.25)	669.42 (563.68)	584.14 (354.28)	647.07 (415.13)	584.14 (354.28)	669.42 (563.68)	647.07 (415.13)
T2	600.98 (373.89)	638.58 (416.3)	533.41 (266.65)	590.45 (342.01)	522.95 (302.24)	565.54 (353.97)	522.95 (302.24)	590.45 (342.01)	565.54 (353.97)
Decoding	666.73 (299.3)								
Comprehension	8.35 (2.26)								

**Table 2** Summary of a generalized linear mixed-effects model predicting reading times for anaphora passages

Predictors	Fixed effects			Random effects				Chi square	p
	β	SE	t	Participant		Item			
				Variance explained	p	Variance explained	p		
Intercept	6.302	0.011	559.231*	0.020	4421.5	< 0.001	< 0.001	146.66	< 0.001
Trial	-0.004	<0.001	-25.172*	-	-	-	-	-	-
Word position (target plus 1)	0.090	0.004	25.667*	-	-	-	-	-	-
Word position (target plus 2)	0.030	0.004	7.876*	-	-	-	-	-	-
Complexity across time (T1-distant)	-0.003	0.004	-0.810	-	-	-	-	-	-
Complexity across time (T2-proximate)	-0.121	0.004	-28.623*	-	-	-	-	-	-
Complexity across time (T2-distant)	-0.124	0.004	-30.210*	-	-	-	-	-	-
Word frequency	-0.274	0.017	-15.931*	-	-	-	-	-	-
Decoding fluency	0.168	0.002	71.583*	-	-	-	-	-	-
Gender (female)	0.048	0.014	3.489*	-	-	-	-	-	-

\*Significant at <0.05

**Table 3** Summary of a generalized linear mixed-effects model predicting reading times for argument overlap passages

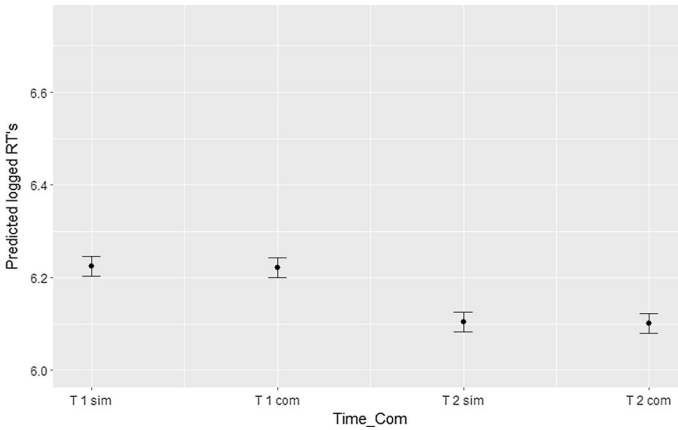
Predictors	Fixed effects				Random effects						
	$\beta$	SE	<i>t</i>	Participant	Variance explained	Chi Square	<i>p</i>	Item	Variance explained	Chi Square	<i>p</i>
Intercept	5.964	0.235	25.358*	0.027	3024.2	< 0.001	848.85	< 0.001			
Trial	-0.006	< 0.001	-28.144*	-	-	-	-	-	-	-	-
Word position (target plus 1)	-0.076	0.010	-7.782*	-	-	-	-	-	-	-	-
Word position (target plus 2)	-0.067	0.010	-6.910*	-	-	-	-	-	-	-	-
Complexity across time (T1-implicit)	0.274	0.011	24.614*	0.009	917.24	< 0.001	-	-	-	-	-
Complexity Across time (T2-explicit)	-0.153	0.012	-13.112*	-	-	-	-	-	-	-	-
Complexity Across time (T2-implicit)	0.119	0.013	9.070*	-	-	-	-	-	-	-	-
Decoding fluency	0.116	0.003	37.114*	-	-	-	-	-	-	-	-
Gender (female)	0.049	0.016	3.138*	-	-	-	-	-	-	-	-
Age	0.040	0.019	2.162*	-	-	-	-	-	-	-	-
Word position (target plus 1) × complexity across time (T1-implicit)	-0.164	0.014	-11.622*	-	-	-	-	-	-	-	-
Word position (target plus 2) × complexity across time (T1-implicit)	-0.237	0.014	-16.833*	-	-	-	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-explicit)	0.041	0.014	2.989*	-	-	-	-	-	-	-	-
Word position (target plus 2) × complexity across time (T2-explicit)	0.047	0.014	3.397*	-	-	-	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-implicit)	-0.124	0.014	-8.817*	-	-	-	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-implicit)	-0.180	0.014	-12.841*	-	-	-	-	-	-	-	-

\*Significant at <0.05

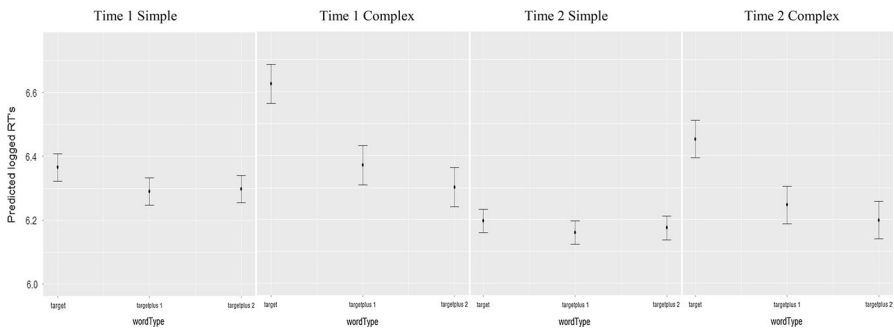
**Table 4** Summary of a generalized linear mixed-effects model predicting reading times for anomalies passages

Predictors	Fixed effects			Random effects				
	$\beta$	SE	<i>t</i>	Participant		Item		
				Variance explained	Chi square	Variance explained	Chi square	
				<i>p</i>		<i>p</i>		
Intercept	6.455	0.016	396.481*	0.031	842.86	<.001	232.46	< 0.001
Trial	-0.005	< 0.001	-31.190*	-	-	-	-	-
Word position (target plus 1)	-0.081	0.008	-10.636*	-	-	-	-	-
Word position (target plus 2)	-0.123	0.008	-15.781*	-	-	-	-	-
Complexity across time (t1-anomaly)	0.039	0.008	4.741*	0.004	232.46	<.001	-	-
Complexity across time (t2-no anomaly)	-0.130	0.010	-12.622*	-	-	-	-	-
Complexity across time (t2-anomaly)	-0.093	0.011	-8.742*	-	-	-	-	-
Word frequency	-0.497	0.011	-44.465*	-	-	-	-	-
Decoding fluency	0.144	0.003	53.589*	-	-	-	-	-
Gender (female)	0.044	0.016	2.788*	-	-	-	-	-
Word position (target plus 1) × complexity across time (T1-anomaly)	0.007	0.011	0.682	-	-	-	-	-
Word position (target plus 2) × complexity across time (T1-anomaly)	0.015	0.011	1.368	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-no anomaly)	0.017	0.011	1.599	-	-	-	-	-
Word position (target plus 2) × complexity across time (T2-no anomaly)	0.021	0.011	1.956	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-anomaly)	0.038	0.011	3.517*	-	-	-	-	-
Word position (target plus 1) × complexity across time (T2-anomaly)	0.033	0.011	3.011*	-	-	-	-	-

\*Significant at <0.05



**Fig. 2** Effect of complexity across time on reading times for anaphora passages. On the X-axis four bars are displayed, representing the complexity across time: respectively time 1 simple (T1 Sim), time 1 complex (T1 Com), time 2 simple (T2 Sim), and T2 complex (T2 Com). On the Y-axis predicted logged reading times are displayed

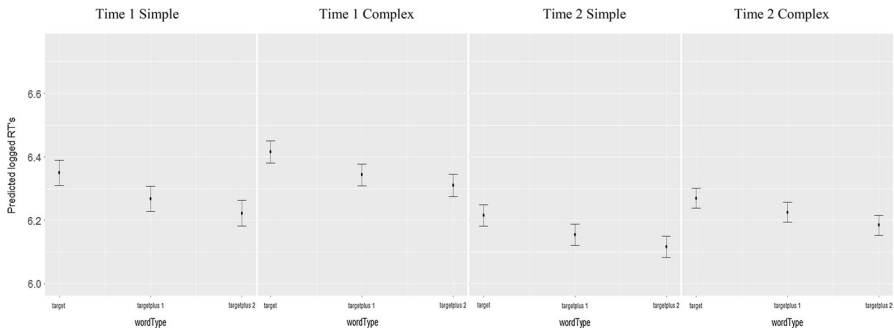


**Fig. 3** Two-way interaction between word position and complexity across time for argument overlap passages. On the X-axis three bars are displayed, representing the word position: respectively target, target plus one, and target plus two. On the Y-axis predicted logged reading times are displayed

## Results

### Task and student predictors of WTI

Students' average word reading times on the critical words (target word, target plus one, and target plus two) divided across types of WTI text manipulation are displayed in Table 1. Separate reading time models were fitted for anaphora resolution, argument overlap, and anomaly detection passages respectively to examine how WTI could be predicted. Tables 2, 3 and 4 give a summary of each generalized linear model with an overview of fixed and random effects, including the amount of variance explained by random effects. The results are graphically represented in Figs. 2, 3 and 4. In these figures, the y-axis refers to logged reading times as predicted by the



**Fig. 4** Two-way interaction between word position and complexity across time for anomaly passages. On the X-axis three bars are displayed, representing the word position: respectively target, target plus one, and target plus two. On the Y-axis predicted logged reading times are displayed

mixed effects model. On the x-axis, three bars are displayed, representing the word position: respectively target, target plus one, and target plus two.

All three models showed significant main effects of *Trial*, anaphora,  $b = -0.004$ , 95% confidence interval (CI)  $[-0.004, -0.003]$ ; argument overlap,  $b = -0.006$ , 95% CI  $[-0.006, -0.005]$ ; anomalies,  $b = -0.005$ , 95% CI  $[-0.005, -0.004]$ , *Decoding Fluency*, anaphora,  $b = 0.168$ , 95% CI  $[0.163, 0.172]$ ; argument overlap  $b = 0.116$ , 95% CI  $[0.109, 0.122]$ ; anomalies  $b = 0.144$ , 95% CI  $[0.138, 0.149]$  and *Gender*, anaphora,  $b = 0.048$ , 95% CI  $[0.021, 0.075]$ ; argument overlap  $b = 0.049$ , 95% CI  $[0.018, 0.080]$ ; anomalies  $b = 0.044$  95% CI  $[0.012, 0.074]$ . The *Trial* effect suggests that students read more slowly at task onset, and faster as they progressed through the task. The effect of *Decoding Fluency* indicates that students with better decoding fluency on the words preceding the target also read faster on the critical words (target, target plus one, target plus two). The *Gender* effect suggests that girls read slower than boys. Further, there was a main effect of *Frequency* on reading times for anaphora resolution,  $b = -0.274$ , 95% CI  $[-0.307, -0.240]$ , and anomaly detection passages,  $b = -0.497$ , 95% CI  $[-0.518, -0.474]$  which indicated that students read faster for more frequent words. The remaining findings (for *Word Position* and *Complexity across Time*) differed across the types of text manipulation and will thus be discussed separately for each text manipulation.

### Anaphora resolution

For anaphora passages, there were main effects of *Word Position* and *Complexity across Time*. The effect of *Word Position* indicated that reading times were quicker on the target compared to the target plus one and target plus two,  $b = 0.090$ , 95% CI  $[0.083, 0.097]$ . The main effect of *Complexity across Time* is visually displayed in Fig. 2. This effect suggested that reading times were slower on T1 than on T2, regardless of word position or passage complexity,  $b = -0.810$ , 95% CI  $[-0.011, 0.004]$ . It seems there were no differences in reading times on simple versus complex passages for anaphora. In other words, for anaphora passages reading times were shorter on higher frequency word and when decoding fluency (reading times

on words preceding target) was better; boys read quicker than girls. There appears to be no complexity effect on reading times and, regardless of complexity, reading times are longest on target plus two for anaphora passages.

### Argument overlap

For the argument overlap passages, we found a main effect of *Age* main effects of and a two-way interaction between *Word Position* and *Complexity across Time*. The effect of *Age* indicates that older students read slower than younger students,  $b = 0.040$ , 95% CI [0.003, 0.076]. The main effects of and two-way interaction between *Word Position* and *Complexity across Time* are shown graphically in Fig. 3. Importantly, the effects indicated that reading times were slower at T1 than at T2 and especially for target words (compared to target plus one and target plus two) in complex compared to simple passages,  $b = -0.164$ , 95% CI [-0.191, -0.136]. To summarize, for argument overlap passage reading times were shorter for students with better decoding fluency skills, for younger students, and for boys compared to girls. Furthermore, reading times were highest on the target word in complex passages, whereas for simple passages reading times remained similar across the word position.

### Anomaly detection

For the anomaly detection passages, we again found main effects of and an interaction between *Word Position* and *Complexity across Time*, which is visually displayed in Fig. 4. These results indicated slower reading times at T1 than at T2 and for complex compared to simple passages and there was a slightly 1 delay in reading times on target in complex passages at T2 compared to the delay on target in complex passages at T1,  $b = 0.007$ , 95% CI [0.022, 0.054]. To summarize, students reading times were shorter for high frequency words, when they had better decoding fluency skills, and for boys compared to girls. Reading times looked similar across complexity, but were higher on T1 than on T2, and on target.

### WTI predicting reading comprehension

To examine the relationship between WTI and reading comprehension, we created an index dividing reading times on complex by reading times on simple passages for each participant for the three types of WTI text manipulations, as described above. We then fitted a linear model with offline reading comprehension as the dependent variable and the WTI indices as the predictors. Table 5 shows the descriptive statistics of the WTI indices at Time 1 and correlations with reading comprehension at Time 2. The final model, presented in Table 6, indicated significant effects of complexity (additional time needed to read complex compared to simple passages). The results suggested that longer reading times

**Table 5** Means and (standard deviations) of WTI process scores (reading times complex/reading times simple) at time 1 on critical words across the three WTI text manipulations, and correlations between the WTI indices, and reading comprehension at time 2 (10)

	M (SD)	1	2	3	4	5	6	7	8	9	10
1. Anaphora, target	1.00 (0.03)	–									
2. Anaphora, target plus one	1.00 (0.04)	0.28***	–								
3. Anaphora, target plus two	1.00 (0.03)	0.21***	0.18***	–							
4. Argument overlap, target	1.04 (0.05)	–0.02	0.08	0.11*	–						
5. Argument overlap, target plus one	1.02 (0.03)	0.11*	0.14**	0.01	0.30***	–					
6. Argument overlap, target plus two	1.00 (0.03)	–0.01	0.07	–0.06	0.24***	0.29***	–				
7. Anomalies, target	1.01 (0.03)	–0.10*	0.07	0.03	0.05	0.03	–0.08	–			
8. Anomalies, target plus one	1.01 (0.03)	–0.08	0.09	0.05	0.11*	0.14**	0.03	0.30***	–		
9. Anomalies, target plus two	1.01 (0.04)	–0.13**	0.09	–0.06	0.16***	0.14**	0.10*	0.16***	0.33***	–	
10. Reading comprehension	6.22 (2.64)	–0.02	–0.01	0.03	0.11*	0.15**	0.06	0.06	0.10*	0.17***	–

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

**Table 6** Linear model of reading comprehension predicted by WTI indices for argument overlap and anomaly detection on target, target plus 1, and target plus 2

Variable	B	SE (B)	<i>t</i>	Sig.( <i>p</i> )	<i>R</i> <sup>2</sup>
Constant	-19.089	5.648	-3.380	<0.001	
Argument overlap target	2.672	2.817	0.94	<i>n.s</i>	
Argument overlap target plus 1	9.242	4.116	2.246	<0.05	
Anomaly detection target plus 1	3.089	4.036	0.765	<i>n.s</i>	
Anomaly detection target plus 2	9.870	3.632	2.717	<0.05	
Total					0.04

for complex argument overlap passages on target plus one and anomaly detection passages on target plus two, relative to simple passages related to stronger reading comprehension skills. The indices for anaphora passages did not significantly predict reading comprehension. A small degree of variance was explained by the model (4%). In other words, students that showed larger processing costs for complex argument overlap and anomaly passages compared to their simple versions, also showed higher reading comprehension scores.

## Discussion

The aim of the present study was to examine how Word-to-Text Integration (WTI) abilities could be measured in early English as a second language (L2) learners by means of a computerized self-paced reading task. We provided a longitudinal perspective on the relationship between three different WTI manipulations with two levels of complexity (simple versus complex) across time, and decoding fluency, controlling for word frequency. The WTI text manipulations were syntactic or semantic in nature. Specifically, they were anaphora resolution: proximate anaphora (simple) versus distant anaphora (complex); argument overlap: explicit repetitions (simple) versus implicit inferences (complex), and anomaly detection: passages without anomalies (simple) versus passages with anomalies (passages with anomalies (complex)). Subsequently, for every participant we created WTI indices for each of the three text manipulations by dividing reading times on complex passages by reading times on simple passages. With these indices, we examined to what extent WTI, as reflected by larger processing costs for complex as compared to simple passages, predicted reading comprehension. A complexity effect was present for argument overlap and anomalies passages, but not for anaphora resolution. Longer reading times for complex (as compared to simple) argument overlap and anomalies (versus continuous) passages were related to offline reading comprehension, and as such could be regarded as an index of WTI proficiency.

Specifically, the first research question was how complexity is reflected in WTI, after controlling for word frequency, students' decoding fluency, gender, and age across the three types of WTI text manipulation. In anaphora resolution passages, there was no significant complexity effect. In argument overlap passages, students

slowed down on target, compared to target plus one and target plus two, especially for complex (implicit inferences) rather than simple (explicit repetitions) passages, suggesting a complexity effect. Similar findings for argument overlap passages in (adult) L2 readers were established by Yang and colleagues (2005, 2007), who interpreted these effects as delayed patterns of integration compared to L1 learners. For anomaly detection passages, we also found effects of complexity mainly for the target word. Findings from the anomalies passages suggest that the immediate effect of anomaly detection, as reflected by higher reading times on the target words, become stronger across time. When the participants were confronted with an anomaly, their reading speed seemed to slow down on the target word and stabilize on the following words. This may be explained by the fact that the early L2 learners in the present study did not only have to inhibit their L1, but they also had to process an anomaly. L1 learners, on the other hand, would only have to process an anomaly, without inhibiting another language (Bialystok, 2009; Hagoort, 2017).

An explanation for the absence of the effect of time on the anaphora resolution passages may be that both the simple and complex passages were very difficult and hence little progress was to be expected. These novice L2 learners have had little exposure to complex syntactic constructions, and therefore may not show a preference to proximate anaphora constructions, whereas more skilled L2 learners, and L1 learners do show this preference (Dussias, 2003).

Online reading times on the critical words seemed to be shorter for relatively high frequency words in anaphora resolution and anomaly detection passages, which is consistent with previous research with L1 learners (e.g., Diependaele et al., 2013; Whitford & Titone, 2017). No such effects were found for argument overlap passages. The absence of frequency effects for argument overlap passages may be explained by the repetition of lexical items in the simple condition, which has been shown to attenuate frequency effects (Clifton et al., 2007). Across the three types of WTI text manipulation, higher decoding fluency was related to shorter reading times on the critical words. Previous studies, focusing on adult L1 and L2 learners, showed higher decoding fluency to be related to better reading comprehension (e.g., Hagoort, 2017; Hoover & Gough, 1990). Furthermore, decoding fluency was found to be a significant predictor of reading comprehension in young Dutch L1 learners (de Jong & van der Leij, 2002). We elaborated on that with our finding that decoding fluency seemed to be related to the WTI process across three different types of WTI text manipulation.

As a second research question, we assessed how WTI is related to reading comprehension. We assumed convergent validity was ensured by relating our WTI measure to reading comprehension (Verhoeven & Perfetti, 2008). First, our results suggest that longer reading times for complex compared to simple argument overlap passages on target plus one and on passages with anomalies, compared to without anomalies on target plus two, were related to better reading comprehension. In other words, students who show longer reading times for complex compared to simple passages, also seem to be better at reading comprehension. This is in line with, for example, findings by Barnes & colleagues (2015), who found that less skilled readers showed less sensitivity to contextual cues. Furthermore, in contrast to L1 learners, L2 learners may not benefit from supportive text-based factors (Degand &

Sanders, 2002). Processing the cues takes time and therefore skilled readers often take more time. This explains why longer reading times for complex compared to simple passages are related to better reading comprehension. Second, the relationship between the WTI indices for anaphora resolution passages and reading comprehension appeared to be absent. This may be explained by the fact that syntactic integration may be dependent on lexical access (Segers & Verhoeven, 2016), which was not controlled for in the present study.

The limitations of this study are the fact that we only had 12 items per text type per complexity. Although we provide evidence that WTI is reflected by longer reading times on complex compared simple passages, it remains unclear how large this difference should be in order to arrive at adequate integration. Future research could focus on different profiles of WTI and how these are related to reading comprehension. Another challenge was a difference in length of the passages across the text types. Namely, the inferencing passages consisted of multiple sentences, whereas in the other manipulations, these consisted of a single sentence. In future research, it would be interesting to also include single sentence argument overlap or multiple sentence anaphora and anomaly passages. Furthermore, after the anomaly passages students were asked to judge plausibility of the passage, and it could be argued that this stimulates students to read the passages using a certain strategy, rather than targeting comprehension (e.g., Cain & Oakhill, 1999). The comprehension questions asked as after the WTI-passages were always identical for each text manipulation, whereas in the reading comprehension task questions differed between the different texts, dependent on the content. Further elaboration of relevant text features, such as syntactic complexity, could also be considered. Furthermore, a limitation of the self-paced reading paradigm is that we were unable to examine students' rereading behavior, whereas previous studies did take this into account (e.g., Clifton et al., 2007). We recommend future research to use the measures of WTI we derived as a predictor of WTI in combination with other measures, such as standardized measures of decoding (e.g., Torgesen, 2012), vocabulary knowledge (e.g., Ouellette, 2006), and language proficiency.

Implications of the present study are that we created a multifaceted measure of WTI, using a WTI-index, with which insight in the development of WTI can be gained and can be related to offline reading comprehension. It must be noted that only a small amount of variance in reading comprehension was explained by WTI. The WTI-index measure turned out to be suitable for a group of early L2 learners of English, for whom L2 reading is often a challenge (Lesaux et al., 2006) which has hardly been examined. Previous studies often focused on either anaphora resolution (Dussias, 2003), argument overlap (Yang et al., 2005), or anomalies (Hagoort et al., 1999). The current study, however, combined these three different text types in one task, to provide a multifaceted perspective on L2 WTI. This WTI-index measure is easily applicable within a school setting. Furthermore, while previous studies demonstrated what the WTI process looks like in younger children learning Dutch as an L2 (van den Bosch et al., 2018; Raudszus, Segers, & Verhoeven, 2018, 2019) and adult learners (Calloway & Perfetti, 2017; Helder et al., 2019; Stafura et al., 2015), the current study added to this

body of literature demonstrating three types of integration in 7<sup>th</sup> grade English as an L2 learners. To conclude, we provided a perspective on word-to-text integration in early English L2 learners. We found longer reading times for complex compared to simple argument overlap and anomalies passages, reflected in a manageable WTI-index, to be related to better reading comprehension.

**Acknowledgements** This research was supported by Grant nr 405-14-304 from the NRO Programme Council for Educational Research (PROO). We thank all university students, participants, schools, and staff that helped to make this project possible. Finally, we thank Charles Perfetti and Bernard Westwell for their useful advice.

**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/>.

## Appendix 1

Stimuli used in the computerized self-paced reading task to measure Word-to-Text Integration. For each item, the complex passage and corresponding critical words (target, target plus 1, target plus two) and after that the simple passage with the critical words are displayed.

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
<i>Anaphora</i>									
1	The dean likes the secretary of the professors who is reading a letter	is	reading	a	2	The dean likes the secretary of the professors who are reading a letter	are	reading	a
3	The young girl loves the driver of the players who is talking to an old woman	is	talking	to	4	The young girl loves the driver of the players who are talking to an old woman	are	talking	to

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
5	The doctor looks at the nurse of the patients who is feeling very tired	is	feeling	very	6	The doctor looks at the nurse of the patients who are feeling very tired	are	feeling	very
7	The director helps the teacher of the school-boys who is writing the reports	is	writing	the	8	The director helps the teacher of the school-boys who are writing the reports	are	writing	the
9	The doctor contacts the nurses of the lawyer who are talking on the phone	are	talking	on	10	The doctor contacts the nurses of the lawyer who is talking on the phone	is	talking	on
11	The secretary sees the drivers of the manager who are dreaming of holidays	are	dreaming	of	12	The secretary sees the drivers of the manager who is dreaming of holidays	is	dreaming	of
13	The journalist calls the pilot of the travelers who is drinking too much	is	drinking	too	14	The journalist calls the pilot of the travelers who are drinking too much	are	drinking	too
15	The judge sees the helpers of the criminal who are lying	are	lying	–	16	The judge sees the helpers of the criminal who is lying	is	lying	–

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
17	The doctor observes the mother of the boys who is reading the newspaper	is	reading	the	18	The doctor observes the mother of the boys who are reading the newspaper	are	reading	the
19	The journalist interviews the daughters of the inspector who are looking very serious	are	looking	very	20	The journalist interviews the daughters of the inspector who is looking very serious	is	looking	very
21	The student photographs the fans of the actress who are looking happy	are	looking	happy	22	The student photographs the fans of the actress who is looking happy	is	looking	happy
23	The woman blames the sisters of the hair-dresser who are smiling all the time	are	smiling	all	24	The woman blames the sisters of the hair-dresser who is smiling all the time	is	smiling	all
<i>Argument overlap</i>									
49	The trapeze artist was very good, but tonight he fell. The plunge resulted in a broken leg	plunge	resulted	in	50	The trapeze artist was very good, but tonight he fell. The fall resulted in a broken leg	fall	resulted	in

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
51	The bomb hit the ground and loudly exploded. The detonation was huge	detonation	was	huge	52	The bomb hit the ground and loudly exploded. The explosion was huge	explosion	was	huge
53	The sun disappeared, the clouds became dark and it rained. The deluge ruined her sweater	deluge	ruined	her	54	The sun disappeared, the clouds became dark and it rained. The rain ruined her sweater	rain	ruined	her
55	Suzie saw her mistake and quickly corrected it. Rectifying it helped her get an A	Rectifying	helped	her	56	Suzie saw her mistake and quickly corrected it. Correcting it helped her get an A	Correcting	helped	her
57	The man opened the door and stole the car. The abducted car was found by the police	abducted	car	was	58	The man opened the door and stole the car. The stolen car was found by the police	stolen	car	was
59	After spotting the spider, John killed it. Assassinating spiders never bothered John	Assassinating	spiders	never	60	After spotting the spider, John killed it. Killing spiders never bothered John	Killing	spiders	never

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
61	When the doctor finished, he washed his hands. Decontaminating his hands felt nice	Decontaminating	his	hands	62	When the doctor finished, he washed his hands. Washing his hands felt nice	Washing	his	hands
63	After a day in the hot sun, the dog drank water. The beverage made him less thirsty	beverage	made	him	64	After a day in the hot sun, the dog drank water. The drink made him less thirsty	drink	made	him
65	When Margret dropped her cigarette, the result was a tragic fire. The blaze ruined many trees	blaze	ruined	many	66	When Margret dropped her cigarette, the result was a tragic fire. The fire ruined many trees	fire	ruined	many
67	Charles forgot the fish he caught in the box for one week, leaving a smell. The scent was awful	scent	was	awful	68	Charles forgot the fish he caught in the box for one week, leaving a smell. The smell was awful	smell	was	awful

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
69	Dick and Nancy had been married for one year and had a fancy dinner. The supper was great	supper	was	great	70	Dick and Nancy had been married for one year and had a fancy dinner. The dinner was great	dinner	was	great
71	Jerry found out he had a low grade, and failed. Having flunked meant he had to redo the exam	flunked	meant	he	72	Jerry found out he had a low grade, and failed. Having failed meant he had to redo the exam	failed	meant	he
<i>Anomalies</i>									
25	Lisa often plays sunshine with her brother	sunshine	with	her	26	Lisa often plays games with her brother	games	with	her
27	On a sunny day there were no dogs in the sky	dogs	in	the	28	On a sunny day there were no clouds in the sky	clouds	in	the
29	The bird built his nest in the fire behind the house	fire	behind	the	30	The bird built his nest in the tree behind the house	tree	behind	the
31	The pilot happily boarded his gun and left	gun	and	left	32	The pilot happily boarded his plane and left	plane	and	left
33	The man with the umbrella walked through the lie alone	lie	alone		34	The man with the umbrella walked through the rain alone	rain	alone	

Item	Complex	Critical word	Critical word plus 1	Critical word plus 2	Item	Simple	Critical word	Critical word plus 1	Critical word plus 2
35	Lieke thinks football is a better long than hockey	long	than	hockey	36	Lieke thinks football is a better sport than hockey	sport	than	hockey
37	On our way to the island we took the joke to the other side	joke	to	the	38	On our way to the island we took the boat to the other side	boat	to	the
39	The nurse was at the bedside of the paper and looked at him	paper	and	looked	40	The nurse was at the bedside of the patient and looked at him	patient	and	looked
41	The artist made a nice reason for the woman	reason	for	the	42	The artist made a nice statue for the woman	statue	for	the
43	Frits reads a book with exciting flowers about pirates	flowers	about	pirates	44	Frits reads a book with exciting stories about pirates	stories	about	pirates
45	The children went sledging in the rest on the hill	rest	on	the	46	The children went sledging in the snow on the hill	snow	on	the
47	Tess repairs the dinosaurs of her bicycle all by herself	dinosaurs	of	her	48	Tess repairs the brakes of her bicycle all by herself	brakes	of	her

## Appendix 2

*Reading comprehension texts and questions. All texts with corresponding Dutch questions are displayed below. All Dutch was translated to English and put between brackets.*

### Engels Begrijpend Lezen

#### [English Reading Comprehension]

Lees de onderstaande teksten en beantwoord de vragen:  
[Read the texts below and answer the questions:]

#### Tekst 1

#### [Text 1]

Vraag 1:

[Question 1]

“Husband shoots wife in leg” (title)

Why?

What happened was

A a sick joke

B a silly accident

C a wilful crime

Husband shoots wife in leg

A man shot his wife in the leg during an evening stroll. Mandy and Jon Hudson were using a gun to scare birds from fishing lakes when Mr Hudson tripped on a rabbit hole and the firearm went off. Mrs Hudson was hit in the ankle.

This happened at Bromeswell lakes, near Woodbridge, Suffolk, where Mr Hudson is leaseholder. Cliff Roberts, 65, a water bailiff, said: “We have a lot of Canada geese down here and a heron and we try to scare them off by firing a gun.”

Mrs Hudson, believed to be in her forties, was in a stable condition in hospital. Her husband was said to be in ‘very deep shock’. Police officers visited the scene but took no action.

**Daily Telegraph**

Vraag 2:

[Question 2:]

Lees onderstaande stellingen:

[Read the following statements:]

1: Mrs Hudson is gevallen

[1 Mrs Hudson fell]

2: Mrs Hudson is gewond

[2 Mrs Hudson is wounded]

Kies het juiste antwoord

[Choose the correct answer]

A Stelling 1 is juist, stelling 2 is onjuist

[A Statement 1 is correct, statement 2 is incorrect]

- B Stelling 1 is onjuist, stelling 2 is juist  
 [B Statement 1 is incorrect, statement 2 is correct]  
 C Beide stellingen zijn juist  
 [C Both statements are correct ]  
 D Beide stellingen zijn onjuist  
 [D Both statements are incorrect]

GA VERDER OP DE VOLGENDE BLADZIJDE  
 [CONTINUE ON THE NEXT PAGE]

Tekst 2  
 [Text 2:]

**Public Library**

1. If you put all the books you own on the street outside your house, you would expect them to disappear in a flash. But one man in Manilla tried it -and found that his collection grew. Hernando Guanlao is a cheerful man in his early 60s, with one great passion - books. They're his pride and joy, which is just as well because, whether he likes it or not, they seem to be taking over his house.  
*Disappear = verdwijnen*
2. Guanlao has set up an informal library outside his home in central Manilla to encourage his local community to share his joy of reading. The idea is simple: readers can take as many books as they want, for as long as they want- even permanently. It's a project you might think would end very quickly - with Guanlao having no books at all. But in the 12 years he's been running his library he's found that his collection has grown rather than diminished, as more and more people donate to the cause.  
*Encourage = stimuleren, diminish = kleiner worden*
3. Guanlao started his library in 2000, shortly after the death of his parents. He was looking for something to honour their remembrance, and that was when he hit upon the idea of promoting the reading habit he'd inherited. "I saw my old textbooks upstairs and came up with the concept of having the public use them," he says. So he put them outside the door of his house to see if anyone wanted to borrow them. They did, and they brought the books back with others to add the collection and the library was born.
4. Somehow, every day, a steady stream of people find their way there, to browse and to contribute. But it's people like Celine who sustain the library. She lives down the road from Guanlao which gives her the opportunity to help by regularly visiting with bulging bags of books. She says she loves the concept because Filipinos – certainly those who are not particularly wealthy – have limited access to books. "Considering the income here, I think parents have other priorities."  
*Contribute = bijdragen*
5. As we sit outside Guanlao's house watching people browse through his collection, he tells me why he thought it was worth spending all his time – even to the point of giving up his job and

surviving purely on his savings – to maintain the library. He has also given several boxes of books to a man trying to set up a similar venture outside Manilla. “You don’t do justice to these books if you put them in a cabinet or a box,” he says. “A book should be used and reused. It has life, it has a message. As a book caretaker, you become a full man.”  
*bbc.co.uk, 2012*

GA VERDER OP DE VOLGENDE BLADZIJDE  
 [CONTINUE ON THE NEXT PAGE]

Vraag 3:  
 [Question 3:]

Hoe introduceert de schrijver het onderwerp in alinea 1?  
 [How does the writer introduce the topic in paragraph 1?]

- A door twee verschillende gebeurtenissen met elkaar te vergelijken  
 [A by comparing two different events]
- B door een tegenstelling met de titel te omschrijven  
 [B by describing a contradiction with the title]
- C door een bijzondere situatie te omschrijven  
 [C by describing a special situation]
- D door zijn steun voor een bijzonder initiatief uit te spreken  
 [D by expressing his support for a special initiative]

Vraag 4:  
 [Question 4:]

Waarom is de boekencollectie alleen maar gegroeid?  
 [Why has the book collection only grown?]

- A omdat Guanlao steeds meer boeken koopt  
 [A because Guanlao is buying more and more books]
- B omdat niemand de boeken meeneemt  
 [B because no one is taking books (from him?) ]
- C omdat er veel schenkingen zijn  
 [C because there have been many donations]
- D omdat mensen de boeken weer terugbrengen  
 [D because people return the books ]

Vraag 5:  
 [Question 5:]

“It’s a project you might assume would end very quickly” (alinea 2)  
 [(paragraph 2)]

- ➔ In welke zin heeft de schrijfster dit al eens eerder gezegd?
- ➔ [In what sentence has the writer mentioned this before?]

Schrijf de eerste twee woorden van die zin hieronder over:  
 [Below, write down the first two words of that sentence:]

---

Vraag 6:

[Question 6:]

Why did Guanlao start his project, according to paragraph 3?

He felt the need to

- A contribute to a better environment for the future
- B do away with all the objects he did not really need.
- C do something suitable in memory of his parents
- D share his great wealth with less fortunate people.

Vraag 7:

[Question 7:]

“Considering the income here, I think parents have other priorities.”

➔ Wat wordt met deze zin bedoeld?

➔ [What is meant by this sentence?]

A dat bewoners van de Filipijnen geen zin hebben om geld uit te geven aan boeken

[A that inhabitants of the Philippines do not feel like spending money on books]

B dat bewoners van de Filipijnen geen geld hebben om aan boeken te besteden

[B that inhabitants of the Philippines do not have money to spend on books]

C dat bewoners van de Filipijnen boeken lezen niet belangrijk vinden

[C that inhabitants of the Philippines do not find reading books important]

D dat bewoners van de Filipijnen veel geld uitgeven aan boeken lezen

[D that inhabitants of the Philippines spend a lot of money on books]

Vraag 8:

[Question 8:]

What does Celine particularly like about Guanlao's library, according to paragraph 4?

A It gives poor people the chance to read books.

B It helps her get rid of books she no longer wants.

C It is in short walking distance from her home.

D It provides reading materials for local schools.

GA VERDER OP DE VOLGENDE BLADZIJDE

[CONTINUE ON THE NEXT PAGE]

Vraag 9:

[Question 9:]

What can be concluded about Guanlao from paragraph 5?

A He dislikes having to depend on gifts from well-wishers.

B He has found a way to earn money with his library.

C He is clearly pleased with the choices he has made.

D He lost his job because he spent too much time on his library

E He wants to make his business profitable by expanding

Tekst 3:

[Text 3:]

Vraag 10:

[Question 10:]

Je hebt onderstaande advertentie gelezen in *The London Paper* en je wilde wel in aanmerking komen voor deze VIP-avond. Je hebt een kansje gewaagd, maar zag te laat dat de termijn al verstreken was.

[You have read the advert below in *The London Paper* and wanted to be considered for this VIP-night. You took a chance, but found out too late, that the term had already passed.]

→ Heeft jouw aanmelding je dan toch geld gekost?

→ [Did your application cost you money?]

Omcirkel 'ja' of 'nee' en citeer (= schrijf over uit de tekst) de eerste twee en de laatste twee woorden van de volledige zin uit de tekst waarop jij je antwoord baseert.

[Circle 'yes' or 'no' and cite (=copy from the tekst) the first two and last two words from the full sentence, on which you base your answer.]

Ja / Nee

[Yes/No]

#### The London Competition

Win a VIP night out

To celebrate the start of The Diners Club, the new restaurant reservation service from londonparties.co.uk, *The London Paper* is giving away a great VIP night out for two on 5 July. The prize includes dinner at Navajo Joe and VIP guest-list entry to Café De Paris including a table and a glass of champagne.



For more information, go to [www.londonparties.co.uk/dinersclub](http://www.londonparties.co.uk/dinersclub)

How to enter:

For your chance to win, simply text VIP and your name to 88855

Terms and conditions:

Winners will be selected at random from all entries by 10am on 2 July 2011. If you text in after the closing date/time of this competition you will not be entered but you will still be charged. Costs: 50p plus your usual network rate. For full terms and conditions visit [thelondonpaper.com/competitions](http://thelondonpaper.com/competitions).

GA VERDER OP DE VOLGENDE BLADZIJDE

[CONTINUE ON THE NEXT PAGE]

Vraag 11:

[Question 11:]

What can you win when you enter?

A a restaurant reservation

B dinner at Café de Paris

C dinner at Navajo Joe

D a VIP night out for four people

## References

- Ahn, S., & Jiang, N. (2018). Automatic semantic integration during L2 sentential reading. *Bilingualism: Language and Cognition*, 21(2), 375–383.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Barnes, M. A., Ahmed, Y., Barth, A., & Francis, D. J. (2015). The relation of knowledge-text integration processes and reading comprehension in 7th-to 12th-grade students. *Scientific Studies of Reading*, 19(4), 253–272.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Van Berkum, J. J. V., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.

- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3–11.
- Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information Processing Letters*, 24(6), 377–380.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.
- Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, 18(3), 453–469.
- Calloway, R. C., & Perfetti, C. A. (2017). Integrative and predictive processes in text reading: the N400 across a sentence boundary. *Language, Cognition and Neuroscience*, 32(8), 1001–1016.
- Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, 11(5–6), 489–503.
- Oakhill, J., Cain, K., & McCarthy, D. (2015). Inference processing in children: The contributions of depth and breadth of vocabulary knowledge. *Inferences During Reading, 2015*, 140–159.
- Chen, L., Fang, X., & Perfetti, C. A. (2017). Word-to-text integration: ERP evidence for semantic and orthographic effects in Chinese. *Journal of Neurolinguistics*, 42, 83–92.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye Movements, 2007*, 341–371.
- College voor Toetsen en Examens—Board for Assessment and Exams. (2016). *Examen Engels KB VMBO, 1e tijdvak 2016 en VMBO-GL/TL, 2e tijdvak 2012 (English Exam 1st Time Block 2016, and 2nd Time Block 2012)*. Retrieved from: <https://www.examenblad.nl>.
- Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, 21(4), 617–641.
- Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words, 1990–present. Retrieved from: <https://corpus.byu.edu/coca/>.
- Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739–757.
- de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6(1), 51–77.
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, 66(5), 843–863.
- Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25(4), 529–557.
- Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, 30, 506–528.
- Ellis, N. C. (2013). Second language acquisition. In *Oxford Handbook of Construction Grammar* (pp. 365–378), G. Trousdale & T. Hoffmann (Eds.). Oxford: Oxford University Press.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 430–445.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883–899.
- Hagoort, P. (2017). The neural basis for primary and acquired language skills. In E. Segers & P. van den Broek (Eds.), *Developmental perspectives in written language and literacy: In Honor of LudoVerhoeven* (pp. 17–27). Amsterdam: John Benjamins Publishing Company.
- Helder, A., Perfetti, C. A., van den Broek, P., Stafura, J. Z., & Calloway, R. C. (2019). ERP Indicators of local and global text influences on word-to-text integration. *Language, Cognition and Neuroscience*, 34, 13–28.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2(2), 127–160.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Kline, P. (2013). *Handbook of psychological testing*. London: Routledge.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57, 1–44.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
- Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing: An Interdisciplinary Journal*, *19*, 99–131.
- Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A., & Federmeier, K. D. (2017). Use of contextual information and prediction by struggling adult readers: Evidence from reading times and event-related potentials. *Scientific Studies of Reading*, *21*(5), 359–375.
- Nahatame, S. (2018). Comprehension and processing of paired sentences in second language reading: A comparison of causal and semantic relatedness. *The Modern Language Journal*, *102*(2), 392–415.
- Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, *87*, 261–276.
- Nguyen, T. Q., Pickren, S. E., Saha, N. M., & Cutting, L. E. (2020). Executive functions and components of oral reading fluency through the lens of text complexity. *Reading and Writing*, *33*, 1037–1073.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37.
- Perfetti, C., Yang, C., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, *22*, 303–318.
- Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology*, *39A*, 657–673.
- Rosenberg, S. (1987). Semantic integration in sentence memory. *The American Journal of Psychology*, *100*(2), 253.
- Raudszus, H., Segers, E., & Verhoeven, L. (2018). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing*, *31*(2), 405–424.
- Raudszus, H., Segers, E., & Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*, *50*, 106–119.
- Segers, E., & Verhoeven, L. (2016). How logical reasoning mediates the relation between lexical quality and reading comprehension. *Reading and Writing*, *29*(4), 577–590.
- Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition*, *27*(3), 219–246.
- Stafura, J. Z., Rickles, B., & Perfetti, C. A. (2015). ERP evidence for memory and predictive mechanisms in word-to-text integration. *Language, Cognition and Neuroscience*, *30*(10), 1273–1290.
- Todaro, S., Millis, K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes*, *47*(5), 421–446.
- Torgesen, J. K. (1986). Computers and cognition in reading: A focus on decoding fluency. *Exceptional Children*, *53*(2), 157–162.
- Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency: (TOWRE-2)*. New York: Pearson Clinical Assessment.
- Tortorelli, L. S. (2020). Beyond first grade: Examining word, sentence, and discourse text factors associated with oral reading rate in informational text in second grade. *Reading and Writing*, *33*, 143–170.
- Van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing: An Interdisciplinary Journal*, *25*, 1665–1690.
- 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.
- Verhoeven, L., & Perfetti, C. A. (2008). Advances in text comprehension: model, process and development. *Applied Cognitive Psychology*, *22*(3), 293–301.
- Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818.
- Viebahn, M. C., McQueen, J. M., Ernestus, M., Frauenfelder, U. H., & Bürki, A. (2018). How much does orthography influence the processing of reduced word forms? Evidence from novel-word learning about French schwa deletion. *Quarterly Journal of Experimental Psychology*, *71*(11), 2378–2394.
- Whitford, V., & Titone, D. (2017). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition*, *22*, 1–20.
- Wurm, L. H., & Fisičaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language and Literacy*, 8(2), 157–181.
- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology*, 33, 55–89.

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