Available online at www.ejal.eu

Eurasian Journal of Applied Linguistics 1(2) (2015) 35–56

EJAL
Eurasian Journal of Applied Linguistics

# Eye movements in reading span tasks to working memory functions and second language reading

Cem Alptekin [a], Gülcan Erçetin [a] *

[a] *Boğaziçi University, Department of Foreign Language Education, Istanbul 34342, Turkey*

**Abstract**

In an eye-tracking experiment with late second language (L2) learners, the present study probed into the relationship between eye-movement (EM) measures and the processing and storage outcomes of reading span tasks (RST) to determine whether EM patterns actually reflect working memory (WM) functions in the L2. Additionally, it examined the relationship between WM capacity as indexed by EMs and L2 reading comprehension to explore whether it was possible to map offline and online data as predictors of L2 reading. The findings reveal that storage performance was negatively affected by fixation durations within the "critical" region of each sentence, indicating a trade-off between processing and storage. Additionally, regressions launched from the sentence-final region were negatively related to not only storage and processing performance but also L2 reading comprehension. These results have implications regarding whether EMs can be instrumental in validating offline span task outcomes and their association with L2 reading.



***Keywords:*** Working memory; eye movements; reading span tasks; L2 reading comprehension

## 1. Introduction

Working memory (WM), as a limited capacity system with dual functions of processing and storage, includes a memory span component with a concurrent processing operation, thereby subserving the active maintenance of information in the face of simultaneous distraction stemming from the intricate nature of the processing task. Thus, WM's modus operandi may be identified through complex span tasks of a verbal or visual-spatial essence that comprise a storage component (primary task) and a processing component (secondary task). While tasks such as word or shape recall aim to quantify the storage component, others such as sentence-level ambiguity resolution, arithmetic problem solving, or visual-spatial tracking identify the processing component.

* Corresponding author. Tel.: +90-212-359-4588
*E-mail address*: gulcaner@boun.edu.tr
http://dx.doi.org/..........................................

## 1. Background information on working memory and reading

An important role has been ascribed to WM in first-language (L1) reading comprehension (Daneman & Carpenter, 1980; Daneman & Merikle, 1996), with WM capacity often measured through complex span tasks, which are said to reflect trade-off patterns between processing and temporary storage of information in the WM system (Daneman & Hannon, 2007).[1] WM plays an important role in reading comprehension because readers retain available previous information while incorporating new information for textual comprehension. As described in Daneman and Merikle's (1996) meta-analysis, which engaged 6,179 participants in 77 separate studies conducted in the L1, span tasks, with their processing and storage functions, are good predictors of comprehension ability. They correlate .41 and .52 with global (textually explicit and implicit aspects of reading) and specific (schematically inferential and evaluative aspects of reading) measures of comprehension, respectively. Span tasks are viewed as measures of a domain-general cognitive capacity, which is said to be language-independent (Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993). This capacity is reported to depend on the ability to control attention between the two span components, keeping relevant information activated in spite of ongoing interference (e.g., Conway, Jarrold, Kane, Miyake, & Towse, 2007; Engle, Cantor, & Carullo, 1992; Engle, Kane, & Tuholski, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane et al., 2004; Kane, Conway, Hambrick, & Engle, 2007; Turner & Engle, 1989; although see Andrews, Birney, & Halford, 2006; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet, Portrat, & Camos, 2011; Towse & Hitch, 1995; Towse, Hitch, & Hutton, 2000, for different perspectives on this view).

One widely used instrument of verbal WM capacity assessment is the reading span task (RST). First designed by Daneman and Carpenter (1980), the original test requires participants to read aloud unrelated active-voice sentences, each 13-15 words in length. Every sentence ends with a high-frequency concrete word. The sentences are presented in sets of two to six designed to assess the processing component. Sentence-final word recall is prompted at the completion of each set as part of measuring the storage component. RSTs have undergone a number of modifications over time. One common variation has been the incorporation of a comprehension check to the processing component of the task, with participants being asked to make a true/false judgment or a sensibility judgment about the sentence read. Currently, it is customary for RSTs to make use of lexicosemantic or morphosyntactic incongruity as "distractors" in processing tasks. As participants process each RST sentence, they are asked to judge, for example, its lexicosemantic or morphosyntactic congruity. At the end of each set they are instructed to recall as many sentence-final words as possible.

Although the RST is considered to be a valid instrument for measuring verbal WM capacity, some concerns have been expressed over its domain generality, with a number of critics arguing that the meaningful relationships found between reading span and reading comprehension may be the result of domain-specific skills in verbal ability (e.g., Ericsson & Kintsch, 1995; Kintsch, 1998; Koda, 2005; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000) while others view it as a domain-general factor that processes and temporarily stores information subserving complex cognitive tasks one of which is reading comprehension (e.g., Daneman & Hannon, 2007).

Similar concerns regarding the RST have also emerged in relation to its significant relationship with measures of second language (L2) reading, particularly in view of the uncertainty involving WM's language independence (e.g., Sanchez et al., 2010; Service, Simola, Metsänheimoi, & Maury, 2002; Van den Noort, Bosch, & Hughdal, 2006). As a case in point, while the scores on RSTs administered in the L2 have been found to correlate with reading comprehension in the L2 (Alptekin & Erçetin, 2011; Service et al., 2002; Walter, 2004), those in the L1 have not shown a direct association with L2 reading except to mediate the relationship between L2 WM capacity and L2 reading, at least in the case of reasonably proficient L2 learners (Alptekin & Erçetin, 2010). It follows that in studies dealing with L2 reading, the use of RSTs in the L2 may generate confounding effects on research outcomes due to the potential influence of verbal proficiency in the target language (e.g., Juffs & Harrington, 2011), while their administration in the L1 may fail to yield conclusive results due to the lack of strong evidence on the relationship between WM capacity measured in the L1 and learning outcomes in the L2 (Alptekin & Erçetin, 2011; Geva & Ryan, 1993; Miyake & Friedman, 1998).

In sum, considering that issues involving RSTs' domain- and language-specificity may affect offline inferences on the trade-off between WM's processing and storage functions on one hand and the widely acclaimed relationship between WM capacity and L2 reading on the other, it is important to examine RST processes online, as they unfold in the L2, through a more direct methodology such as eye tracking in order to understand the moment-by-moment operations of linguistic factors in the temporary maintenance of information amid simultaneous distraction caused by the processing load. Hence, the question arises whether further insights into the role WM capacity plays in L2 reading comprehension could be gained through the use of eye-tracking methodology, particularly as regards how eye movements (EMs) in processing L2 RSTs relate to L2 reading comprehension.

## 2. Eye movements and reading

The application of EM monitoring to the L2 reading domain is fairly new. Less than a decade ago, Frenck-Mestre (2005) lamented that only a handful of EM studies on L2 reading had been published despite eye tracking and recording being an invaluable tool in the study of text processing. As a tracking and recording technique, EM monitoring has recently begun to be used in L2 research (see the special issue of *Studies in Second Language Acquisition*, Winke, Godfroid, & Gass, 2013) primarily to examine lexicosemantic and/or morphosyntactic processing of bilinguals at various stages of L2 proficiency (e.g., Dussias, 2010; Dussias & Sagarra, 2007; Felser & Cunnings, 2012; Foucart & Frenck-Mestre, 2012; Kaakinen & Hyönä, 2007;

Kaushanskaya & Marian, 2007; Keating, 2009; Roberts, Gullberg, & Indefrey, 2008; Roberts & Siyanova-Chanturia, 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Whitford & Titone, 2012; Winke, Gass, & Sydorenko, 2013; Witzel, Witzel, & Nicol, 2012). The focus on lexicon as an index of meaning has been on bilinguals' activation of one or both of their languages when they are exposed to input in one of their languages alone. In matters of syntax, the focus has been on investigating the type of processing operations (i.e., based on morphosyntactic or lexicosemantic computations) L2 learners make use of in linguistic incongruity detection in L2 reading. The primary reason for the use of EM monitoring in L2 research has to do with its allowing researchers to obtain evidence about what is happening moment-by-moment during the reading comprehension process "without significantly altering the normal characteristics of either the task or the presentation of the stimuli" (Dussias, 2010, pp. 153-154). As such, eye tracking can be said to simulate natural reading conditions in experimental research better than procedures (e.g., event-related potentials or self-paced reading) that require participants to read sentences word by word and prevent them from regressing to parts of the text that has already been processed (Sagarra & Ellis, 2013, p. 267). Given that eye tracking does not disrupt the reading process, we thought that it may offer an online record of the cognitive operations underlying the relationship between (verbal) WM capacity and reading comprehension in the L2, as evidenced in behavioral studies.

In fact, eye-tracking use in research on native language (L1) comprehension has been found to provide a relatively direct (online) means of obtaining insights into many cognitive processes underlying textual comprehension, to the extent that EMs have even been referred to "as a window into language and cognition" (Spivey, Richardson, & Dale, 2009, p. 225). In general, the method is based on examining readers' execution of a series of saccadic movements, which involve the saccade itself (a rapid and jerky EM between fixations), the fixation pause (the state of the eyes being relatively still), and the regression (the backward right-to-left movement of the eyes for those using a left-to-right script). What is investigated with reference to reading stems from the fixation-saccade-regression sequence, with such factors as the location and duration of fixations as well as the frequency of regressions being taken into consideration to determine variables such as text processing difficulties and cognitive resource demands.[2] The fewer the number of fixations and regressions and the shorter the fixation duration, the less the reader can be said to experience difficulty in reading comprehension. For example, when lexical access is easy, that is, a word in the text is easily identifiable and comprehensible, readers are likely not to feel the need to use fixations or, in the event they do, fixation duration will be relatively short. Fixation duration characterizes the time needed to process the information in the region of interest (ROI) and the time required to plan the next saccade.

Approximately 10 to 15 percent of all saccades are expected to move the eyes back to previous parts of the text. These backward movements, referred to as "regressive saccades" or simply "regressions," are usually indicative of processing problems that are associated with readers' difficulties in understanding and interpreting the text, thereby paving the way for fixations. It is generally held that the more difficult the text, the more the regressions, the longer the fixations, and the shorter the saccades will be (Rayner & Sereno, 1994). In addition to problems involving lexical access owing to such factors as word frequency, word familiarity, lexical ambiguity, contextual constraint, other discourse variables such as morphosyntactic, lexicosemantic, and pragmatic features may also play a role in readers' encountering comprehension difficulties (e.g., Binder & Morris, 2011; Van Gompel, Fischer, Murray, & Hill, 2007). A case in point would involve a reader misinterpreting a word or a phrase in the text due to lexical ambiguity, which would generate a regression, leading the reader to the part of the text, or the ROI, where the word or phrase is located. What this indicates is the tendency of readers to go back to what they deem to be a "critical" section of the context which they reread due to their understanding of that section being inadequate.

Technically speaking, extended fixations and regressions in processing texts are known to reflect the reader's inability to encode the material online, thereby being symptomatic of increased cognitive demands on WM capacity. This may result not only from individual variations in span capacity but also from the conceptual and linguistic complexity of the sentence that is processed. In the event textual requirements exceed the upper bound of capacity constraints, there may be serious deterioration of comprehension processes, which is likely to lead to higher incidences of relatively long fixations and frequent regressions. Thus, tracking EMs can provide a detailed view of the time-course of processing as well as the launch sites of fixations and regressions, with a view to generating a temporally and spatially specific representation of online processing that underlies offline span task performance.

Normally, EM records consist of both early- and late-stage EM measures (Boland & Blodgett, 2002; Clifton, Staub, & Rayner, 2007; Mitchell, 2004; Pickering, Frisson, McElree, & Traxler, 2004; Rayner, 2009; Whitford & Titone, 2012), as time spent in a ROI and the immediacy or lateness of specific fixations appear to be sensitive to cognitive processing. Early-stage measures, which are said to index "processes that occur in the initial stages of sentence processing" (Clifton et al., 2007, p. 349), involve first pass reading times (FP), which refer to the summed duration of fixations (and possible refixations) made within a ROI from first entering it until first leaving it with a saccade—either to the left or to the right. Also referred to as "gaze duration" when the ROI consists of a single word, this measure is informative in revealing detections of morphosyntactic inaccuracies (Dussias, 2010) and initial lexical problems (Roberts & Siyanova-Chanturia, 2013; Whitford & Titone, 2012; Winke, Godfroid, & Gass, 2013).

Late-stage measures, on the other hand, are taken to reflect higher-order processes such as formulating sentence-level semantic integration, information reanalysis, and recovery from processing problems (Clifton et al., 2007; Roberts & Siyanova-Chanturia, 2013). In addition to several late-stage measures (e.g., second pass reading time or regression path duration), total reading time (TT) is perhaps the best late-stage measure on account of its being sensitive to linguistic processes that operate after a word has been identified, thereby being indicative of sentence comprehension as a whole (Holmqvist et al., 2011, p. 389; Roberts & Siyanova-Chanturia, 2013, p. 217). Total reading time refers to the summed duration of all fixations made within a ROI (Cook, Colbert-Getz, & Kircher, 2013; Dussias, 2010) and often signals "an interruption to the normal reading process" (Winke, Godfroid, & Gass, 2013, p. 206). It follows that the longer the duration of the sum of all fixations in a ROI the higher the degree of its processing load is likely to be, thereby hampering reading comprehension.

In addition to measures of processing time, EM research makes use of regressions, which refer to the reader's returning to a previous ROI from a currently fixated ROI. One particular type that involves backward EMs is the "regression-out" (Holmqvist et al., 2011, p. 426), which normally occurs when the reader experiences initial difficulty interpreting a particular ROI, in which early lexical or syntactic problems hinder access. The frequency of regressions-out (RO) indexes increased cognitive load (Roberts & Siyanova-Chanturia, 2013, p. 220) and could be taken as a late-stage measure, as it is often indicative of sentence-final cognitive processes that point to the processor's inability to incorporate material on-line.[3]

Early- and late-stage processing time as well as the frequency of ROs provide useful measures in revealing whether a particular difficulty in a given ROI affects the initial processing of that region or exerts an effect during the later stages in sentence comprehension. They thus offer insights into the cost of overcoming processing difficulties encountered in different regions in a sentence and may be instrumental in understanding the linguistic decisions individuals make during reading.

## 3. The study

Based on the above considerations, the research underpinning the present study aimed to probe into the relationship between EM measures and RST's processing and storage outcomes to determine whether EM patterns actually reflect WM functions in the L2, as seems to be the case in the L1 (Just & Carpenter, 1980). The study further examined the relationship between L2 reading comprehension and WM capacity as indexed by EMs in L2 span task processing, with a view to exploring the match between online and offline data as predictors of L2 reading.

## 4. Research questions and hypotheses

The following questions were posed:

1) Are online EM measures related to offline RST outcomes of processing and storage?
2) If EMs are indicators of RST functions, are they related to L2 reading comprehension?

As regards the first question, eye-tracking literature concerning sentence congruity judgments suggests that the process is incremental and aims at achieving semantic and pragmatic fit by the end of the sentence (e.g., Roberts, 2010). It is not surprising, therefore, that long fixation durations in the sentence-final region along with high incidences of ROs from it have been observed in the literature (Boland & Blodgett, 2002; Braze, Shankweiler, Ni, & Palumbo, 2002; Dussias & Piñar, 2010; Keating, 2009; Ni, Fodor, Crain, & Shankweiler, 1998; Roberts, 2012). Particularly in the case of RST sentences, the reader's focus on the end of the sentence is enhanced in view of the need to recall the sentence-final word as part of the span task, thereby increasing the possibility for long fixation durations to take place at sentence end. Similarly, long fixation durations involving local lexicosemantic or morphosyntactic difficulties can also be observed within the "critical" region of each sentence (Dussias, 2010; Rayner, Warren, Juhasz, & Liversedge, 2004; Roberts & Siyanova-Chanturia, 2013; Whitford & Titone, 2012). As for ROs, they are expected to be launched from the sentence-final region rather than the "critical" region, given that semantic and pragmatic fit normally takes place by the end of the sentence. Thus, by applying eye-tracking into the examination of WM functions, we expect EMs to reflect trade-off relationships between WM's storage and processing components in the L2 context, as this context is rich for investigating the links between EMs and cognitive operations due to processing problems that are endemic in L2 learners (Winke, Godfroid, & Gass, 2013, p. 207).

Specifically, we predict longer fixation durations within the "critical" region of each sentence, where the verb argument structure may present a case of lexicosemantic incongruity. These fixations would, in turn, reflect how WM's storage function could be negatively affected by the focus on local processing. In addition, we expect longer fixation durations in the sentence-final region of the RST sentences, as mentioned above. However, we do not anticipate these durations to be indicative of a trade-off between processing and storage because the reader's focus on sentence end is likely to reflect both the processing of sentence-level semantic and pragmatic information and the sentence-final word to be remembered. Instead, we predict L2 learners' ROs from the sentence-final region to correlate negatively with storage performance, as ROs are taken to reflect processing difficulties characterizing the incremental assessment of sentential plausibility, with a view to achieving semantic and pragmatic fit by the end of the sentence (Hypothesis 1).

As regards the second question, L2 reading comprehension, which normally entails complex cognitive operations, is expected to correlate negatively with late-stage fixation durations as well as RO frequencies because lexicosemantic congruity judgment would require a relatively significant degree of higher-order (e.g., sentence-level semantic and pragmatic integration) processing skills, as it results from coping with a sentence in terms of its overall plausibility that rests on the conceptually congruent relationships of its constituents (Hypothesis 2).

## 5. Methodology

### *5.1. Participants*

Thirty Turkish first-year students majoring in English in an English-medium university in Turkey participated in the study, for which they received course credit.

They had been successful on the TOEFL iBT, whose cut-off point for admission to first-year courses is taken to be 80, corresponding to the "Threshold" (B1) level on the Common European Framework of Reference for Languages.[4] They had also obtained high scores on the Turkish equivalent of the verbal sections of the SAT Reasoning Test, administered countrywide as part of the national university admission examination system. All participants had normal or corrected-to-normal vision and their ages ranged from 20 to 22, with a mean of 21.5. The data for 25 (21 female, 4 male) of the 30 participants were analyzed because they were complete.

*5.2. Materials and procedures*

The RST used in the study was a modified version of the original Daneman and Carpenter (1980) test. It made use of a span task that involved the assessment of lexicosemantic incongruity, measured through an equal number of lexicosemantically plausible and implausible sentences. The participants had to judge whether the verb argument structures in the RSTs were congruent or incongruent as part of processing the sentences and commit to memory the sentence-final noun. These words were selected from the British National Corpus on account of their high-frequency and concreteness (e.g., *door*),[5] with every sentence ending with a different noun. Care was taken not to deploy any cognates.

The test was administered in the L2 (English) due to research findings pointing to the existence of a direct relationship between WM capacity and L2 reading, with L1 WM capacity being chiefly a mediator for reasonably proficient L2 learners, as indicated earlier.[6] Otherwise, in line with the conventional format of RSTs in the literature, the test consisted of 42 unrelated simple sentences in the active voice, each 12-14 words in length. It comprised four levels, starting at level two and extending up to five, with each level containing three trials. There were 21 lexicosemantically congruent (e.g., *The dean's office ought to give new work permits to all students)* and 21 incongruent sentences (e.g., *The little bird will often dance its lovely songs by the window*). The sentences were arranged randomly.

Following a practice session prior to the experiment, participants were individually tested on the same sets of sentences and each sentence was presented only once. After completing all 3 trials at a given level, they moved on to trial 1 of the next level, with their EMs recorded throughout the experiment. Using the Cronbach alpha, the internal consistency reliability coefficients for the processing and storage tasks on the RST were found to be .770 and .845, respectively.

The sentences in the RSTs were modelled on the design in a similar study done with native speakers of English by Braze and colleagues (2002), with one exception. In the Braze et al. study, despite the application of a length correction procedure, there was an unequal length of ROIs with an uneven number of words per region. Given that this procedure is considered not particularly reliable when the "critical" region is short (Rayner et al., 2004), we decided to divide each sentence in the present study into four ROIs for statistical analyses, with ROI/3 being the "critical" region. Each region comprised meaningful phrase structure units of 10 to 14 characters. Otherwise, based on the relevant eye-tracking data (Braze et al., 2002; Ni et al., 1998), cases of lexicosemantic (in)congruity were generated through the verb argument structure, in which the verbs used were all four-to-five characters each and of high frequency (e.g., *The little bird/ will often dance/ its lovely songs/ by the window*).

The RSTs were administered in a computer lab so that they could be delivered online. Sentences were displayed at 7-second intervals until all in a set had been viewed.[7] While processing the sentences, the participants pressed one of two computer keys to indicate whether a given sentence was appropriate in terms of its lexicosemantic content. After the participants viewed all the sentences in a set, a field box appeared on the screen for them to enter the sentence-final words that they were able to recall. While the participants' judgments concerning the lexicosemantic congruity of the sentences represented the processing measure of their reading span, the total number of correctly recalled sentence-final words was taken as the measure of storage.

Participants' reading knowledge in the L2 was measured through the use of the comprehension section of the Nelson-Denny Reading Test (NDRT). The test consisted of seven reading passages with a total of 38 multiple-choice questions. The passages were taken from science, social science, and humanities texts used at high school and college levels in the Unites States. The skills tested included understanding explicit details, understanding main ideas, drawing conclusions, and generating inferences.

### *5.3. Apparatus and procedures*

Participants' EMs were recorded from the right eye using a D6 Remote Optics Applied Science Laboratories (ASL) eye tracker, at a sampling rate of 60 Hz. The D6 optics module is placed on a stand directly under a 19-inch subject display monitor on which the stimulus is displayed. The participant sat in front of the subject display monitor at an optimal distance, which is approximately 24 inches from the eye. The eye tracker was calibrated for each participant at the beginning of the experimental session. Following the calibration procedure, which required participants to fixate their gaze on nine points presented on the computer screen, the RST was run using E-prime (Schneider, Eschman, & Zuccolotto, 2002). Viewing was binocular but EMs were recorded from the left eye only. In order to synchronize the eye data recorded by the eye tracker with the E-prime experiment, a different integer value was given for each sentence on the RST, which was sent to the eye tracker via a parallel port. This value was recorded as External Data (XDAT) along with each gaze data record.

### *5.4. Data analysis*

The analyses were conducted using the ASL Results Program, which provides both fixation sequence and duration. Fixation durations under 80 ms were automatically eliminated and those greater than 1000 ms were not counted during the analyses (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Traxler, Pickering, & Clifton, 1998). This procedure was adopted because very short fixation durations do not allow readers to extract much information during reading. In fact, they are said to reflect oculomotor programming times rather than cognitive processes (Binder & Morris, 2011, p. 314). As for individual fixation durations greater than 1000 ms, they are found to reflect instances where the eye-tracker loses track of the reader's gaze location (Rayner & Pollatsek, 1989).[8] Thus, extreme losses in the corneal and pupil reflections during the experiment resulted in some data being excluded from the analyses.

Two standard reading time measures of EMs in ms were analyzed in line with what is commonly used in eye-tracking studies involving early- and late-stage measures related to incongruity resolution (Clifton et al., 2007; Dussias, 2010; Mitchell, 2004; Pickering et al., 2004; Whitford & Titone, 2012). These were first-pass reading time FPs and TTs: FP refers to the summed duration of fixations made within a ROI from first entering it until first leaving it with a saccade, either to the left or to the right. This measure is said to be informative in revealing detections of syntactic inaccuracies (Dussias, 2010) and in reflecting initial processes of lexical access (Roberts & Siyanova-Chanturia, 2013; Whitford & Titone, 2012). On the other hand, TT is the summed duration of all fixations made within a ROI (Cook, Colbert-Getz, & Kircher, 2013; Dussias, 2010). Often being indicative of higher-order processing, this measure reflects sentence-level semantic integration, information reanalysis, and recovery from processing problems (Clifton et al., 2007; Roberts & Siyanova-Chanturia, 2013).

Additionally, the probability of making ROs was taken into consideration, as an index of increased processing load (Roberts & Siyanova-Chanturia, 2013, p. 220). ROs are defined as the number of eye movements in which a FP fixation made within a ROI is followed by a regression to an earlier region. As indicated earlier, they normally occur when the reader experiences initial difficulty understanding or interpreting a particular ROI, in which early lexical or syntactic problems hinder access.

The number of ROs for each sentence was computed for each participant. Since it was not possible to regress from the first region of the RST, only ROs from regions 2, 3, and 4 were calculated. The RO means are based on the total number of regressions per condition and per participant, divided by the number of trials in each condition. These measures were computed for each of the four regions manually. FP and TT durations as well as the frequency of ROs involving all four regions were analyzed.

In order to determine whether these EM measures were related to offline RST outcomes, their correlations with processing scores and storage scores were obtained for each ROI. As for the relationship between EMs and reading comprehension, the correlations of the same EM measures with scores on the NDRT were computed.

## 6. Results

### *6.1. Eye movements and WM functions*

Table 1 provides the descriptive statistics for the FP and TT values (in ms) as well as the percentage of ROs across the four regions.

Table 1. Descriptive statistics for EM measures

| | FP (ms) | | TT (ms) | | RO (%) | |
|---|---|---|---|---|---|---|
| Sentences | M | SD | M | SD | M | SD |
| Region 1 | 417.61 | 160.05 | 480.13 | 171.58 | - | - |
| Region 2 | 358.56 | 106.12 | 488.23 | 127.23 | 3.83 | 4.38 |
| Region 3 | 371.42 | 114.19 | 563.59 | 177.35 | 5.23 | 6.18 |
| Region 4 | 530.01 | 182.59 | 643.86 | 235.66 | 45.80 | 17.99 |

*Note*. FP = First pass reading time; TT = Total reading time; RO = Regression out.

FP and TT durations in Table 1 indicate that the longest fixation duration is observed in ROI/4. Similarly, the most frequent regressions are launched from ROI/4. A one-way repeated measures ANOVA conducted on fixation durations and ROs indicated a significant effect of region for FP duration $F(3, 72) = 16.30$, $p < .001$, partial eta$^2$ = .40, TT duration, $F(3, 72) = 7.94$, $p < .01$, partial eta$^2$ = .25, as well as ROs, $F(2, 48) = 114.04$, $p < .001$, partial eta$^2$ = .83. Post-hoc comparisons through the Bonferroni procedure revealed that FP duration in ROI/4 was significantly longer than those in the other three regions. Similarly, ROs from this region were significantly more frequent compared to the other regions. In addition, TT durations in ROI/3 and ROI/4 were significantly longer than those in the other regions. However, the difference between these two regions was not significant.

In order to determine whether these EM measures were related to offline RST outcomes, Pearson product moment correlations of fixation durations and regressions with the scores on the primary storage task ($M$ = 21.04, $SD$ = 7.09) and the secondary processing task ($M$ = 29.08, $SD$ = 4.68) were obtained (see Table 2).

Table 2. Correlations between EMs and RST components

| | FP | | TT | | RO | |
|---|---|---|---|---|---|---|
| | Processing | Storage | Processing | Storage | Processing | Storage |
| Region 1 | -04 | .07 | -.03 | -.01 | - | - |
| Region 2 | -.24 | -.13 | -.38 | -.32 | .19 | .28 |
| Region 3 | -18 | -.35 | -.31 | -.41* | -.01 | .15 |
| Region 4 | -.27 | -.01 | -.31 | -.02 | -.44* | -.40* |

*Note*. FP = First pass reading time; TT = Total reading time; RO = Regression out.

* $p < .05$.

The correlations in Table 2 suggest a trade-off relationship between processing and storage in terms of TT duration and RO frequency. Specifically, when eye fixation is on ROI/3, storage seems to suffer. As for ROs, those launched from ROI/4 appear to affect both processing and storage performance negatively. No significant correlations are found between WM functions and FP fixation durations made within any region.

To summarize, it is in ROI/4 that the longest initial FP and late-stage TT durations are observed, not to mention the most frequent ROs from this region, because of the pressing need for reanalysis and integration of information to achieve semantic-pragmatic fit at the end of the sentence. It follows that by the time participants tackle the RST tasks, their lower- and higher-order processing skills are taxed considerably. Our first hypothesis predicted that the trade-off between WM functions would be observed in terms of TT duration and RO frequency. As predicted, a significant negative relationship was observed between TT duration in ROI/3 and storage performance. Additionally, the ROs launched from ROI/4 had a negative effect on storage, as we anticipated, but also on processing—which was not expected. By and large, then, EM measures reflect the trade-off patterns between WM's processing and storage components through late-stage effects.

### *6.2. Eye movements and L2 reading*

An investigation of the performance data revealed that L2 reading comprehension ($M$ = 32.45, $SD$ = 2.69, *Min.* = 26, *Max.* = 37) correlated substantially with storage scores ($r = .51$, $p < .01$) but not with processing scores ($r = .19$, $p > .05$). With a view to determining whether EM measures obtained from RSTs could be used as indices of WM functions similar to the performance scores obtained from RSTs themselves, Pearson product moment correlations between EM measures and L2 reading comprehension scores were computed. Findings indicated a significant negative correlation between ROs and L2 reading comprehension ($r = -.43$, $p < .05$) while FP duration ($r = .17$, $p > .05$) and TT duration ($r = .-01$, $p > .05$) had negligible relationships with L2 reading. Thus, our second hypothesis, which predicted significant correlations of L2 reading with TT durations as well as ROs was partially confirmed.

## 7. Discussion

On the whole, our first hypothesis was confirmed. True to our predictions, we obtained negative correlations between storage capability and the frequency of ROs induced from ROI/4 (sentence-final region). As mentioned earlier, these EMs reflect higher-order processes of post-lexical access (see Whitford & Titone, 2012), representing readers' attempts for discourse integration and coherence checking, with a view to deriving a meaningful interpretation of the target sentence. Such attempts, particularly if the sentence presents a case of lexicosemantic incongruity, seem to require readers to spend considerable time to read and reread it, which clearly leads to a negative effect on storage.

It is likely that the frequent ROs from ROI/4 come from "sentence wrap-up effects" (Just & Carpenter, 1980; Mitchell & Green, 1978; Rayner, Juhasz, Ashby, & Clifton, 2003), whereby readers are said to become involved in high-level information reanalysis and discourse integration. Given the case of lexicosemantic incongruity as a potential "intruder," readers possibly feel the urge to induce ROs from ROI/4 in order to achieve semantic and pragmatic fit by trying to clarify ambiguities unresolved during FP fixations. As this entails a high degree of element interactivity (lexical access, structure assignment, semantically- and pragmatically-oriented heuristics), it would normally place a considerable burden on WM's limited resources

(Beckman, 2010). Thus, the reanalysis and integration of information, deferred until sentence end, has a negative influence on the reader's ability to recall sentence-final words in RST tasks. In this sense, the findings of the present study are in concert with those of previous research conducted with native speakers (Braze et al., 2002; Ni et al., 1998).

Storage capacity is further hampered by the reader's extended focus on ROI/3. It is possible that readers' extended TT durations in the "critical" region are indicative of their efforts to resolve the lexicosemantic incongruity which, in turn, has a negative effect on their storage performance, in line with trade-off relative to WM's processing and storage functions. Long TT fixations that index processing difficulties of a lexicosemantic nature are compounded by the necessity to recall the sentence-final words, leading to deficiencies in task performance requiring memorization.

In fact, lexicosemantic incongruity presents serious processing challenges because, unlike early-stage effects reminiscent of initial lexical access or structure assignment, it involves the violation (*dance its lovely songs*) of a semantically congruent (formulaic) chunk with a high co-occurrence frequency (*sing its lovely songs*), which requires a global sense creation process concomitant with the totality of sentence-based propositional content. Sense creation is a higher-order reasoning process necessitating lexical encoding of information that lacks semantic-pragmatic congruity and formulaic chunking (Caillies & Declercq, 2011). As a late-stage process, it stems from readers' compensatory efforts to generate meaning when they are unable to achieve lexical access in their initial processing due to a number of factors (e.g., unfamiliar content, lexicosemantic implausibility, task complexity). It follows that, with their moderate proficiency level in the L2, the late learners in the present study show a propensity for deploying lexicosemantically-oriented heuristics for incongruity resolution while trying to commit to memory the sentence-final words.

A finding that was not anticipated with respect to trade-off was the negative relationship between processing performance and ROs from ROI/4. Since a trade-off effect is based on the premise that an increase in demands in one of the WM components would necessitate a decrease in the sources allocated for the other, the fact that the frequency of ROs from ROI/4, representative of higher-order processes, was instrumental in a decline in *processing* capacity failed to initially make sense from a trade-off perspective. However, we do not view this as evidence against the trade-off effect. Research findings in EMs conducted with native speakers suggest that in sentences with lexicosemantic incongruity there is a tendency in readers to initiate ROs from the end of the sentence to generate a meaningful interpretation of the whole, as mentioned above. This often leads to a gradual rise in ROs that peak at the ends of sentences, causing delays in the interpretation of information (Boland & Blodgett, 2002; Braze et al., 2002, Ni et al., 1998). Given that in the present study, readers had to chunk together a number of phrases that were lexicosemantically unrelated and operate under time pressure, it is likely that the high frequency of ROs launched from ROI/4 was indicative of the need for additional processing resources which, in turn, affected not only their storage performance (as indicated) but also their overall processing performance negatively. That is, RST requirements may have exceeded the upper bound of WM's limited capacity constraints at times, resulting in a cognitive overload that seriously impeded the reader's ability to actively maintain and process the stimuli under conditions of interference. After all, trade-off effects, as Towse and Hitch (1995) indicate, normally take place "when working memory is not fully occupied by processing" (p. 123).

Our second hypothesis, which predicted L2 reading to correlate with late-stage measures of EMs was partially confirmed. Only in the case of the frequency of ROs a negative relationship was observed between the two constructs. The more ROs readers launched to previous regions in the sentence the less satisfactory their reading comprehension performance appeared to be. Otherwise, the Pearson product moment correlations between early- (FP) and late-stage (TT) EM measures on the RST and L2 reading scores based on the NDRT failed to indicate any statistically significant relationships. Thus, there was no transferability of the meaningful association observed between extended TT durations in ROI/3 and storage into reading comprehension, casting doubt on the extent EMs may be indicators of RST functions as these underlie reading processes.

On the other hand, the negative relationship between ROs and reading comprehension appears to be in line with the negative associations between ROs and both processing and storage. L2 readers launch ROs from ROI/4, often to the "critical" region, due to the need to resolve comprehension and interpretation problems. These are higher-level cognitive processes demanding additional resources of processing while also deploying resources from the storage component, as suggested by the results to the first research question. The foregone conclusion is that an increase in cognitive load hampers reading comprehension--a well-known notion from offline research (e.g., Sweller, 1994), which is also captured by eye-tracking data.

In sum, we agree with Rayner and Sereno's (1994) observation that EM data are not perfect reflections of the cognitive operations associated with comprehension even though they may be informative with respect to understanding reading. As a case in point, they reveal how long it takes readers to read and re-read "critical" target regions in a sentence and how much they regress to earlier regions to resolve incongruity, in the same way they would do in a natural reading task. But how far these EM patterns reflect processes of reading comprehension at the sentence level is, in our view, an open-ended issue, even though one senses that any supposition hinting at the absence of EM role in understanding textual meaning would be counter-intuitive.

In this respect, late L2 learners' eye-tracking patterns show that they process the input incrementally by reading the entire sentence, evidently trying to integrate new information they encounter into the ongoing semantically- and pragmatically-oriented analysis of the sentence which they update continuously, more like native speakers, as reported by Roberts (2010). As such, sentences containing lexicosemantic incongruity in the L2 seem to offer a relatively similar picture to how this is resolved in the L1, except perhaps for the readers' limited ability to apply their lower- and higher-order processing skills in full to reading.

## 8. Conclusion

The findings of the present study demonstrate that eye tracking may provide an online measure of the time-course of processing and storage components characterizing WM operations with regard to L2 sentence-level comprehension. In this context, late L2 learners, with a moderate degree of proficiency in the target language, are seen to make use of higher-order processing skills to resolve incongruities of a lexicosemantic nature. In fact, lexicosemantic incongruity is found to be processed as an integral part of a sentence-level phenomenon, requiring L2 learners to read the entire sentence to piece together its meaning before focusing on the "critical" region with a view to disambiguating it. Late-stage TT durations in the "critical" region along with high incidences of regressive EMs from the sentence-final region point to efforts to establish a semantically and pragmatically coherent fit through higher-order cognitive processes.

Nevertheless, despite the promise EMs offer in reflecting the processing skills that underlie WM functions, "it should be clear that there has been no definitive mapping between specific cognitive processes and eye behavior," as stated by Boland (2004, p. 63). This may be due to the fact that EMs are preceded by an attentional shift to the saccadic target, with visual attention being slightly ahead of the eye (between 100-250 ms) and causing the eyes to follow its shifts to new positions (Hoffman & Subramaniam, 1995). Thus, the lack of steady isomorphicity between EMs in real time and both WM functions and reading outcomes involving time durations (FP and TT durations) may not be surprising. In our view, the frequency of ROs seems to be a more valid index of processing difficulties in L2 reading.

This casts doubt on claims that EMs can perhaps be instrumental in validating offline span task outcomes as these relate to the cognitive processes underlying L2 reading comprehension. For this to happen, it is clear that more research needs to be done—the type of research that will provide answers to how EMs are guided by different cognitive operations in sentence comprehension. In conclusion, we concur with Pickering and colleagues (2004) when they say that, at this point in time, "we do not typically know how to uniquely map cognitive events on to the different eye-movement measures, even when we have good grounds for believing that the eye-mind assumption generally holds" (pp. 35-36). It follows that it would be rather premature to make use of EM measures as direct indices of WM functions in L2 reading comprehension, despite the important information they offer regarding the processes underlying WM operations.

**Acknowledgements**

This study was funded by a research grant (Grant 110K261) from the Scientific and Technological Research Council of Turkey (TUBITAK).

## References

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*, 206–219. **http://doi.org/fsfjxc**

Alptekin, C., & Erçetin, G. (2011). The effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly*, *45*, 235–266. **http://doi.org/dzdjjp**

Andrews, G., Birney, D., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition*, *34*, 1325–1340.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570–585. **http://doi.org/cm22dw**

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*, 175–192. **http://doi.org/bq7rsv**

Beckman, J. F. (2010). Taming a beast of burden – On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction*, *20*, 250–264. **http://doi.org/dmczgh**

Binder, K. S., & Morris, R. K. (2011). An eye-movement analysis of ambiguity resolution: Beyond meaning access. *Discourse Processes*, *48*, 305-330. **http://doi.org/cq35kp**

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 51–76). New York: Psychology Press.

Boland, J. E., & Blodgett, A. (2002). *Eye movement as a measure of syntactic and semantic incongruity in unambiguous sentences*. Unpublished manuscript, University of Michigan.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31,* 25–44.

Caillies, S., & Declercq, C. (2011). Kill the song—steal the show: What does distinguish predicative metaphors from decomposable idioms? *Journal of Psycholinguistic Research*, *40*, 205–223. **http://doi.org/ck8fqg**

Cappelletti, M., Fregni, F., Shapiro, K., Pascual-Leone, A., & Caramazza, A. (2008). Processing nouns and verbs in the left frontal cortex: A transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *20*, 707–720.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Oxford: Elsevier.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). Variation in working memory: An introduction. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 3-17). New York: Oxford University Press.

Conway. A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Cook, A. E., Colbert-Getz, J., & Kircher, J. C. (2013). Number-of-features effects during reading: Evidence from eye movements. *Discourse Processes*, 50, 210–225. **http://doi.org/5jn**

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. **http://doi.org/fk6kpz**

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 21–42). New York: Oxford University Press.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: Meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433.

Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology*, *54A*, 31–48. **http://doi.org/cf35mw**

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149–166. **http://doi.org/bwpk82**

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English second language speakers. *Second Language Research*, *26*, 443–472. **http://doi.org/dbnqhs**

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116. **http://doi.org/fcs9h5**

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 972–992.

Engle, R. W., Kane M. J., & Tuholski S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Felser, C., & Cunnings, I. (2012). Processing reflexives in English as a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics*, *33*, 571–603. **http://doi.org/fch5nv**

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226–248. **http://doi.org/5jp**

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *21*, 175–198. **http://doi.org/b4bv6f**

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Geva, E., & Ryan, E. B. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language Learning*, *43*, 5–42. **http://doi.org/bcm3bn**

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*, 787–795.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44,* 137–166. **http://doi.org/bb3nb5**

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149. **http://doi.org/c6595h**

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*, 634–646. **http://doi.org/cgwcz3**

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183. **http://doi.org/c6rmmg**

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217. **http://doi.org/cntzrs**

Kaushanskaya, M., & Marian, V. (2007). Non-target language recognition and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning*, *57*, 119-163. **http://doi.org/cfdmfq**

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*, 503-535. **http://doi.org/b8fqmk**

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* New York: Cambridge University Press.

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 15–32). New York: Psychology Press.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, *30*, 609–636.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, *27*, 515–539.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island contraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588. **http://doi.org/bhcqwx**

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*, 287–289.

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31,* 117–118.

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 33-50). New York: Psychology Press.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457-1506. **http://doi.org/d4pf7r**

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, NJ: Erlbaum.

Rayner, K., & Sereno, S. C. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–82). San Diego, CA: Academic Press.

Rayner, K., Juhasz, B., Ashby, J., & Clifton, C. (2003). Inhibition of saccade return in reading. *Vision Research*, *43*, 1027-1034. **http://doi.org/c67ghm**

Rayner, K., Warren, T., Juhasz, B. J., Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1290-1301.

Roberts, L. (2010). Parsing the L2 input, an overview: Investigating L2 learners' processing of syntactic ambiguities and dependencies in real-time comprehension. In G. D. Véronique (Ed.), *Language, interaction and acquisition* [Special issue], (pp. 189–205). Amsterdam: Benjamins.

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*, 113–127. **http://doi.org/fxz8ck**

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*, 213–235.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30,* 333–357.

Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, *35*, 261-290.

Sanchez, C. A., Wiley, J., Miura, T. K., Colflesh, G. J. H., Ricks, T. R., Jensen, M. S., & Conway, A. R. A. (2010). Assessing working memory capacity in a non-native language. *Learning and Individual Differences*, *20*, 488–493.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime reference guide.* Pittsburgh, PA: Psychology Software Tools Inc.

Seigneuric, A., Ehrlich, M., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*, 81-103.

Service, E., Simola, M., Metsänheimoi, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14*, 383–407.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*, 251-272. **http://doi.org/bspr64**

Spivey, M., Richardson, D., & Dale, R. (2009). The movement of eye and hand as a window into language and cognition. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds). *Oxford handbook of human action* (pp. 225–249). Oxford: Oxford University Press.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The timecourse of plausibility effects on eye movements in reading: Evidence from noun–noun compounds. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition, 33*, 1162–1169. **http://doi.org/cqppsd**

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4,* 295–312. **http://doi.org/dhh7hs**

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology, 48*, 108–124. **http://doi.org/b8m625**

Towse, J. N., Hitch, G. J., & Hutton, U. (2002). On the nature of the relationship between processing activity and item retention in children. *Journal of Experimental Child Psychology*, *82*, 156–184.

Towse, J. N., Hitch, G. J., Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory and Cognition*, *28*, 341–348.

Traxler, M. J. (2012). *Introduction to psycholinguistics*. Chichester, West Sussex: Wiley-Blackwell.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39,* 558–592.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.

Van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*, 289–296. **http://doi.org/b486r4**

Van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (2007). Eye movement research: An overview of current and past developments. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 1–28). Oxford: Elsevier.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25,* 315–339. **http://doi.org/fnj6wz**

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin Review*, *19*, 73–80. **http://doi.org/c948jp**

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, *97*, 254–275. **http://doi.org/5jq**

Winke, P., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eye movement recordings in second language research. *Studies in Second Language Acquisition*, *35*, 205–212. **http://doi.org/5jr**

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*, 419–456. **http://doi.org/d8qjhc**

## NOTES

[1] The trade-off between processing and storage operating interdependently within WM is based on the notion that WM's limited pool of resources are shared continuously between the two functions, with an increase in processing demands resulting in a correlative decrease in the resources left available for storage and vice versa (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992). Controversies exist, however, about the continuous interdependence of the two components. Some critics argue that the two components are functionally independent, each relying on different aspects of the cognitive system (e.g., Duff & Logie, 2001). Others advocate alternation as the key to the operation of the components: they maintain that the critical determinant of span is task switching dependent on the time spent processing versus recalling. That is, the timing of the processing component (or the time spent away from storage) is what drives span performance (e.g., Towse & Hitch, 1995; Towse, Hitch, & Hutton, 2002).

[2] As the mechanisms responsible for determining when to move the eyes are chiefly independent of those responsible for determining where to move the eyes next and how often to move them, the location, duration, and frequency of EMs are the key variables pointing to the degree of processing difficulty in reading and, more importantly, the linguistic nature of the difficulty itself.

[3] Regressions-out are said to be "the most randomly varying eye movement behavior involved in reading" because factors that trigger them are difficult to capture (Traxler, 2012, p. 405). Thus, regression-out measures have been described as both early- and late-stage measures (e.g., Clifton et al., 2007; Holmqvist et al., 2011; Sagarra & Ellis, 2013; Winke, Godfroid, & Gass, 2013). Given that the frequency of ROs from the sentence-final region is taken into account in the present study, we deemed it appropriate to treat RO measures as manifestations of late-stage processing skills as distinct from early-stage measures like FPs (see Van Gompel & Pickering, 2001).

[4] The B1 level on the CEFR corresponds to a TOEFL iBT score of 57-86. Given the rather wide range, care was taken to select participants all of whom had obtained a score of 80 on the test in order to ensure that the learners were equal in L2 proficiency level, with a view to minimizing its influence. We hypothesized that 80 constituted a score that was on the upper end of the "Threshold" range in CEFR, making it possible to refer to our participants as "upper-intermediate" learners.

[5] Developed in the 1980s, the British National Corpus is a 100 million word corpus, which is stored and distributed by Oxford University Computer Service (http://www.natcorp.ox.ac.uk).

[6] There were other reasons for administering the test in English. One stemmed from the typological differences between Turkish and English. Because the former, as an agglutinative language, is morphologically very rich compared with the latter, the same proposition can be expressed in considerably fewer words in Turkish (e.g., *Polis memuru yoldaki koyunu ezen dikkatsiz kamyoncuyu tutuklamak istedi*) than in English (e.g., The police officer wanted to arrest the careless trucker who had run over the sheep on the road). This certainly engenders serious obstacles to maintaining the number of words around 12-14 words per RST sentence, which is normally a given in RST research conducted in English as L1. The other reason involved the effect for word order: unlike English, Turkish is SOV, making it necessary for RST sentences to end in a verb. However, this could be a confound in that even in

L1 acquisition studies verbs are found to develop more slowly than nouns due to their greater conceptual complexities (e.g., Gentner, 1982), not to mention the neurally evidenced noun-verb dissociation displayed by functional neuroimaging studies with behavioral implications for retrieval of one category over the other (e.g., Cappelletti, Fregni, Shapiro, Pascual-Leone, & Caramazza, 2008).

[7] Given that an educated native speaker of English is found to need between 225 to 400 ms to process a single word in English (Rayner & Pollatsek, 1989), we estimated that our post-intermediate L2 learners would need a total of approximately 7 sec to process each RST sentence. This was based on the premise of 14 words x 400 ms equaling 5.6 seconds allocated for sentence processing, and 1.5 sec allocated for making the plausibility judgment (see Conway et al., 2005, p. 772). Our estimation proved to be accurate based on the results of a pilot study conducted with a similar group of eight learners.

[8] Rayner (2009) indicates that the average fixation duration in reading is approximately 225-250 ms, with fixation durations being nonetheless as short as 50-75 ms and as long as 500-600 ms (p. 1460).