ABSTRACT

The present study brings together methods form extensive reading studies and eye movement research to track the cognitive effects of exposure frequency on vocabulary processing and learning. Forty-two advanced second language learners of English read a stage 1 graded reader, '*Goodbye Mr. Hollywood'*, on a desk-mounted eye tracker screen followed by comprehension questions and vocabulary posttests. Target vocabulary consisted of 20 pseudo words and 20 known words with a range of repetition from 1 to 30. Eye-movement data showed that readers spent more time on pseudo words than on familiar words and that fixation times decreased across encounters with more attention given to target words on early encounters. Repeated exposure supported form recognition but was not as significant for meaning recall and recognition. Total times spent on each encounter was positively associated with learning success in all vocabulary measures. The amount of attention, as reflected in total reading times on each pseudo word across all its encounters, positively predicted learning outcomes above and beyond total exposure. Results of the study add a cognitive dimension to the concept of engagement in lexical learning and provide implications on the process of incidental learning from extensive reading.
Keywords: incidental vocabulary learning; extensive reading; eye movement; exposure frequency; attention; total reading times

Research on extensive reading has provided ample evidence on the role of repetition in lexical learning from L2 reading (e.g. Chen, & Truscott, 2010; Horst, 2005; Rott, 1999; Waring, & Nation, 2004; Webb, 2007, 2008). On the other hand, eye movement studies on reading behavior documented the cognitive effects of repetition on lexical processing and associated lexical retention in terms of online processing patterns and the eye-mind link hypothesis (Godforoid et al., 2013; Juhasz, & Pollatsek, 2011; Pellicer-Sanchez, 2015; Rayner, & Well, 1996). The present study aimed to bring together methods from both strands to investigate incidental vocabulary acquisition from L2 reading and track the cognitive effects of repetition on the development of different aspects of vocabulary knowledge.

BACKGROUND

Extensive reading and L2 vocabulary

Although incidental learning has been challenged as slow and inefficient in terms of acquisition and retention (e.g., Laufer, 2003, 2005; Macaro, 2003; Read, 2004), many researchers and teachers believe it is an essential supplement for learners to expand their vocabulary independently (see Schmitt, 2008, 2010) However, when learners are directed to extensive reading of authentic text, they usually face a lexical coverage challenge. Nation (2001, 2006) calculated that the percentage of known words in a text should range between 95% and 98% in order for learners to obtain a sufficient comprehension level. It was thus calculated that authentic novels require at least a vocabulary size of 8000 to 9000 word families for adequate comprehension and new vocabulary intake (Hu & Nation, 2000; Nation & Wang, 1999; Waring & Nation, 2004).

Because it can take several years for L2 learners to reach higher levels of vocabulary size, extensive reading programs have taken advantage of simplified graded readers that are systemically adjusted to different levels. One important advantage of these readers is that they can provide spaced repeated exposures to new and low frequent vocabulary and reinforce partially known words, which is an ideal setting for incremental vocabulary development. Grabe and Stoller (2011) defined extensive reading as reading that exposes learners to "large quantities of material within their linguistic competence" (p.259). Proponents of extensive reading reported its value in increasing reading fluency, reading comprehension, and speed of access to frequent words as well as providing opportunities to meet new words, infer new meanings and build larger mental lexicons (Day & Bamford, 1998; Elley, 1991; Horst, 2005; Lai, 1993; Parry, 1991). One important benefit of extensive reading was reported by Uden, Schmitt, and Schmitt (2014) who found evidence that graded readers can support a smooth transition to authentic novel reading.

Several studies investigated the potential of lexical gains from graded readers and authentic novels. The classic study of Saragi, Nation and Meister (1978) used the novel *A Clockwork Orange* (1962) by Anthony Burgess. It was of particular interest because it included Russian slang words, referred to as *Nadsat*, which were targeted in reading experiments. They found that native English speakers were able to learn an average of 76 % of 90 Russian slang words used in the novel. Pitts, White and Krashen (1989) used one chapter of the same novel with second language readers and found modest rates of learning, about 6.4 % to 8.1 % of 30 target Russian words. Day, Omura and Hiramatsu (1991) reported that Japanese EFL learners learned an average of 3 words out of 17 target words encountered in a simplified short story, *The Mystery of the African Mask*. Horst, Cobb and Meara (1998) had learners read a simplified

version of *The Mayor of Casterbridge*, and reported that learners could pick up an average of 5 words out of the 45 target words. Horst (2005) showed that readers picked up around 51 % of the target words from selected extracts of graded readers. A common factor among all these studies was frequency of exposure in that learning chances increased as learners encountered target words more times in the text.

In addition to word meaning, the acquisition of other aspects of lexical knowledge was also investigated in extensive reading. Waring and Takaki (2003) used the 400 headwords graded reader *A Little Princess*. They found that learners scored higher in meaning recognition of the target words than productive translation and that scores in both tests dropped sharply after three months. Pigada and Schmitt (2006) found that a French learner showed considerable improvement in word spelling but a lesser command of meaning and grammatical knowledge after one month of extensive reading especially as exposures with target words increased. Webb (2007) reported that vocabulary encounters in reading or writing positively reinforced spelling, associations, syntax, grammatical functions, and form-meaning mapping. He found that the group that encountered the target words more than 10 times showed a better grasp of different aspects of word knowledge than other groups who received fewer exposures. Pellicer-Sanchez & Schmitt (2010) investigated vocabulary learning outcome from an authentic novel *Things Fall Apart*, and found that meaning recognition reached 84 % after ten exposures while meaning recall was still around 55 %.

Cognitive perspectives of lexical learning

Based on Schmidt's (1990) noticing hypothesis, vocabulary researchers assume that readers need to notice novel words in context based on text properties or lexical features, and that this pattern of noticing would determine the nature of learning outcomes. However, it is

difficult to test this assumption offline because retrospective measures that have been used to track noticing such as note taking, underlining or think-aloud protocols can be less sensitive in capturing moment-by-moment processing of context. Godfroid et al (2013) reviewed these measures, concluding that a more precise and complete account of cognitive processing during reading can be fulfilled by the eye tracking technique, which can provide a more sensitive measure of the amount and locus of attention during processing.

Reviews of eye tracking research show that eye movements provide an accurate representation of the cognitive processes in the reader's mind. This assumption was coined the 'eye-mind' link, which proposes a connection between overt and covert attention (Rayner, 1998, 2009). In reading research, many variables were tested such as word properties, such as frequency, predictability, familiarity and other context variables in order to examine their effects on reading behavior as measured by eye tracking.

Eye movement models

A large amount of research used recordings of eye movements to explore the psychological processes that control the reading behavior of adult skilled readers (see Rayner, 1998, 2009 for a review). Several computational models were developed to explain the characteristics of reading behavior based on the assumption that there is a strong relationship between lexical encoding and eye fixation measures (Liversedge, Gilchrist, & Everling, 2011; Van Gompel, Fischer, Murray, & Hill, 2007). These models were categorized into serial-attention and parallel-attention models. Serial attention models assume that attention is allocated sequentially to support lexical processing of one word at a time and that lexical processing causes the eye to move from one word to the next (e.g. Reader model: Just, & Carpenter, 1980; EMMA model: Salvucci, 2001; E-Z Reader model, Reichle, Rayner, & Pollatsek, 2003).In

parallel attention models, processing is shared to neighboring words due to their specific characteristics (e.g. SWIFT model: Engbert et al., 2002).

Although no model was claimed to account for the whole picture, the E-Z reader model was found to be the most comprehensive in linking lexical recognition process to eye fixations because it provided assumptions as necessary to account for sophisticated observations in reading behavior (Liversedge, Gilchrist, & Everling, 2011). Simulations of eye movements in reading studies showed that the E-Z reader assumptions and the serial attention hypothesis is sufficient to account for reading behavior in alphabetic and non-alphabetic languages (see Pollatsek, Reichle & Rayner, 2006; Rayner, Ashby, Pollatsek, & Richle, 2004 for a full review).

Eye movement research in reading

Many eye movement studies have looked at native and nonnative speakers' processing of written input and responding to different lexical and contextual features (e.g., Ashby, Rayner, & Clifton, 2005; Clifton, Staub, & Rayner, 2007). Hyönä and Niemi (1990) used the repeated reading paradigm with Finnish readers. The readers' fixation times decreased consistently from first to third encounter with target sentences, and the number of their progressive fixations and regressions also decreased. Similarly, Raney and Rayner (1995) investigated the effects of repeated exposure on native-English speaker's second reading performance. They found that individuals had shorter reading times, made fewer fixations, and had longer saccades during the second reading of the same text. Moreover, shorter fixation durations were associated with high frequency words, suggesting independent effects of word frequency and repetition on reading times. Rayner, Raney, and Pollatsek (1995) found similar results regarding the effect of three repetitions of lexical items in a given text, and they also found frequency effects after the first two repetitions, but no further differences occurred after that, which indicated that word

frequency was mitigated by repetition. Recently, Joseph, Wonnacott, & Nation (2014) found significant decreases in reading times as a function of repated exposures and shorter reading times for novel words that were presented earlier in the text than later items. Early presneted words were remembered more accurately in an offline post test. In their results, they advocated an important effect of age of aqcuisition on lexical processing and learning.

Few studies have investigated a potential association between online processing patterns and learning new words. Chaffin, Morris, and Seely (2001) found that the familiarity of target words and context quality (informative or neutral) determined the amount of time readers spent on the target words in that learners fixated the most on novel words encountered in neutral contexts. Williams and Morris (2004) examined the effect of word familiarity in reading comprehension and word recognition. They found that readers spent more processing time on novel words than familiar words, and that there was a systematic relationship between online processing patterns (i.e. reading times), and retention of new word meanings. Brusnighan and Folk (2012) conducted a self-paced reading study on incidental vocabulary learning. They found that readers spent more time processing sentences that contained novel compound words, and that they were able to retain new word meanings from a single exposure. They made a case for a strong relationship between increased processing times and accuracy in vocabulary retention measures. They stated that skilled readers spend extra time on difficult items to establish form-meaning connections, which results in memory traces which are available for later recall. Godfroid et al. (2013). They operationalized attention to novel pseudo words as a quantitative variable reflected in the participants' eye fixation times during reading. Twenty-eight advanced EFL learners read 12 paragraphs in English with target areas that consisted of known words, pseudo words or a combination of both. Results showed that readers fixated longer on pseudo

words than on known words, regardless of whether these pseudo words were combined with appositive cues. There was a significant association between the total fixation time on pseudo words and subsequent recognition of these words in a surprise posttest. Recently, Pellicer-Sanchez (2015) tracked the role of repeated exposure to novel words as compared to familiar words, finding that repetition increased reading fluency and that reading times were associated with vocabulary gains of form and meaning.

Goals of the study

The goal of the present study is to bring together methods from extensive reading tradition and eye movement research to investigate the online aspects of incidental vocabulary learning from L2 reading with a focus on the effects of repeated on processing and vocabulary intake. Tracking moment-by-moment interaction with the text can provide a cognitive picture of the factors that increase or decrease attention to target words, as reflected in online fixation measures. Participants in the present study read a stage 1 graded reader on an eye tracker screen where they encountered novel and familiar words in a range of repetition form 1 to 30. They were subsequently tested on form recognition, meaning recognition and meaning recall of target words. To investigate the holistic effects of attention and exposure, I tested the role of summed online measures, the total times readers spent on individual target words, in predicting the variance in vocabulary outcomes, and whether these processing aspects override or support the roles of total exposures of novel words in L2 reading environment.

CURRENT STUDY

Research questions

The current study is guided by the following research questions:

RQ1. How do learners of English in the study process novel lexical items in silent reading relative to known control items? And how do exposure frequency influence lexical processing of pseudo words compared to control words in the text?

RQ2. What is the effect of exposure frequency on the acquisition of receptive and productive knowledge of form and meaning of target words in vocabulary posttests?

RQ3. To what extent do eye fixation times and total reading times on target words predict the learning gains of L2 readers in the vocabulary knowledge posttests?

Participants

The participants in this study were 42 advanced second language learners of English (22 females and 20 males) ranging in age from 19 to 35 (M = 22, SD = 4.2). Thirty participants were undergraduate international students with diverse majors who were also enrolled in advanced ESL reading and writing classes, and 12 participants were graduate students mostly majoring in scientific and engineering fields. Participants represented different language backgrounds including Chinese (N=13), Arabic (N=4), Spanish (N=5), Portuguese (N=5), Japanese (N=5), African languages (N=5), Hindi (N=2), in addition to single representations for Korean, Polish and Russian. Proficiency levels were determined based on self-reports of recent TOEFL IBT scores that ranged from 79 to 100 (M =89, SD =7.3). The minimum of 79 in TOEFL is the cut-

off required for undergraduate studies at MSU. Their vocabulary sizes, measured at the 5k level using Meara's (1992) vocabulary size test, yielded an average of 3908 (*SD* = 659).

Material

Vocabulary size test. To confirm that students' vocabulary levels matched the selected reading material, a yes-no vocabulary size measure, adapted from Meara (1992), was planned to be administered prior to the experimental session for each participant. The test comprised 5 levels targeting the first 5,000 most frequent words according to Nation (2001). Each level contained 60 words (40 real words and 20 non-words). The score on each level is calculated based on the estimation of hits (real words checked as known) against false alarms (non-words checked as known). A participant's vocabulary size at 5k is estimated as the sum of scores across the five levels multiplied by 10. This particular test was selected for its quick administration besides the fact that the experimental reading material will be targeting the 5k level in lexical coverage. Examples of this test can be found online at (http://www.lextutor.ca/tests/).

Reading material. The search for a graded reader involved consultations with ESL teachers and browsing the library resources of the English Language Center. Several short novels were inspected for content and length and then run through the Range software (Heatley, Nation and Coxhead, 2002), which lists the words in a given text according to their frequency and word families. The final selection was a short novel Goodbye Mr. Hollywood by John Escott, which is a stage 1 (400 headwords) graded reader made available through the Bookworms Library, Oxford University Press. It is available in print with a word count of 5400 (642 types and 372 word families) and classified under thriller and adventure stories. The text was cut down to 4649 words (595 types and 394 word families) by adjusting encounters of target words and taking out unnecessary details. Accordingly, the lexical density (types/tokens ratio) was not high (12.9%).

Range output confirmed that the lexical coverage of the story is at the 5,000 word level. Table 1 outlines the lexical distribution of the text across frequency levels.

TABLE 1

Lexical Profile of "Goodbye Mr. Hollywood"

Word List	Tokens (*percentage*)	Types (*percentage*)	Families
1000	4074 (87.6%)	479 (80.5%)	328
2000	309 (6.65 %)	61 (10.25%)	49
3000	38 (0.82%)	12 (2.02%)	9
4000	13 (0.28%)	6 (1.01%)	4
5000	22 (0.47)	6 (1.01%)	4
Not in the lists	193 (4.15%)	31 (5.21%)	
Total	4649	595	394

Target words. The list of the target words consisted of 40 items with occurrences ranging from 1 to 30. These words were equally split into two lists (20 items each), of which a given participant saw one list as experimental items (i.e., pseudo words) and the other list as familiar English controls. Each vocabulary item in the first list matched another item in the second list in part of speech, and number of letters and syllables. Because the graded reader contained all familiar words that were estimated to be a part of participants' lexical repertoire, the experimental items in each version were replaced by matching pseudo words retrieved from online resources especially the ARC Nonword Database (http://www.cogsci.mq.edu.au/~nwdb/) and previous vocabulary research (Godfroid et al., 2013; Webb, 2007, 2008). The pseudo words in one version of the story appeared in their familiar forms in the other version and vice versa. With this procedure, the two versions were counterbalanced and every pseudo word in a given

context had a familiar counterpart in the other text version. To minimize item effect, a single pseudo form was made to substitute two different words: one in each story version. Each pseudo item matched the real word in number of letters and syllables to minimize visual effects on eye movements. Table 2 outlines the target words in the two versions, the number of times they appeared in text and their substitute pseudo words.

The total number of pseudo tokens in each version was 121, which accounted for 2.6 % of the total tokens in the text. This guaranteed that the reading material provided approximately 97.4% of lexical coverage, which falls within the recommended lexical coverage range of (95%-98%) to ensure reading comprehension and the ability to guess novel words from context (Nation, 2006). Based on these criteria, pseudo words were inserted and the text was divided into shorter parts (seven chapters) and shorter paragraphs in preparation for programming.

Comprehension packet. A 50-item comprehension test (5-8 items per chapter) was created to monitor readers' understanding of the main content of the story. The items included a combination of true/false statements and multiple choice questions depending on the content of each chapter. The test was printed out in seven pages (one page per chapter) along with characters' illustrations copied from the story book to foster reader engagement and visualize the content.

TABLE 2

Pseudo Forms and their Frequency in the Reading Text

Version A targets	Version B targets	Pseudo words	Number of encounters
hotel	table	fozle	30
café	room	gube	18
face	desk	mave	10
stop	meet	tund	9
tall	busy	leam	7
kill	push	blef	6
party	money	toker	6
pocket	window	bannow	5
bag	gun	mot	5
picture	airport	fonteen	4
quiet	happy	dangy	4
garden	letter	windle	4
shirt	dress	neech	3
accident	hospital	redaster	3
rich	cold	dook	2
sleep	drink	tance	1
cinema	camera	pamery	1
famous	hungry	tantic	1
plane	noise	dorch	1
chair	shoes	smick	1

Vocabulary tests. Three vocabulary tests were prepared to measure form recognition, meaning recognition and meaning recall of the target pseudo words. In general, only these target words were identical in all the tests while distracter items differed. The form recognition test comprised 100 vocabulary items including the 20 target pseudo words, familiar words from the text and other sources and pseudo words out of the text. The instruction for the task is to circle only the words that were seen in the reading material. the Meaning recall test included the 20 target pseudo words in addition to 10 distracter items that represented pseudo words not in the text, familiar words from the list and other low frequency English words. The task was to recall meanings, synonyms, related words or semantic fields for the given items. Meaning recognition test was a multiple choice test with 30 items covering the target words along with other additional pseudo words, familiar words and low frequency words. Each item had five meaning options in addition to an 'I don't know' option to minimize guessing.

Procedure

Apparatus. Before any participants were invited into the lab, the reading material was programmed into the desk-mounted EyeLink 1000, an eye-tracker manufactured by SR Research (http://www.sr-research.com/). The story was copied into the Experiment Builder and set up in two versions so that a participant can selectively be assigned to one experiment file at the time of participation. The text was typed in Courier New font size 18, on a 19-inch computer monitor set up 55 cm from the participants' eyes. The font color was black on a light grey background.

The full experiment file consisted of 87 screens including introductory pages, instructions and break transitions. The main story content was thus provided in 70 screens, each containing

60-70 words in double spaced text. Minor editing was performed on the displayed text to confirm that target words did not appear in the beginning of slides and/or at the beginning and end of sentences. Each chapter was captured in a range of 7 to 11 screens. Breaks were offered at the end of each chapter in the story. Eye calibration was set to be performed at the beginning of the experiment and after the return from breaks. Participants moved across screens using a button on the right side of a hand-held controller. Drift correction was set up at the beginning of each page. Participants placed their heads on a chin and forehead rest during reading time to minimize head movements.

The reading session. Participants were randomly assigned to either version A or B in the experiment builder. Once a participant is done with a chapter, a break prompt appears on the screen. At the beginning of the break, he/she would take the comprehension packet on the side of the desk and respond only to the questions on the chapter he/she has just finished. Whenever he/she was ready, the participant would return to the chin rest and perform calibration for the following chapter. The same procedure continued with the rest of the chapters. When the participant reached the end of the story, he/she would complete the last page in the comprehension packet then move away from the eye tracker to another desk in the lab. The reading session for each participant including calibration, breaks and comprehension check took an average of 45 to 70 minutes.

The testing session. Participants took the vocabulary tests in the following order to avoid transfer effects: form recognition, meaning recall and meaning recognition. The testing session

for an average participant took a maximum of 10 to15 minutes and it was the final task required from participants.

Analyses

Definition of variables. I distinguish between online and offline effects on vocabulary outcomes. Online variables refer to the information in eye movement records that includes early measures of processing (e.g.; first fixation and first pass time) and late processing measures (e.g.; gaze duration and total time). First fixation duration captures the time of the first look at the target area (for example, a novel vocabulary word) when encountered for the first time during forward reading. Gaze duration combines first fixation duration along with any other fixation made on the target area at the initial visit before the eyes move forward or backward to the next target area. Total reading time is the sum of all fixation durations on the target area (see Winke, Godfroid, & Gass, 2013). I also report skipping rates, regressions-in and regressions-out of the interest areas. Regressions-in refer to instances when readers returned to a target word after first pass. Regressions-out refer to times when readers went back to a previous part of the sentence on first pass. These processing measures are reported for each of the target tokens as well as summed over vocabulary items to test if eye movement behavior predicts learning outcomes in token-based and item-based analyses.

Total exposure is treated as an offline item-based factor that represents the number of times a vocabulary item was seen in the text. Based on exposure, each item contributed different number of tokens. The instance of meeting a single token was labeled as an ‘encounter’. Because there are two versions of the reading material where target and control words were counterbalanced, each participant contributed reading times to two conditions: experimental and control. The factor of condition was used to describe differences in processing measures and

reading behavior between the target words (pseudo words) and control words (familiar English words). Vocabulary outcomes, scored as 0 or 1 for each item, represented three categorical dependent variables in the statistical models, one for form recognition, one for meaning recognition and the third for meaning recall.

Data structure. The data sheet was organized by subjects and item information. Each subject reported 242 observations, representing the total number of experimental and control tokens. In this fashion, the layout of the data showed items nested within subjects, and encounters nested within items. Figure 1 shows an example of this structure for a given reader in the experiment.

Figure 1 here

Figure 1. Data structure for participants and target words

Based on this hierarchical structure, I adopted a Generalized Linear Mixed Model (GLMM) to fit the appropriate regression that can accommodate multiple levels (Heck, Thomas, & Tabata, 2012). In the light of Figure 1, GLMM is conducted with two levels when we test by vocabulary item so that the model would only include subject variables and item variable in repeated measures. The model expands to three levels when we need to test the level of encounter including information about all the tokens of all items.

Reporting results. The GLMM output calculates the probability of the incidence of a dependent variable in terms of an odds ratio (OR), quantifying the predicted change in the

dependent measure as a function of a one unit increase in a given predictor (Ferguson, 2009). An OR larger than 1 indicates a positive relationship and an OR less than one indicates a negative relationship. The interpretation of OR varies according to the type of the dependent variable. For example, if encounter predicted fixation times (continuous variable) with an OR of 0.25, this would indicate that one additional encounter predicts a percent decrease in fixation times by 75 % (1 – 0.25 * 100%). On the other hand, if repeated exposure predicted form recognition (binary categorical variable) with an OR of 1.75, this would indicate that one extra exposure was associated with an increase in the odds of correct responses in form recognition by 75 % (1.75 – 1 * 100 %). In addition to the odds ratio (OR), I report the 95% confidence interval of the effect size of the predictor variable. The predictor was considered significant at the .05 level while the strength of the relationship was interepreted through OR. A strong relationship starts at OR < 0.33 or OR > 3 (Ferguson, 2009; Menard, 2010; Powers, & Xie, 2008).

RESULTS

Online reading patterns

Online reading measures were entered as continuous dependent variables in a GLMM to run a Gamma regression analysis with condition and encounter as predictors and word length as a control variable. Descriptive statistics and regression outputs are presented below for each measure.

First fixation durations. Figure 2 shows that first fixation durations on target words started at an average of 264 ms (SD = 124) and ended with 215 ms (SD = 88) while first fixations on control words started at 227 ms (SD = 86) and ended at 218 ms (SD = 88). Visually,

there were no major changes in fixation times from first to last encounter or major differences between conditions.

Figure 2 here

Figure 2. Mean fixation times (in milliseconds) on target and control words by encounter

Condition significantly predicted first fixation (*OR* = 1.08, 95% CI = [1.04, 1.12], $p < .001$). Comparing the odds ratio against the odds of the intercept, fixation times in the experimental condition was significantly longer than in the control condition and that the probability of fixating longer on a target words increases by about 2 % when the word is unfamiliar. Encounter was slightly associated with a decrease in FFD (*OR* =0.98, 95% CI = [0.967, 0.993], $p < .001$), implying that adding more encounters was associated with a negative change in first fixation durations by a factor of 1 %. There was a significant interaction between encounter and condition (*OR* = 1.01, 95% CI = [1.005, 1.03], $p < .001$), which implied that target and control words started to behave similarly as encounters increased. Table 3 summarizes the regression output for the effect of text-based factors on first fixation durations.

TABLE 3

Effects of Text-Based Factors on First Fixation Durations (FFD)

	Odds	OR	95% CI		*p*
Intercept	5.022		4.82	5.22	< .001 ***
Condition		1.08	1.04	1.12	< .001 ***
Encounter		0.98	0.96	0.99	< .001 ***
Condition * Encounter		1.01	1.005	1.03	< .001 ***

Note: The (*) marks signify the level of significance of the *p* value

Gaze durations. Figure 3 shows that gaze durations (GD) on target words started at 393 ms (SD = 282) and ended at 237 (SD = 118). In the light of Table 4, condition was a significant predictor of the variance in gaze durations (*OR* = 1.24, 95% CI = [1.17, 1.32], $p < .001$). Comparing the odds of the intercept with odds ratio, the probability of longer gaze durations on pseudo words increased by around 16 % when target words were unfamiliar. Each additional encounter predicted a decrease in gaze duration by about 5 % (*OR* = 0.95, 95% CI = [0.94, 0.97], $p < .001$). There was a small interaction between encounter and condition (*OR* = 1.04, 95% CI = [1.01, 1.10], $p < .001$). Table 4 summarizes predictors of gaze durations.

Figure 3 here

Figure 3. Mean gaze durations (in milliseconds) on target and control words by encounter

TABLE 4

Effects of Text-Based Factors on Gaze Durations (GD)

	Odds	OR	95% CI		*p*
Intercept	4.18		3.84	4.53	< .001 ***
Condition		1.24	1.17	1.32	< .001 ***
Encounter		0.95	0.94	0.97	< .001 ***
Condition * Encounter		1.04	1.01	1.10	< .001 ***

Note: The (*) marks signify the level of significance of the *p* value

Total reading times. Figure 4 shows that total reading times recorded highest on the first encounter of target words (M = 702 ms, SD = 512) and lowest by the final encounter (M=265 ms, SD=130). Table 5 shows that condition was a significant predictor of the variance in total reading times (*OR* = 1.68, 95% CI = [1.42, 1.77], $p < .001$), indicating that readers took longer times on target words than known words. Each additional encounter was associated with a

decrease in total times (*OR* = 0.88, 95% CI = [0.87, 0.91], $p < .001$), which was modulated by a small interaction between encounter and condition (*OR* = 1.020, 95% CI = [1.016, 1.023], $p < .001$). Table 5 summarizes predictors of total reading times over encounters.

Figure 4 here

Figure 4. Mean total reading times (in milliseconds) for target and control words by encounter

Table 5

Effects of Text-Based Factors on Total Reading Times (TFD)

	Odds	OR	95% CI		*p*
Intercept	3.30		2.88	3.73	< .001 ***
Condition		1.68	1.42	1.77	< .001 ***
Encounter		0.88	0.87	0.91	< .001 ***
Condition * Encounter		1.020	1.016	1.023	< .001 ***

Note: The (*) marks signify the level of significance of the *p* value

Skipping. Skipping was more frequent in the control condition than in the experimental condition (around 21% of target occurrences and almost 26% of control occurrences). Condition was a significant predictor of skipping (*OR* = 0.82, 95% CI = [0.71, 0.95], *p* = .007), meaning that the odds of skipping decreased by around 18 % when the target words were familiar. The effect of encounter was not significant (*OR* = 0.96, 95% CI = [0.90, 1.03], *p* = .273).

Regressions-in. Readers returned more to target items (almost 25%) than to control items (almost 15%). Condition was a strong predictor of regression-in rates (*OR* = 2.79, 95% CI = [2.42, 3.22], *p* < .001), which indicated that the odds of regressing-in significantly increased by 2.79 times when the target was unfamiliar. Each additional encounter decreased the odds of regressing to the target word by about 28 % (*OR* = 0.72, 95% CI = [0.67, 0.77], *p* < .001), implying that regressions-in were more frequent in initial encounters.

Regressions-out. Regressions-out occurred on almost 27 % of target observations and on 22 % of control observations. Condition was a significant predictor of regression-out rates (*OR* = 1.21, 95% CI = [1.09, 1.34], *p* < .001), which shows that the odds of regressing-out increased by about 21% when the target was unfamiliar. Each additional encounter decreased the odds of regressing out of the interest area by 2 % (*OR* = 0.98, 95% CI = [0.97, 0.99], *p* = .010). Table 6 summarizes text-based effects on behavior patterns.

TABLE 6

Effects of Text-Based Factors on Skipping and Regression Rates

	Skipping		Regression-in		Regression-out	
	OR	*p*	*OR*	*p*	*OR*	*p*
Condition	0.82	.007 **	2.79	< .001 ***	1.21	<.001***
Encounter	0.96	.273	0.72	< .001 ***	0.98	.010 *
Condition * Encounter	0.98	.582	0.90	< .001***	0.96	.187

Note: The (*) marks signify the level of significance of the *p* value

Vocabulary knowledge gains from reading

In overall vocabulary measures participants reported the highest gains in form recognition, followed by meaning recognition and finally meaning recall. Table 7 indicates that participants were able to retain the forms of an average 42 % of target words while they recognized the meanings of 30 % of the words and recalled the meanings of only 13 % of the same target items.

TABLE 7

Average Word Gains in the Vocabulary Post Tests

Test	*M* (*SD*)	Percentages (%)	Minimum	Maximum
Form recognition	8.36 (3.16)	41.8	1 (5%)	16 (80%)
Meaning Recognition	6.06 (3.27)	30.3	1 (5 %)	13 (65%)
Meaning recall	2.59 (2.32)	12.9	0 (0 %)	8 (40 %)

To investigate the effect of amount of exposure on vocabulary learning, I analyzed participants' responses by exposure bands (refer to Table 2) to estimate how many hits (correct

responses) each item received from participants in each test. Figure 5 reveals a wide difference between highest and lowest exposure bands but variable patterns were noted for middle bands particularly in meaning recognition and recall.

Figure 5 here

Figure 5. Mean percentages of vocabulary gains in the vocabulary posttests by exposure bands

Text-based characteristics and vocabulary learning

Controlling for item effects and word length, logistic regression output showed that total exposure was a significant predictor for all the vocabulary outcomes but to somewhat different degrees: form recognition (*OR* = 1.21, 95% CI = [1.05, 1.40], $p = .010$), meaning recognition (*OR* = 1.29, 95% CI = [1.15, 1.44], $p < .001$), and meaning recall (*OR* = 1.42, 95% CI = [1.27, 1.61], $p < .001$). By comparing the odds ratios with the odds of the intercept in the three models, we calculate the difference between the probability of learning outcomes and the baseline probability of the intercept [OR * odds/ (odds+1)]. Regression output in Table 8 indicated that each additional exposure increased the probability of form recognition by around 2 %, meaning recognition by around 3 % and meaning recall by 2 %.

TABLE 8

Regression output for the effects of exposure on vocabulary knowledge gains

	Odds	OR	95% CI		*p*
Intercept	0.13		0.054	0.30	< .001 ***
Form recognition		1.21	1.05	1.40	.010 **
Meaning recognition		1.29	1.15	1.44	< .001 ***
Meaning recall		1.43	1.27	1.61	< .001 ***

Note: The (*) marks signify the level of significance of the *p* value

Real time processing and vocabulary learning

In this section, I investigate real time processing of the target words in text; in other words testing incidental learning while reading and how moment-by-moment eye movement measures can predict that certain vocabulary items will be acquired from reading. Because the data included items nested within subjects and encounters nested within items, I fitted a binary logistic regression in a three-level GLMM for each vocabulary test. Online reading measures were entered as fixed factors with subjects and items as random factors and word length and total exposure as control variables.

Results yielded significant positive relationships of form recognition with first fixation durations (*OR* = 1.21, 95% CI = [1.13, 1.32], $p = .035$) and total reading times (*OR* = 1.42, 95% CI = [1.12, 1.80], $p = .004$), indicating that a one second increase in first fixations and total times spent on a target occurrence increased the probability of form recognition success by 4 % and 7 % respectively. Meaning recognition results pointed to a positive effect of total reading times on vocabulary outcomes (*OR* =1.33, 95% CI = [1.03, 1.72], $p = .029$). A one second increase in reading times of each token increased the probability of meaning recognition by 3 %. Meaning

recall was significantly predicted by gaze durations (*OR* = 2.19, 95% CI = [1.22, 3.77], *p* = .005) and total reading times (*OR* = 1.73, 95% CI = [1.14, 2.63], *p* = .010). One additional second spent on target tokens increased the probability of meaning recall by 3 %. Table 9-11 summarize token-based predictors of vocabulary gains.

Table 9

Token-based predictors of form recognition

	Odds	OR	95% CI		*p*
Intercept	0.26		.075	0.93	.038 *
First fixation duration		1.21	1.13	1.32	.035 *
Gaze duration		1.16	0.73	1.82	.533
Total time		1.42	1.12	1.80	.004 **

Note: The (*) marks signify the level of significance of the *p* value

Table 10

Token-based predictors of meaning recognition

	Odds	OR	95% CI		*p*
Intercept	0.10		.034	0.32	< .001 ***
First fixation duration		2.65	0.81	8.67	.106
Gaze duration		1.54	0.71	1.88	.560
Total time		1.33	1.03	1.72	.029 *

Note: The (*) marks signify the level of significance of the *p* value.

Table 11

Token-based predictors of meaning recall

	Odds	OR	95% CI		*p*
Intercept	.051		.008	0.30	.002 **
First fixation duration		0.77	0.26	2.31	.650
Gaze duration		2.19	1.22	3.77	.005 **
Total time		1.73	1.14	2.63	.010 *

Note: The (*) marks signify the level of significance of the *p* value

The role of cumulative online processing in vocabulary learning

Because summed fixation times reflected the cumulative processing effort devoted to target words, it was interesting to test how these measures would compare with offline total exposure in explaining the variance in vocabulary outcomes. I fitted binary logistic regressions using a two-level GLMM because holistic effects based on items rather than the encounter level are of interest.

Total reading times spent on target words positively increased the chances of learning form (*OR* = 2.16, 95% CI = [1.21, 3.38], $p < .001$), indicating that looking for one extra second at target words increased the probability of form recognition success by 13 %. At the level of text-based features, total exposure positively influenced form recognition although the effect was somewhat smaller than online processing times (*OR* = 1.29, 95% CI = [1.18, 1.41], $p < .001$).

Meaning recognition was significantly predicted by summed reading times (*OR* =1.47, 95% CI = [1.25, 1.72], $p < .001$) and total exposure (*OR* =1.38, 95% CI = [1.21, 1.58], $p < .001$). Meaning recall followed the same pattern with total reading times (*OR* =3.27, 95% CI = [1.28,

5.33], $p < .001$) and total exposure (*OR* =1.27, 95% CI = [1.13, 1.41], $p < .001$). Table summarizes online and offline predictors of vocabulary learning.

Table 12

Regression output of the online vs. text-based predictors of vocabulary gains

	Odds	*OR*	95% CI		*p*
Form recognition					
Intercept	*0.18*		0.094	0.35	< .001 ***
Summed TFD		2.16	1.21	3.38	< .001 ***
Total exposure		1.29	1.18	1.41	< .001 ***
Meaning recognition					
Intercept	*0.15*		.060	0.38	<.001 ***
Summed TFD		1.47	1.25	1.72	< .001 ***
Total exposure		1.38	1.21	1.58	< .001 ***
Meaning recall					
Intercept	*0.033*		0.007	0.15	<.001 ***
Summed TFD		3.27	1.28	5.33	<.001 ***
Total exposure		1.27	1.13	1.41	<.001***

Note: The (*) marks signify the level of significance of the *p* value

Overview of results indicate that holding the effect of total exposure constant, summed reading times strongly predicted learning success in all vocabulary measures particularly in form and meaning recall. This might suggest that individual attention on the part of the reader can be more important in explaining vocabulary learning above and beyond repeated exposure.

General summary of results

Participants in this study read the graded reader '*Goodbye Mr. Hollywood'*, a stage 1 story of 4649 words, which was found to be well within their current English proficiency levels.

The percentage of pseudo tokens in the story was below 3 %, and the lexical coverage of the story was satisfactory (Nation, 2001, 2006). Readers met pseudo words and familiar control words in equal number of exposures ranging from 1 to 30, yielding 121 tokens in each condition. Eye movements were recorded during reading to compare online and text-based effects on incidental vocabulary learning.

Online reading patterns pointed to significant differences between attention to target items and familiar items. First fixations, gaze durations and total times decreased as a result of additional encounters to target words, pointing to a gradual increase in familiarity with pseudo words in the text as they were repeated. The decrease in reading times was more significant in early encounters (1-12) than in later encounters. Analyses of regressions and skips confirmed, as was to be expected, the extra attention devoted to pseudo words in early encounters.

Readers displayed the most learning outcomes in form recognition followed by meaning recognition and finally meaning recall. Total exposure predicted all vocabulary outcomes but with varying degrees. Token-based online processing measures demonstrated that Total time was a positive indicator of learning success in all vocabulary tests while first fixations only predicted form recognition and gaze durations only predicted meaning recall. When aggregating processing measures on all encounters, it was shown that only summed total time was positively associated with learning outcomes. After accounting for total exposure, it was estimated that a one second increase in total times can predict vocabulary learning outcomes beyond total exposure per se.

DISCUSSION

Lexical processing in repeated encounters

The first research question sought to investigate how second language readers processed unknown words in the graded reader '*Goodbye Mr. Hollywood'*, and what textual factors influenced their reading patterns in real time. It was shown that readers gave relatively more attention to pseudo words as compared to familiar words, particularly in early encounters. Gaze durations and total times were inflated between encounters 1 and 12, after which target and control words started to exhibit similar processing patterns. Steady decreases were more significant in early encounters than in later encounters. A possible explanation can be provided in the light of the E-Z Reader model that postulates different stages of lexical processing (e.g. Pollatsek, Reichle, & Rayner, 2006). In this model, first fixations reflect an early stage of familiarity check, and do not capture later events of reanalysis, word recognition or form-meaning mapping. The unfamiliarity of target words triggered subsequent fixations that fed into gaze durations, and the reported frequent regressions to target words ultimately fed into total times. This scenario may have caused the notable rise of attention exhibited in early encounters.

The fact that readers did pay more attention to pseudo words and particularly on early encounters was also confirmed by other evidence from reading behavior. In particular, skipping was less frequent on pseudo words while regressions occurred more frequently particularly in early encounters. Skipping instances occurring at pseudo words does not contradict with the fact that readers attended more to novel items. Parafoveal processing may have occurred for new words making them less likely to be skipped at first pass. Less skipping and more regressions indicated increased processing and reanalysis of target words, which may have supported the form-meaning mapping process. Repeated encounters were associated with shorter fixation

times, which is consistent with previous eye movement research (e.g. Joseph et al., 2014, Pellicer-Sanchez, 2015), suggesting a gradual increase in familiarity with target forms over time.

Early indicators of vocabulary intake

In line with previous literature on vocabulary acquisition (Nation, 2001; Schmitt, 2008, 2010), knowledge of form seemed to be the first component to develop followed by meaning recognition and finally meaning recall. These differential learning rates can be explained in terms of a progression from the lowest to the highest cognitive demands on the learner's memory. Controlling for total exposure, it was found that total time spent on individual tokens was associated with successful intake in all the three vocabulary measures. Additionally, first fixations predicted form recognition while gaze durations predicted meaning recall. The kind of associations found between online processing and different types of vocabulary gain aligns with the claim that different eye movement measures tap into different cognitive processes. Within the framework of the E-Z reader model (Reichle, Rayner, & Pollatsek, 2003; Pollatsek, Reichle, & Rayner, 2006), lexical processing has been posited to proceed in two stages: an early stage called 'familiarity check', and a later stage referred to as the completion of lexical access. The fact that first fixation durations predicted form recognition conforms to this hypothesis in that early lexical processing is largely form-focused (Reichle, Warren, & McConnell, 2009). Gaze duration, as the total duration of early processing, predicted meaning recall, which may indicate that subsequent lexical processing of form-meaning mapping and encoding into memory becomes more important with subsequent fixations on the target word. The same principle would explain why total time, as a late measure, predicted all types of vocabulary learning. Because total time marks the completion of lexical access and sentence integration, it was indicative of the total attention devoted to each individual token in the text. As the total time spent on every

encounter of a target word increased, there was more chance that the reader would retain that word in all vocabulary measures.

Combined measures of attention and exposure

It was clearly shown that summed total reading times positively predicted learning outcomes in all vocabulary measures. This confirmed an association between online processing and lexical retention as documented by previous research (Godfroid et al., 2013, Pellicer-Sanchez, 2015). The fact that summed total times were a strong predictor of learning outcomes after controlling for total exposure might indicate that individual attention to target words can explain the variance in vocabulary learning above and beyond mere repeated exposures. This finding aligns with lexical processing data which showed that readers invested more time in initial encounters checking for familiarity and reanalyzing context. From a reader's perspective, exposures were not equal in the amount of context and information they provided about target words. Thus, when we compare online times with total exposure, we are actually comparing two dimensions of exposure that I may distinguish as dynamic versus static exposure. Dynamic exposure involves the sum of all the information that readers have accrued from all encounters with a given word while static exposure mainly represents an offline scale variable; that is, a number. In the present study, the dynamic exposure captured readers' interaction with target words and all the stages of lexical integration (Reichle et al, 2009) that have contributed to the incremental development of word knowledge as a byproduct of exposure. From this perspective, it was plausible to find that the way readers utilized their repeated encounters with target words strongly predicted learning outcomes beyond encounters per se.

CONCLUSION

The present study sheds more light on the cognitive aspects of engagement (Schmitt, 2008) and involvement (Laufer, & Hulstijn, 2001), which were emphasized in vocabulary acquisition research and particularly within the incidental learning framework. Reader engagement with lexical items is reflected in online measures which capture ongoing processing of new vocabulary in different contexts. This adds another dimension to extensive reading as a source of vocabulary development, distinguishing between learning opportunities offered by the text and the expected learning outcomes based on textual features and readers' engagement.

Practical and pedagogical implications

The results of the study are mostly relevant to second language vocabulary learning and teaching. Maximizing exposure to vocabulary in rich contexts is a recommended strategy to ensure the best conditions for internalizing partially known words or acquiring new vocabulary. Exposure is not only confined to reading, but can also be extended to task-based learning where different input modalities (speaking, listening, reading and writing) can integrate vocabulary learning goals in variable contexts (Brown, Waring, & Donkaewbua, 2008).

The present study corroborates previous research on the role of extensive reading mainly in developing reading fluency along with creating possible opportunities for learning new vocabulary. Increasing reading fluency can be an early stage that sets the scene for acquiring new vocabulary. One relevant implication for extensive reading is that it can afford more familiarity with new lexical items but this does not guarantee successful internalization of new word meanings in a limited time frame. This is particularly true in light of the fact that the effects of extensive reading are longitudinal in nature. Reading programs should be evaluated over longer

periods of time, considering all factors of input, textual features and individual reading behavior as well as learners' motivation.

Limitations and further research

The current study provides additional insights in SLA vocabulary research and extends further understanding of the cognitive aspects of incidental vocabulary acquisition. As a newly integrated technology in second language vocabulary research, the eye-tracking technique can answer specific questions about learners' interaction with L2 material with considerable temporal and spatial accuracy. Implementing eye-tracking methodology in SLA is likely to open new avenues of investigation to uncover detailed cognitive processes in language acquisition in general and vocabulary development in particular.

Vocabulary learning from reading is not a byproduct of a single factor but it is rather influenced by multiple variables with variable effect sizes. What makes this type of statistical analysis complicated is that the model controls for the effects of other variables in the equation before assessing the effect of the variable of interest. Therefore, it may moderate or reduce other effects. For example, text-based variables were found to be good predictors of learning. However, their roles were moderated, showing that there were factors beyond the text that had more important roles, particularly reader's processing behavior.

A significant variable not covered in the present study is context quality. The patterns of results on the effect of exposure frequency, particularly in meaning recognition and recall, indicate there may be other factors at play in determining the quality of learning (refer to Figure 5). The fact that different eye measures predicted different vocabulary knowledge outcomes points to a possible interaction between encounters and context richness or the level of

predictability of individual tokens within the text. Further replications can consider this variable for a more inclusive explanation of lexical gains from context.

Some methodological issues need to be discussed regarding the nature of tasks and participants in the present study. Using a head mount and a chin rest during the reading task might have interfered with the natural reading behavior of readers to some extent. Further eye-tracking research can make use of more advanced techniques to maximize the ecological validity of task performance without jeopardizing the accuracy of eye movement measures. The second point concerns the use of pseudo words for the study. As learners were expected to know the real words for the target items, they may have concluded that the novel words they encountered in reading were less frequent synonyms of the words they already knew, an impression that may have reduced their motivation or cognitive effort to incorporate the new lexical items. Moreover, the lab-controlled experiment condensed the number of exposures into one experimental session, which may not exactly match the typical incremental route that learners go through in incidental learning, where repeated exposures are spaced over longer periods of time.

Vocabulary acquisition from L2 reading is usually characterized as incidental when learners are not forewarned of a vocabulary test after receiving input. In the current study, the amount of attention measured through eye movements seemed to be learner-driven because there was no external motivation that manipulated the existence or amount of attention on target vocabulary. Future research can examine how drawing attention of readers to focus on novel words in L2 input can yield different processing patterns and subsequently reflect on the amount of vocabulary gains. However, this kind of methodological manipulation should point to vocabulary gains in terms of a clear distinction between incidental and intentional learning setting.

Finally, the ideal extensive reading study will be longitudinal in nature and it evaluates learning outcomes from several readings over longer periods of times (Horst, 2005). The present study provided a model for further large-scale research that can consider a wider variety of reading material and more authentic texts with different populations of second language learners. Although eye movement research can provide precise quantitative account of lexical processing, it would be an additional asset in future studies to apply stimulated recalls or think-aloud protocols to explore qualitative aspects of attention to target words and reading fluency and their relationship to vocabulary acquisition (Rott, 2005; Rott, & Williams, 2003). Generally speaking, combining quantitative and qualitative methods to explore lexical learning from reading would add to our understanding of attention and engagement in reading comprehension and provide further implications on the process of incidental vocabulary learning from L2 reading.

REFERENCES

Brown, R., Waring, R., and Donkaewbua, S. (2008). Vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, (2)136–163.

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial processing. *Reading Research Quarterly, 47*(2), 172-190.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, amq031.

Clifton, C., Jr., 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). Amsterdam: Elsevier, North Holland.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

Day, R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7*, 541–551.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174–187.

Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*, 375– 411.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. New York: Routledge.

Grabe, W., & Stoller, F. (2011). *Teaching and researching reading* (2nd ed.). Harlow, Essex:Pearson Education.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Inside the learner's mind: Cognitive processing and second language acquisition* (pp. 169-197). Amsterdam/Philadelphia: John Benjamins.

Godfroid, A. (2012). Eye tracking. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 234-236). New York/London: Routledge.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 35*(3), 483-517. doi: 10.1017/S0272263113000119

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from http://www.victoria.ac.nz/lals /staff/paul-nation/nation.aspx

Heck, R. H., Thomas, S., & Tabata, L. N. (2012*). Multilevel modeling of categorical outcomes using IBM SPSS*. New York: Routledge.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review, 61*(3), 355-382. doi: 10.3138/cmlr.61.3.355

Horst, M., Cobb, T., &Meara, P.(1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Hyönä, J. & Niemi, P. (1990). Eye movements in repeated movements of a text. *Acta Psychologica, 73*, 259-280.

Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition, 133*(1), 238-248.

Juhasz, B. J. & Pollatsek, A. P. (2011). Lexical influences on eye movements in reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook on eye movements* (pp.873–893). Oxford: Oxford University Press.

Lai, F. K. (1993). The effect of a summer reading course on reading and writing skills. *System, 21*, 87–100.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*, 567 587.

Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics, 22*(1), 1-26.

Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook, 5*, 223-250. doi: 10.1075/eurosla.5.11lau.

Liversedge, Simon P. and Rayner, Keith (2011) Linguistic and cognitive influences on eye movements during reading. In, Gilchrist, Iain and Everling, Stefan (eds.) The Oxford Handbook of Eye Movements. Oxford, GB, Oxford University Press, 751-766.

Liversedge, S., Gilchrist, I., & Everling, S. (Eds.). (2011). *The Oxford handbook of eye movements*. Oxford University Press.

Macaro, E. (2003). *Teaching and learning a second language* : A review of recent research. London; New York: Continuum.

McCullagh, P., Nelder, J. A., & McCullagh, P. (1989). *Generalized linear models* (Vol. 2). London: Chapman and Hall.

Meara, P. (1992). EFL vocabulary test. Swansea, UK: University College, Centre for Applied Language Studies.

Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Thousand Oaks, CA: SAGE Publications, Inc.

Nation, P., & Wang, K. (1999). Graded readers and vocabulary. Reading in a foreign language, 12(2), 355-380.

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge; New York: Cambridge University Press. doi: 10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82.

Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly, 25*, 629–653.

Pellicer-Sanchez, A., & Schmidt. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*(1), 31-55.

Pellicer-Sánchez, A. (2015). Incidental L2 Vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, 1-34.

Pigada, M., &Schmitt, N.(2006). Vocabulry acquisition from extensive reading: A case study. *Reading in a Foreign Language, 22*(1), 1-28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in a Foreign Language, 5*, 271–275.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*, 1–56.

Powers, D. A., & Xie, Y. (2008). *Statistical methods for categorical data analysis*. Bingley, UK: Emerald.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 720–732

Raney, G., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49*(2), 151-172. doi: 10.1037/1196-1961.49.2.151

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch and E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9-36). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rayner,K.,&Well,A.D.(1996).Effects of contextual constraint on Eye movements in reading:A further examination. *Psychonomic Bulletin & Review, 3*,504-509.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161. doi: 10.1017/S0267190504000078

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparison to other models. *Brain and Behavioral Sciences, 26*, 445–476.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic bulletin & review, 16*(1), 1-21.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(3), 589-619.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language, 17*(2), 95-124.

Rott, S., & Williams, J. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language, 15*(1), 45-75.

Saragi, T., Nation, I. S. P., & Meister, F. (1978). Vocabulary learning and reading. *System, 6*, 72–78.

Schaffin, R., Morris, R.K., & Seely, R.E. (2001). Learning new words in context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 225-235. doi: 10.1037/0278-7393.27.1.225

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. doi: 10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. doi: 10.1093/applin/11.2.129

Uden, J., Schmitt, D., & Schmitt, N. (2014). Can learners make the jump from the highest graded readers to ungraded novels?: Four case studies. *Reading in a Foreign Language, 26*, 1, 1-28.

Van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (EDS.) (2007).*Eye movements: A window on mind and brain*. Oxford: Elsevier.

Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English speaking world, 4*, 97-110.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from a graded reader? *Reading in a Foreign Langugae, 15*, 130-163.

Webb, S. (2007).The effects of repetition on vocabulary knowledge. *Applied Linguistics,28*,46-65. doi: 10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232-245.

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*, 312-339.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue. Eye-movement recordings in second language research. *Studies in Second Language Acquisition, 35*(2), 205-212. doi: 10.1017/S027226311200085X