LEARNING WORDS UNDER INCIDENTAL AND INTENTIONAL LEARNING CONDITIONS: AN EYE-TRACKING STUDY

By

Ina Choi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2018

ABSTRACT

LEARNING WORDS UNDER INCIDENTAL AND INTENTIONAL LEARNING CONDITIONS: AN EYE-TRACKING STUDY

By

Ina Choi

The present study investigated the cognitive processes of vocabulary learning under incidental and intentional conditions using eye-tracking. It aims to find out the extent to which intentionality and time restrictions are associated with vocabulary learning; as well as the mechanism through which these relationships are mediated by attention, controlling for the effects of word length, predictability, and part of speech of target words.

Forty-four high-intermediate L2 English learners were randomly assigned to one of three different groups: no test announcement with time restriction (Group 1: Incidental Timed), test announcement without time restriction (Group 2: Intentional Untimed), and test announcement with time restriction (Group 3: Intentional Timed). The participants read an 1100-word-long reading passage twice while their eyes were being tracked. Twelve low-frequency English words in the text served as targets for word learning. In order to accurately measure noticing, two eye-tracking measures were used: total fixation duration and the difference between the observed and expected duration. After reading, participants received three surprise vocabulary tests in the following order: form recognition, meaning recall, meaning recognition.

The descriptive statistics confirmed a pattern of incremental vocabulary development with the highest scores on form recognition, followed by meaning recognition, and then meaning recall. Eye-movement data showed that Intentional Untimed and Intentional Timed (Group 2 and 3) spent similar amounts of time on target words while Incidental Timed (Group 1) paid significantly less attention to targets than the other groups. More importantly, multivariate

multilevel mediation model demonstrated the importance of attention in predicting learning success. Effects of the test announcement and the time limit were completely mediated by the total reading time on target words. The results further support the hypothesis that intentional and incidental learning differ quantitatively by showing that the effect of the test announcement was significant for the total reading time, but not on the extra attentional processing time.

Copyright by
INA CHOI
2018

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and sincere appreciation to my advisor, Dr. Aline Godfroid, for the support and guidance she has provided during the entire process of writing my dissertation. I am deeply indebted to her for her extreme patience and constant encouragement, which have helped me stay on track and keep pushing forward when the going got tough. I would like to extend my gratitude to my committee members, Dr. Susan Gass, Dr. Shawn Loewen, and Dr. Paula Winke, for their understanding, support, and feedback during this project.

My appreciation also goes to Unhee Ju and Hope Akaeze for their valuable advice on the statistical analysis and Jennifer Majorana for reading my dissertation. I am grateful for the financial support provided for the dissertation by the Journal of Language Learning with a Language Learning Dissertation Grant and by the College of Arts and Letters at Michigan State University with a Dissertation Completion Fellowship.

I am especially thankful for my colleagues and friends who have cheered me on and helped me stay steady through these many years – Yaqiong Cui, Talip Gonulal, Jihyun Park, and Lorena Valmori. The relationships and memories that I have developed with you all will always be priceless for me and I will cherish them for life.

Most importantly, none of this would have been possible without the love and encouragement of my family, including Yoon-A and In-hyuck. My special thank goes to my mother for her unconditional love, faith, and confidence in me. My most important source of support and strength has been my husband, Sunil. This is truly an accomplishment that belongs

to both of us and I would never have achieved it without your support, your love, and your belief in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Building vocabulary knowledge is the most basic and essential element of language learning and language use. According to Nation (2001), language learners need to know at least 6,000 word families to understand spoken language and 8,000 word families to understand written language. Learners' achievement of this requirement cannot be explained by explicit language learning in language classes alone. Instead, extensive reading serves as a key to increasing the size of learners' vocabulary. Extensive reading often involves a high level of incidental learning because learners do not have the intention of learning lexical items when reading, but they may pick up words incidentally in the process.

Many researchers have long recognized that incidental and intentional vocabulary learning differ in their effectiveness, but it is unclear whether such differences reflect quantitative or qualitative differences in the underlying cognitive processes. Provided that intentional learning often involves more time on tasks and attracts more attention on targets than incidental learning, learning can be just more challenging under disadvantageous conditions, suggesting that the facilitative effect of intentional learning may simply due to the longer time paid to targeted lexical forms.

Using an eye-tracking method, the current study addresses these concerns by investigating the cognitive processes of vocabulary learning under incidental and intentional conditions. Results are expected to inform researchers and practitioners as to whether intentional and incidental learning differ qualitatively or only quantitatively, contributing to a growing body of research on the second language vocabulary learning.

CHAPTER 1: REVIEW OF THE LITERATURE

1.1 Attention and awareness in Second Language Acquisition

The constructs of noticing, attention, and awareness have been explored in Second Language Acquisition (SLA) since the 1980s (e.g., Hulstijn, 1989). However, the first serious discussions and analyses of noticing emerged during the 1990s with Richard Schmidt's noticing hypothesis (Schmidt, 1990, 1994, 1995, 2001, 2010), which has proved a powerful concept in the cognitive-oriented research in SLA ever since. It has served as a theoretical framework to interpret various pedagogical phenomena including input and interaction (e.g., Gass, 1997; Long, 1996), Focus on Form (e.g., Doughty & Williams, 1998; R. Ellis, 2002; Williams, 2005), and implicit and explicit language learning (e.g., N. Ellis, 1994).

According to the noticing hypothesis, noticing is "the necessary and sufficient condition for the conversion of input to intake" (Schmidt, 1990, p.129). The hypothesis is based on Schmidt's own language learning experience in Brazil studying Portuguese (Schmidt & Frota, 1986). Analyzing his journals and recordings of his conversations with Brazilian interlocutors, Schmidt and Frota found that the linguistic features that he was able to incorporate into his speech were generally those that he had consciously noticed in Schmidt's speech. Forms were not used in production if they were not noticed although they were present in the input. Based on his findings, Schmidt maintained that noticing is a prerequisite for learning to take place and is necessarily a conscious process.

This early version of the noticing hypothesis (Schmidt, 1990) has received some criticism for several reasons. First, the weakness of the hypothesis came from the fact that the term "intake" was not clearly defined. Godfroid, Housen, and Boers (2010) contended, "this characterization does not specify exactly what intake is, other than that it is the product of noticing and an intermediary step in the acquisition process" (p. 170). Second, researchers disagreed with the premise that noticing is a necessary condition for learning. For example, in Gass, Svetics, and Lemelin's (2003) study, Italian learners in the non-focused-attention group showed greater gains than those who in the focused-attention group. Third, Schmidt's (1995) idea that unconscious learning does not exist has been contested by a series of studies and reviews (e.g., Hama & Leow, 2010; Leow et al., 2008; Rosa & Leow, 2004; Williams, 2005). Williams (2005) reported that learning without awareness can occur, and Robinson (1995) also suggested that both focal attention and awareness are required for the representation of novel linguistic forms. In later publications, Schmidt (2001) weakened his argument, proposing that noticing is at least a facilitative, if not a necessary and sufficient, condition for L2 development. However, what is important for him may be the fact that "more awareness leads to more learning" (p. 8) rather than whether awareness is necessary or not.

Due to its theoretical confusion, cognitive psychologists retired the term "noticing," and started adopting the constructs of "attention" and "awareness" (Godfroid, Boers, & Housen, 2013). "Noticing" now serves as an umbrella term that involves both attention and awareness. Schmidt also implied these two concepts in a recent publication, saying "the idea that SLA is largely driven by what learners pay attention to and become aware of in target language input seems the essence of common sense" (Schmidt, 2010, p. 721). Schmidt (2001) also highlighted the role of attention, stating that "There is no doubt that attended learning is far superior, and for all practical purposes, attention is necessary for all aspects of second language learning" (Schmidt, 2001, pp. 1-2).

In the field of SLA, many researchers have examined the role of awareness in the noticing studies through various techniques such as think-alouds, underlining, and stimulated recall interviews (e.g., Hanaoka, 2007; Godfroid & Spino, 2015, Izumi and Bigelow, 2000). However, the current study aims to focus on noticing as attention by adopting eye-tracking methodology.

1.2 Involvement Load Hypothesis

Among a variety of approaches and methods of vocabulary teaching, Laufer and Hulstijn (2001) formulated the Involvement Load Hypothesis, claiming that the degree of involvement is a key to better retention of unknown words. Involvement is regarded as a combination of a motivational and cognitive construct, which has three main elements: *need*, *search*, and *evaluation*.

According to Laufer & Hulstijn (2001), the *need* means the need to achieve, indicating how much students need the word to complete the task. T he *search* is the attempt to find the meaning, concerning with whether students have to look for the meaning or it is given directly. The *evaluation* refers to assessment if the meaning is correct, inferring whether students have to look at different meaning and figure out the correct one. The need and evaluation components can be divided into three levels (0 ~2) depending on whether learners are intrinsically motivated or externally motivated and whether the degree of cognitive processing is moderate or strong. The cognitive component search can either be absent (0) or present (1). Calculating the sum of these three components, any tasks can be rated in the range from a minimum of 0 to a maximum of 5. That is, tasks with the higher rating are considered to be more effective vocabulary tasks according to the Involvement Load Hypothesis.

Most of studies referring to the hypothesis have appeared in the area of incidental vocabulary learning and task-based learning. Word retention was found to be longer when words were looked up in a dictionary compared to when words were explained (Cho & Krashen, 1994). The marginal glosses were also found to be facilitative when compared to the control group (Hulstijn, Hollander, & Greidanus, 1996). Findings from Joe (1995, 1998)'s research were in consistent with the previous research in that engagement in tasks enhanced the acquisition of vocabulary. A series of experiments in Laufer's (2005, 2006) study supports filling the blank in the sentences using the target words led better retention than reading a text for comprehension. All in all, the results of the research seem to suggest that varying degrees of involvement load have some effects on vocabulary learning as the Involvement Load Hypothesis predicts. While those research addressed above were interpreted in light of the involvement load explanation, Hulstijn and Laufer (2001) and Kim (2008) were designed to test the involvement load hypothesis directly and empirically.

Hulstijn and Laufer (2001) explored the effects of three different tasks with the total of 225 English language learners in Israel and Netherlands. Those three tasks were designed to represent three different involvement loads: (1) reading comprehension with marginal glosses, (2) comprehension plus filling in target words, and (3) composition-writing with target words. Participants were intended to learn ten target words incidentally through the tasks. Unexpected vocabulary tests were administered right after the completion of the task and 1-2 weeks later to measure short-term and long-term retention of the words. The results of Hulstijn and Laufer (2001) were different in two groups of participants (Israel and Netherlands). In case of Israeli participants, the findings are in accordance with the involvement load hypothesis. The composition task yielded the highest score, lower score for the reading with filling the blanks task and the lowest score for the reading with marginal glossing task. In the experiment in Netherlands, however, scores from the fill-in-the-blank task and the reading with glossing task were not significantly different although participants outperformed significantly in the composition task compared to two other tasks.

The main concern of Hulstijn and Laufer (2001)'s study lies in the fact that the assigned time of each task was different: 40-45min for reading plus marginal glosses, 50-55min for fill-in-the-blank, and 70-80min for composition. Although the authors stateed that time on task is considered as an inherent property of a task, it is unclear that the difference in scores from three tasks was attributed to either time or involvement load. It is possible that the composition task yielded higher score because participants had more time to learn and remember the target vocabulary. Also, when operationalizing the levels of involvement load (need, search, and evaluate), the hypothesis assumes that the degree is the same for each element to affect involvement load. For example, the impact of strong need might be different from the impact of strong evaluation to involvement load although both are represented as the same number/degree (2) in the involvement load index. However, the ecological generalizability for the study is fairly high. The experiments were conducted during normal classes and participants were randomly selected and assigned to each task in two different countries. Considering that most of studies in SLA field have limited number of participants, Hulstijn and Laufer (2001)'s research can be considered to have a competitive number of participants (97 and 128 in each countries).

Kim (2008) also explored the effect of task involvement load on second language vocabulary learning, including participants with two different levels of proficiency (Experiment1) two different types of tasks with the same involvement load (Experiment2). Sixty-four participants were recruited for Experiment 1 and forty participants for Experiment 2. Following Hulstijn and Laufer (2001), three similar tasks were designed to operationalize different levels of involvement load: reading, gap-fill, and composition and ten words were served as targets. The difference from Hulstijn and Laufer (2001)'s study was that time was the same for all three tasks. The results of Experiment 1 and 2 were in line with the Involvement Load Hypothesis, revealing that tasks with the same involvement load induce similar outcomes in vocabulary learning and higher involvement loads results in greater vocabulary gains over time.

In Kim (2008)'s study, the Vocabulary Knowledge Scale (VKS) was used to measure participants' long-term and short-term retention of target words. The VKS, developed by Wesche & Paribakht (1996), is a self-report scale, containing five stages of differing degrees of knowledge. Wesche and Paribakht's stages are listed below:

1. I don't remember having seen this word before.
2. I have seen is word before but I don't know what it means.
3. I have seen is word before and I think it means _____________________.
4. I know this word. It means _____________________.
5. I can use this word in a sentence. For example, _____________________________.

In the VKS, the five stages are considered as a progression or succession of word learning based on the assumption that a word that was successfully produced (stage 5) is learned better than a word that was recognized. However, the use of VKS has been contested in recent years for several reasons (see Schmitt, 2010 for details). Meara (1996) claimed that the single, unidimentional scale may not accurately represent lexical development of the targeted words, Read (2000) noted that the increments between each stage cannot be assumed to indicate an equal interval, and Schmitt (2010) stated that the self-report data on stage 1 and 2 are not reliable, which does not disclose the direct representation of learners' vocabulary knowledge.

Schmitt (2010) admits, however, “no current scale gives a full account of the incremental path of mastery of a lexical item, and perhaps acquisition is too complex to be so described” (p. 224).

1.3 Understanding incidental vocabulary learning

Use of the notions, *incidental* and *intentional*, dates back to the early twentieth-century i(Hulstijn, 2003). Since the end of the century, these constructs have started receiving elevated interest in the SLA field, specifically in the domain of vocabulary (e.g., Ellis, R., 1994; Gass, 1999; Godfroid et al., 2017; Hucklin & Coady, 1999; Hulstijn, 2001, 2003; Laufer, 2005; Rieder, 2003; Schmitt, 2008, 2010; Pellicer-Sánchez & Schmitt, 2010). Despite its popularity, Husltijn (2003) commented, “incidental learning has often been rather loosely interpreted in common terms, not firmly rooted in a particular theory” (p. 357). Interpretation of incidental vocabulary learning in the existing literature can be categorized in one of two ways: *classroom-oriented* and *attention-oriented* (Sok, 2014).

First, in the *classroom-oriented interpretation*, incidental and intentional learning is explained in the frame of classroom instructions. Incidental learning refers to the learning that occurs when the pedagogical purpose of instruction is not language, whereas intentional learning is described as the type of learning that is designed and intended to focus on the formal information being learnt. Content-based instruction (CBI), learning of language while studying content matter subjects, is a good illustration of how incidental learning is viewed from the classroom-oriented perspective. Grabe and Stoller (1997) report theoretical and experimental support for CBI in second language acquisition research, stating “language is best acquired incidentally through extensive exposure to comprehensible input in content-based classrooms” (p.6).

The classroom-oriented distinction between incidental and intentional learning can be related to the two major types of form-focused instruction (FFI): Focus on Form (FonF) and Focus on Forms (FonFs). Although there is still debate on the definition and operationalization of the two instructional practices (Loewen, 2011), FonF is generally considered a teaching approach that puts a primary focus on communication and meaning and occasional while incidental attention to linguistic forms is provided when the need arises. In contrast, FonFs means lessons in which language features are taught or practiced in isolation without contextual connections, which seems to coincide with the construct of intentional learning. Laufer (2006), for example, compared the effectiveness of FonF and FonFs approaches in vocabulary learning with 158 English learners in Israel. Participants in the FonF condition were invited to read a 165 word-length text and encouraged to use the bilingual dictionary when it was needed while participants in the FonFs condition studied a list of target words with their meanings and explanations in English and completed two word-focused exercises. A surprise vocabulary test was administered to the participants in both groups and their scores were subsequently analyzed by the researcher. Laufer found that the FonFs group scored significantly higher than the FonF group. Based on her findings, Laufer claimed that FonFs is indispensable to vocabulary instruction due to the nature of lexical competence; it "has major importance in any learning context that cannot recreate the input conditions of first-language acquisition" (p.162).

Another example of a study that investigated incidental learning from a pedagogically-oriented perspective is Coll's (2002) study of 40 low-intermediate English language learners in a hypermedia-assisted learning environment. The participants were exposed to a set of multimedia lessons, including chemistry-related video segments. Various comprehension tools (e.g., L1 translations of questions and answers, L2 video transcript, translation of transcript sentences, etc.) were provided to make sure that the participants learn the meaning of the words in a contextualized form. The researcher recorded and analyzed the participants' actions to find out what tools were more frequently used, and administered pre- and post-treatment vocabulary tests to evaluate the learning gains. Coll concluded that the hypermedia-based instruction can be an effective way to enhance learners' retention of words when it is associated with word-related activities. Incidental learning was thus operationalized in the context of the study as "vocabulary is learned incidentally when the learning focus is on listening comprehension training" (p. 266). Coll added "vocabulary was not taught explicitly, but rather implicitly by providing the learner with verbal as well as visual input (p. 268).

On the other hand, other researchers have taken an *attention-oriented interpretation* of incidental and intentional learning. This approach views incidental vocabulary learning as the absence of conscious intention (Barcroft, 2004), directly contrasted with intentional vocabulary learning, which refers to any activity geared toward committing lexical information to memory (Hulstijn, 2001, p. 271). Thus, incidental learning is often described as learning which accrues as a "by-product" (Schmitt, 2010, p. 29) or as the unplanned picking-up of vocabulary within an activity where meaning is the primary focus. This psycholinguistic approach carries the underlying assumption that learners' attention is drawn to *meaning* during incidental learning and to *form* during intentional learning.

One of the earliest studies that looked at incidental learning using an experimental setup was Saragi, Nation, and Meister (1978). In this study, 20 native English speakers were asked to read Anthony Brugess's *A Clockwork Orange*, in which 241 Russian slang words (*nadsat*) were embedded. To keep the purpose of the experiment hidden, the researchers forewarned the participants of a comprehension and literacy criticism test afterwards. Some questions were also presented to students while reading to ensure that the students read the text for comprehension. In other words, through the forewarning of posttests and the while-reading questions, the researchers manipulated learners' intention and attention to the content of the story rather than to the individual vocabulary items. Still, results revealed an average of 76% learning gains of the 90 Russian slang words used in the novel. The surprisingly high learning gains, however, could not be reproduced in replication studies such as Pitts, White, and Krashen (1989) and Hulstijn (1992) where second language learners recorded less than 10% gains of new words. Horst, Cobb, and Meara (1998) explained that the case of native English speakers learning Russian words from an L1 context does not accurately represent the L2 learning condition.

Another study that subscribed to the attention-oriented definition of incidental learning was Waring and Takaki (2003). Waring and Takaki were concerned with the rate at which learners learn and retain new words from reading a graded reader and with the effect of frequency of exposure rates on incidental vocabulary learning. As in Saragi et al.'s (1978) study, reading was the main activity for their participants, 15 university students in Japan. Participants were told to "read the story as usual and enjoy it" (p. 141), and were also informed that they would be tested after reading without being told what kind of test it would be. Three different types of vocabulary tests (word-form recognition, prompted meaning recognition, and unprompted meaning recognition) were conducted immediately after reading, and one and three months later to measure retention. However, the researchers did not administer the comprehension test to confirm that participants' attention was directed towards on meaning nor did the researchers take any measure to prevent participants from paying deliberate attention to target words. In addition, using artificial words as targets, as Waring and Takaki did, may invite more attention because of the nonwords' saliency. In turn, this may lead participants to expect that the following test would be vocabulary-related. These limitations weaken the authors' claim that the learning gains are from incidental learning/ leisure reading, suggesting that it might be insufficient to simply assume that participants in the experimental setting would read naturally as they would do in their real life.

Although the *classroom-oriented* and *attention-oriented* perspectives of incidental vocabulary learning are not identical, we can say that vocabulary learnt as a by-product of other activities implies learning without intention to learn the lexical items. For example, a learner may pick up some words while reading a text for comprehension, and the learnt words can be regarded as a by-product. However, it is difficult to investigate what the learner actually does when encountering new lexical items while reading. That is, the learner may have intentionally and voluntarily tried to infer the meanings of certain words while reading for pleasure or simply paid attention to the unknown items without the intention to learn. That is, it is ambiguous if some degree of intentionality is involved in a supposedly incidental condition (Bruton, Garcia Lopez, & Esquiliche Mesa, 2011; Godfroid et al., 2017). In a similar vein, Huckin and Coady (1999) stated that incidental learning is not entirely incidental, as the learner must pay at least some attention to individual words (p. 190). According to Barcroft (2004), moreover, vocabulary learning is neither purely incidental nor purely intentional in a real-world context (p. 201), so incidental and intentional learning should be viewed as a continuum from highly incidental to highly intentional. Several researchers (e.g., Gass, 1999; Hulstijn, 2001, 2003) have also pointed out that the lack of consensus over the constructs of incidental and intentional learning in terms of attention and awareness, coupled with the ill-informed understanding of the terms "incidental" and "intentional," all of which may lead to misguided pedagogical implications (Hulstijn, 2001, p. 261). These controversies regarding the role of attention in incidental learning have led many researchers to prefer the *method-oriented definition* of incidental learning, which will be introduced and discussed in the next section.

1.4 Hulstijn's methodological operationalization

Due to the lack of a finely established theory and of a satisfactory operationalization of intent and learning (Hulstijn, 2003), researchers in SLA have attempted to operationalize intentional and incidental learning experimentally. On this view, intentional and incidental learning are distinguished simply based on the presence and absence of an explicit instruction to learn (Hulstijn, 2003), assuming that forewarning of a post-test invites more intentional learning. The authors that adopted the *classroom-oriented* and *attention-oriented* interpretations in the previous section also designed the incidental learning environments through conducting unexpected vocabulary tests after meaning-oriented activities such as reading or listening. However, the *method-oriented definition* is mainly concerned about the absence and presence of instruction to learn the vocabulary. That is, there can be posttests, but the explicit information about the posttests is the only criterion to determine the two types of learning within the study (Sok, 2014). Accordingly, whether learners' attention is drawn to meaning or form is neither the major concern nor the assumption behind this approach. Barcroft (2009) and Peters, Hulstijn, Sercu, and Lutjeharms (2009) adopted this methodological operationalization of incidental and intentional learning.

Barcroft (2009) compared incidental and intentional conditions in relation to the effects of synonym generation on L2 vocabulary learning. Spanish-speaking learners of English were asked to read an English text with 10 target words and their translations in parentheses next to each target word. Participants were divided into four groups according to whether an announcement of a pending vocabulary posttest was made before reading (i.e., intentional vs. incidental learning group) and whether participants were asked to write a synonym of the target words. Barcroft found that intentional learning yielded higher L2-word-form learning from reading than incidental learning. In addition, negative effects of synonym generation on L2-word-form learning were found in both incidental and intentional conditions.

Peters and her colleagues (2009) investigated how three techniques (vocabulary test announcement, task-induced word relevance, and vocabulary task) affected learners' behavior of looking up words in an online dictionary and what the effects on subsequent word retention were. The three techniques were designed to enhance learners' attention to lexical items a) by informing the learners about a pending vocabulary test to be given after the reading, b) by having them complete comprehension questions, and c) by giving them an additional vocabulary task. The results indicated that announcing the type of test to be taken after reading positively affected learners' performance on the form recognition test but not on the meaning recall tests. More interestingly, however, the authors found more significant effects on overall vocabulary learning when target words were relevant to the comprehension questions. In other words, manipulating students' attention by the test announcement does not appear to be as influential as increasing target words' salience through external intervention in vocabulary learning. In another study, Peters (2006) examined whether learners' performance is influenced by four different task instructions—announcement of a vocabulary test, combined with announcement of a comprehension test. Results revealed that participants approached the experiment similarly regardless of task instructions, which indicates that test announcement itself cannot control learners' behavior or cognitive processing.

The problem of distinguishing between incidental and intentional learning grounds in Pellicer-Sánchez and Schmitt's (2010) study. Although none of the participants in the study were informed of the post-reading vocabulary tests after reading a novel, the participant who received the highest score on the vocabulary tests expressed that she expected that the knowledge of the foreign words would be examined after reading. Consequently, she paid more attention to the target words by underlining unknown words in the novel and revisiting the words after finishing the reading. This anecdotal evidence clearly shows that learners may actually intend to learn some words when they are not supposedly induced to learn lexical items (Bruton, Garcia Lopez, & Esquiliche Mesa, 2011). Likewise, informing participants of a vocabulary test before reading does not necessarily lead learners to learn the target items during reading. Considering the limitations of the methodological operationalization of incidental and intentional learning, Hulstijn (2001, 2003) commented that the usage of the terms should not be expanded to understand learners' attention but instead of limited to explaining and discussing experimental procedures. The reason for this suggestion is that the underlying concept of the incidental and intentional learning distinction cannot be completely explained yet on a theoretical level although the distinction is fairly straightforward in operational terms. Several researchers in earlier times asserted that the dichotomous distinction between incidental and intentional learning is not valid (Postma, 1964, as cited in Hulstijn, 2001), or that complete incidental learning in an absolute sense does not exist (McGeoch, 1942, as cited in Hulstijn, 2001) and thus, the distinction should be viewed with regards to the degree of attention (R. Ellis, 1994). These issues can now be addressed with the eye-tracking methodology, which is a tool to measure the amount of attention.

1.5 Understanding eye-tracking methodology

Eye-tracking refers to the recording of an individual's eye movements. Eye-movements have been a useful data source in cognitive research since the mid-seventies to investigate underlying processes during scene perception, reading, and visual search (Rayner, 1998, 2009). Research adopting eye-tracking technology is based on the assumption that eye gaze (overt attention) provides information about cognitive processes (covert attention). According to this assumption of an eye-mind link (Reichle, Pollatsek, & Rayner, 2012), an individual's cognition in the primary drives of when and where the eyes move (duration and location). In other words, increased processing demands are associated with longer processing time and longer processing times are believed to be reflected by longer fixation durations or a larger number of fixations.

Eye-movements provide several important advantages as a measure of reading behavior relative to measuring overall reading times. Specifically, monitoring eye-movements offers multiple aspects of eye-movement data (e.g., fixations, saccade, regression). Eye-movements also reflect text features (e.g., word length, predictability) and individual reader differences (e.g., age, proficiency). However, the most favorable aspect of the eye-tracking method is that it produces "a good moment-to-moment indication of cognitive processes during reading" (Rayner, 2009, p. 1461). To assess learners' cognitive activities during performance of a certain language task, several online and offline methods have been used in SLA, including think-alouds, self-recording, note taking, and underlining. A well-known problem with these measures is reactivity (Bowles, 2010, Fox, Ericsoon, and Best, 2011). Reactivity describes the phenomenon that research treatments or instruments alter participants' performance or behavior. For example, in Sachs and Polio's (2007) study, a think-alouds protocol was employed to examine learners' attentional processes in relation to written feedback on a L2 writing revision task. The results showed that learners who were not required to make verbal reports produced significantly more accurate revisions than those who were instructed to speak their thoughts aloud while processing the feedback. Sachs and Polio acknowledged the think-alouds were reactive in their study, concluding that SLA researchers should implement and interpret think-aloud protocols with caution. Bowles (2010) conducted a meta-analysis on 12 reactivity studies, and concluded that tthink-alouds has a very small effects on task performance, but that the effects may depend on tasks and subjects.

Reviewing the reactivity issues in SLA literature, Godfroid, Boers, and Housen (2013) proposed that the eye-tracking technique can provide a more sensitive measure of the amount and locus of attention during processing. Eye-tracking methodology is beneficial in reading research in that it makes it possible to investigate readers' ongoing cognitive processing of readers without significantly altering the original characteristics of either the task or the presentation of the stimuli (Dussias, 2010). Since eye movements occur naturally as a part of reading, recording eye movements does not alter the thought processes. Although forehead and chin rests, screen layout, and font size and type may influence the reading process in eye-tracking studies (Godfroid & Spino, 2015), researchers seem to be in agreement that it is "probably the closest experimental operationalization of natural reading" (Van Assche, Drieghe, Duyck, Welvaert, and Hartsuiker, 2011, p. 93).

1.6 Eye-tracking and second language vocabulary acquisition

Eye-movement recordings in second language research have gained popularity as techniques to investigate various aspects of SLA theories (for an overview, see Conkin & Pellicer-Sánchez, 2016; Dussias, 2010; Frenck-Mestre, 2005; Siyanova-Chanturia & Roberts, 2013; Winke, Godfroid, & Gass, 2013). Godfroid and Schmidtke (2013) and Godfroid, Boers, and Housen (2013) claimed that eye-movement registration is a valuable tool in helping researchers examine theoretical models of the language learners' minds and understand the cognitive process of L2 development. Since eye-tracking methodology has been integrated into SLA research quite recently, only a handful of studies have attempted to link eye-movement behavior and second language vocabulary acquisition.

Godfroid, Housen and Boers (2010) and Godfroid, Boers, and Housen (2013) introduced eye tracking to L2 vocabulary studies and provided the evidence for the noticing hypothesis in relation to second language vocabulary. The authors investigated the role of attention in incidental vocabulary learning through four different input conditions. Participants' eye movements were monitored as they read 20 English texts that included 12 target words, whereby contextual support to infer the meanings of target words was manipulated within subjects. After the reading task, participants took a surprise vocabulary test to measure their learning. They found that participants fixated longer on novel words than on known words, regardless of whether the novel words were presented with appositive contextual cues. The results also support Schmidt's noticing hypothesis that more attention to a novel word leads to better recognition of that word on the posttest. Mohamed (2017) later extended this finding to meaning recognition and meaning recall.

In a similar vein, Godfroid and Schmidtke (2013) analyzed verbal reports from participants, showing that awareness (verbal reports) and attention (eye-fixation) are closely related. Depending of the level of awareness, eye fixation durations were found to vary. Overall, there seems to be some evidence that increased fixation during later stages of processing (i.e., second pass time and total reading time) is a positive indicator of learning success.

Several studies have specifically looked at other factors affecting the eye fixation times during reading such as frequency, familiarity, predictability, word length, and part of speech (e.g., Elgort & Warren, 2014; Godfroid et al. 2017; Mohamed 2017; Pellicer-Sánchez, 2015; Waring & Takaki, 2003; Webb, 2007).

Similar decreasing patterns in reading times across repeated exposures were observed by Pellicer-Sánchez (2015), who examined three components of vocabulary knowledge (i.e., form recognition, meaning recognition, and meaning recall) acquired incidentally from reading. In line with previous studies (Pellicer-Sánchez & Schmitt, 2010), receptive aspects of vocabulary knowledge were found to be easier to acquire than productive knowledge. Another major finding of the study was that L2 learners needed at least eight encounters to read unknown words in a fashion similar to how they read known words. Results also showed that participants who had longer reading times scored higher on the meaning recall test, again highlighting the important role of attention in vocabulary learning.

Regarding the role of frequency to target words in reading, Godfroid and colleagues (2017) extended the eye-movement investigation to a longer, more authentic reading passage and highlighted the role of word repetition in vocabulary learning. Thirty-five advanced English language learners and 19 native speakers of English read an English novel *A Thousand Splendid Suns* containing Dari words. After reading, participants performed a comprehension test and three surprise vocabulary tests. The number of exposures was found to be a predictor for all three types of vocabulary knowledge. The results further showed that the tests scores increased and processing time decreased as the readers encountered the targets more often in the text.

While there is sufficient evidence to demonstrate that frequency of exposure plays a beneficial role in vocabulary processing and acquisition, the effect of contextual clues seem to be inconclusive outside the realm of eye-tracking studies. For example, Zahar, Cobb and Spada (2001) and Schwanenflugel, Stahl and McFalls (1997) stated that the role of contextual support in assisting vocabulary learning was unclear. On the other hand, Webb (2008) found that contextual richness positively influenced meaning-related word-learning rather than form-related word-learning. This view was further supported by Hu (2013), who argues that the quality of context supports form-meaning connection and grammatical features, whereas repetition is associated with the knowledge of form. Although there has been little agreement on the effect of context on learning, eye-tracking researchers have shown that a high context predictability invites higher skipping rates and reduced processing time measured by different types of eye-fixations (e.g., Calvo & Meseguer, 2002; Brysbaert, Drieghe, & Vitu, 2005; Drieghe, Brysbaert, Desmet, & De Baecke, 2004; Rayner, Ashby, Pollatsek, & Reichle, 2004).

In L2 vocabulary research, Mohamed (2017) recorded 42 English language learners' eye-movements as they read a graded reader, *Goodbye Mr. Hollywood*, containing 20 pseudo words and 20 known words. The targets varied in the number of occurances and the level of predictability. After reading, participants were asked to take comprehension questions and vocabulary posttests. The results showed that words with rich context clues required less processing time. Mohamed also found evidence for the role of contextual support for meaning recognition and recall, which is consistent with Webb (2008) and Hu's (2013) findings. Another noteworthy finding is that the role of context predictability played a more important role in later encounters than in early encounters with target words. Mohamed (2017) explained that new, unknown words required more repetition to be recognized before participants were able to utilize the context clues to infer the meanings. So far, no previous study has directly compared the incidental and intentional learning using eye-tracking technology.

In the field of psychology, there is already a large volume of studies available about the role of word length in fixation durations (e.g., Rayner, Sereno, & Raney, 1996; Schilling, Rayner, & Chumbley, 1998). Rayner (2009) simply stated, "As word length increases, the probability of fixating a word increases" (p. 1461). In the field of SLA, Godfroid and her colleagues (2017) included word length and part of speech as control variables. They found that word length negatively affected all types of vocabulary learning and the meanings of noun were learnt more than those of other words.

1.7 Conclusion

Incidental and intentional learning are the two key mechanisms through which language learners build up their lexical repertoire. As this review has shown, many previous scholars have investigated incidental and intentional vocabulary learning in regards of its effectiveness, although they have conceptualized and operationalized incidental and intentional learning in different ways. Moreover, the use of eye-tracking technology has been growing in recent years, but most of studies have only focused on incidental learning. Consequently, to what extent incidental and intentional learning differ in terms of the underlying cognitive process still remains unclear. Therefore, in the current study, by using the eye-tracking methodology, I aim to answer to the question whether the distinction between intentional and incidental learning reflects a quantitative difference in reading time or a qualitative difference in the degrees of learners' intentionality.

CHAPTER 2: THE CURRENT STUDY

2.1 Overview of the research design

The current study is a between-subject design with three conditions: no test announcement with time restriction (Group 1: Timed Incidental), test announcement without time restriction (Group 2: Untimed Intentional), and test announcement with time restriction (Group 3: Timed Intentional). Group 1 performed the reading task without instructions to learn lexical items and then was given three vocabulary tests with no prior announcement. On the other hand, Group 2 and Group 3 were told in advance that they would be tested on vocabulary knowledge. However, Group 2 was told to complete the reading task at their own pace whereas Group 1 and 3 were asked to finish the reading in a limited time. The purpose of including the time restriction is to find out whether the longer time learners spend results in the beneficial effect of intentional learning. If intentional and incidental learning differs not in the cognitive processes but in the amount time allotted for the task, Group 3 and Group 1 would perform in a similar manner. While the independent variables are the presence or absence of a test announcement and time restrictions, the dependent variables include test scores and participants' eye-fixations on target lexical items. See Table 1 for an overview of the three groups.

Table 1

Group descriptions

	Test announcement	Time restriction
Group 1 (Timed Incidental)	No	Yes
Group 2 (Untimed Intentional)	Yes	No
Group 3 (Timed Intentional)	Yes	Yes

2.2 Materials

2.2.1 Experimental text.

I adapted and modified a passage — "Smart Cars, Intelligent Highways" — from the textbook *World Class Readings 3 Student Book: A Reading Skills Text* (Rogers, 2005). I used two tools to examine whether the reading would be appropriate for the participants' proficiency level: vocabulary profiles and readability. A vocabulary profile indicates how large a vocabulary is needed to read a text and which words readers are unlikely to know, while a readability index reveals how complex a passage is to read based on sentence length and other factors.

First, I used Compleat Web Vocabulary Profiler, an online research tool developed by Tom Cobb at the University of Quebec at Montreal (http://www.lextutor.ca/vp/eng/). This tool brakes down a text into 25 frequency bands and provides the percentage of lexical coverage in each band, based on the Corpus of Contemporary American English (COCA: Davies, 2008) and the British National Corpus (BNC) (Nation, 2005). For example, the K-1 band includes the most frequent 1000 words of English and K-2 includes the second most frequent 1000 words of English (i.e., 1001- 2000). Before entering the text into the tool, I re-categorized proper nouns and target words to high-frequency words for an accurate analysis. Overall, K-1 to K-2 words composed approximately 90% of the text and that figure rose to approximately 94% when K-3

Table 2
Vocabulary Profiles for Experimental Text

	K-1	K-2	K-3	K-4	K-5	Others	Total
Number of tokens	890	137	59	17	16	29	1151
Cumulative tokens (%)	77.32	89.22	94.35	95.83	97.22	100	

words were included. In other words, if a reader knew the first 3000 words of English, he or she would be able to understand 94% of this passage. Previous research has reported that learners need to know at least 90 to 95% of the words in a text (Hu & Nation, 2000; Laufer, 1997; Stahl, 1999) to be able to infer and generate the meaning of unknown words. As all the participants achieved at least 87.24% accuracy on the 3K level of the Levels Test, it is clear that the lexical demand of the reading text was suitable for them. Table 3 reports the results of these analyses.

Table 3

New Vocabulary Levels Test Results

	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6
Mean % (SD)	96.06 (0.75)	90.08 (1.40)	87.24 (1.81)	72.57 (2.19)	59.62 (3.32)	72.32 (5.53)

Second, to estimate the degree of difficulty in reading the text, I adopted three readability measures: Flesch-Kincaid Grade Level, the SMOG Index, and the Flesch Reading Ease. These traditional readability formulas are based on word and sentence lengths and approximate the age or number of years of education needed to understand the text. According to the Flesch- Kincaid Grade level and the SMOG Index, approximately nine to ten years of education in the United States is required to be able to read and comprehend the experimental text without difficulty. Next, in the Flesch Reading Ease test, possible scores range from 100 (indicating the easiest) to 0 (indicating the most difficult). The experimental reading passage scored 54.7 on this test, denoting that it can be considered comprehensible for 10th to 12th grade students. Thus, it seems suitable for the high-intermediate learners of English.

Table 4

Readability Assessment of Experimental Text

Flesch-Kincaid Grade Level (0–12)	**9.9**
The SMOG Index	**9.3**
Flesch Reading Ease (0–100)	**54.7**

2.2.2 Target words.

To ensure lack of previous knowledge of the target items, 12 words in the text were replaced with low-frequency words. These words were composed of three nouns, five verbs, and four adjectives. Each word occurred once in the passage. All of the target words belonged to the

Table 5

Target Words

Part of Speech	Target words	Definition	Frequency Level
nouns	gizmo	device	K-13
	fatality	death	K-4
	calamity	accident	K-9
adjectives	perilous	dangerous	K-5
	incessant	constant	K-8
	bewildering	distracting	K-5
	staggering	overwhelming	K-4
verbs	apprise	inform	K-14
	decipher	figure out	K-7
	succumb	die	K-6
	chauffeur	drive	K-8
	sip	drink a little	K-4

range of K-4 to K-14 level based on Compleat Web Vocabulary Profiler (Cobb, 2013). Table 5 displays the list of the target words and meanings.

2.2.3 Language background questionnaire.

A paper-based background questionnaire was prepared for the participants. The questionnaire elicited basic information about participants' gender, age, year in college, major, English language use, and English/ second language learning experience. In addition, participants provided standardized test scores (e.g., TOEFL, IELTs, and/or TOEIC scores) and self-rated their speaking, writing, reading, and listening proficiency levels (see Appendix B).

2.2.4 Prescreening vocabulary test.

As a prescreening measure, I adopted the New Vocabulary Levels Test (NVLT), developed by McLean and Kramer (2015): www.lextutor.ca/tests/. This NVLT is a diagnostic tool for measuring learners' receptive knowledge of the most frequent 5,000 word families. The test is composed of five 24-item parts in a multiple-choice format, one part each for representing the 1000-, 2000-, 3000-, 4000-, and 5000- word level. Additionally, the sixth part includes thirty items to measure academic word knowledge.

The prescreening vocabulary test served two purposes: 1) to measure participants' vocabulary knowledge and 2) to control for participants' pre-existing knowledge of the target words. For the former, I used the first three parts of the NVLT. For the latter, I randomly added the 12 target words to Part 6 of the NVLT, resulting in the 30 original items and 12 newly added items. The total score on the first five parts of the New Vocabulary Levels Test (NVLT) was used to ensure participants in each group had similar amounts of vocabulary knowledge. In sum, the final version of the prescreening vocabulary test comprised 174 items in total, 24 items for the first five parts and 54 items for the last part.

2.2.5 Reading proficiency test.

A reading proficiency test was administered to control for a possible effect of reading ability on participants' learning in the main experiment. I used the 2013 Sample Test Materials of the Examination for the Certificate of Competency in English (ECCE) developed by Cambridge Michigan Language Assessments (CAMLA). A standardized test for high-intermediate English language learners, the ECCE is divided into four sections: speaking, listening, GVR (grammar/vocabulary/reading), and writing. For the purpose of the current study and because of time limitations, participants took Part 1 of the reading section only. It included two reading passages followed by 5 multiple-choice comprehension check questions each. One point was assigned for correct answers, and zero points were given for incorrect answers. The cut-off score for inclusion in the study was set at six points. The reliability coefficient (Cronbach's α) obtained for all participants was .73 for the 10 items. This was considered to indicate acceptable test reliability (Field, 2013).

2.2.6 Comprehension test.

A set of paper-based comprehension questions was developed to ensure that participant actually read the text and to measure participants' understanding of the text. The test included 10 statements with three possible answers: true, false, and I don't know. I piloted the initial version on ten native speakers and five advanced non-native speakers of English. Based on their performance and opinions, some items were revised. One point was assigned for correct answers, and zero points were given for incorrect and "I don't know" answers. See Appendix E for a copy of the comprehension tests.

2.2.7 Vocabulary tests.

In this study, learning was operationalized as the ability to recognize and produce the form and meaning of target words. To measure the different aspects of participants' vocabulary learning, three vocabulary tests were designed and administered in the following order: form recognition, meaning recall, meaning recognition (see Appendices X, Y, and Z for each test). All three tests assessed participants' knowledge of the 12 target words, so the maximum score of each test is 12. Items on each test were presented in a random order.

Form recognition test. A form recognition test assessed participants' ability to recognize the target words. The test contained 12 items, each with five options: one target word, three distractors, and "I don't know." The distractors were selected randomly from low-frequency words. Participants were asked to circle one word they remembered seeing in the reading for each item and were encouraged not to guess. Participants earned one point for correct answers and zero points for false and "I don't know" answers. The reliability coefficient of the test was good, $\alpha = .79$.

[Example] Circle words you saw in the reading. If you do not know the answer, do not guess. There is a penalty for wrong answers.

(a) veer (b) distend (c) cardinal (d) sip (e) I don't know.

Meaning recall test. A translation-type meaning recall tests measured the learners' ability to recall the meanings of the 12 target words. Participants were asked to write down anything they remembered about the meaning of each word. When asked, I allowed them to write in their native language. Participants earned one point for the correct meanings, close synonyms, and related words and zero points for irrelevant answers or blanks. Half marks were not allocated. Spelling and grammar mistakes in responses were ignored because the test was intended to measure knowledge of meanings of words. The reliability coefficient of the test was good, α = .72.

[Example] For each word, write down anything you can remember about its meaning.

sip ________________________

Meaning recognition test. A meaning recognition test was administered to examine participants' ability to identify the meanings of target items. The test is regarded as easier than the meaning recall test because it is receptive in nature. Participants were instructed to circle one of five possible options: one correct answer, three possible but incorrect choices, and "I don't know." The three distractors were selected to match the target words in terms of the part of speech. For example, if the target word was a concrete noun, all three distractors were concrete nouns as well. Distractors were also chosen to not be close in meaning to the correct answer so that partial knowledge could be demonstrated. I awarded one point to correct answers and zero points if the target words were not circled. The reliability coefficient of the test was high, α = .81.

[Example] Circle the correct meaning of each given word. If you do not know the meaning, please circle "I don't know".

sip (1) brew (2) order (3) serve (4) drink (5) I don't know

2.3 Participants

Participants in the study were non-native speakers of English and came from two sources within a large Midwestern university: 47 were high-intermediate learners of English enrolled in Level 4 or 5 classes of a five level pre-university English program, where they took IEP or EAP classes, separately. Thirty-three participants were regularly matriculate students. From an original pool of 80 potential participants, 36 participants were excluded due to the following reasons: (1) having recognized three or more target words during the prescreening vocabulary test (n = 2), (2) failing to achieve an overall accuracy of 60% or more in the reading proficiency test (n = 2), (3) failing to attend the second session of the experiment (n = 1), (4) having technical errors during the experiment, including the unsuccessful calibration of the eye-tracker and the unexpected shutdown of the eye-tracking program (n = 5), (5) having produced inaccurate eye-tracking data (n = 26).

The final sample of 44 participants (28 females and 16 males) were mostly freshmen (n = 39) with a few sophomores (n = 4) and a single graduate student (n = 1), pursuing 32 different academic specializations, e.g. music, advertisement, business administration, computer science, and engineering. A majority of the participants were from China (n = 34), two came from Saudi Arabia, and 8 from other parts of the world, which reflects the population of the intensive English program at the university where the data was collected. Their age ranged from 17 to 30 (M = 19.43, SD = 2.43). Based on their self-rated English proficiency level on a 5-point scale, they were more confident in their reading (M = 3.62, SD = 0.78) and listening skills (M = 3.38, SD = 0.76) than they were in speaking (M = 2.67, SD = 0.57) and writing (M = 3.16, SD = 2.67).

2.4 Procedure

2.4.1 Apparatus.

The eye-tracking reading task was programmed using Experiment Builder and performed through EyeLink 1000, a desk-mounted eyetracker (SR Research Ltd. http://www.sr-research.com/). The eyetracker sampled gaze data 1,000 times per second. However, fixations below 120 milliseconds were eliminated from analysis because fixations below 120 milliseconds are considered not to reflect cognitive processing of words (Ashby, Rayner, & Clifton, 2005; Reichle, Rayner, & Pollatsek, 2003). The full experiment consisted of 5 screens for practice texts, 17 screens for the main text, and 8 screens for instructions and a break page. For the text, each screen contained an average of 66.88 words (SD=10.31) and included between five and ten lines of text. The entire text was presented in regular Consolas font, size 18, double-spaced. Position of target words on screens was controlled, so that no target words appeared in the first or last line, and on the left or right side of the screen.

Participants were seated in front of a computer monitor with their head placed against a chin and forehead rest to ensure the highest levels of accuracy and spatial resolution. The spacebar on the keyboard was used for participants to move from one screen to the next. Participants were not allowed go back to the previous screens while reading. While the eye tracker was calibrated at the beginning of the experiment and after the return from breaks, drift correction was set to be performed at the beginning of each screen.

2.4.2 Pretests.

The first session took place a week prior to the main reading experiment. In this session, two to three participants came to the office together. After signing the consent form and completing the background questionnaire, they took the New Vocabulary Levels Test (NVLT) and the reading proficiency test as prescreening measures. Participants who met the three conditions were qualified to continue participating in the next session of the study following: (a) achieving 90% or higher accuracy in Part 1 and Part 2 of the New Vocabulary Levels Test (NVLT), (b) having recognized two or fewer target words in Part 6 on the Vocabulary Levels Test (NVLT), and (c) achieving an overall accuracy of 60% or higher on the reading proficiency test. The first session took about an hour for the participants to complete. Test takers who did not meet the eligibility criteria were given $10 for their time.

2.4.3 Reading experiment.

In the second session, participants came to the eye-tracking laboratory individually and completed the main reading experiment and posttests. I started by giving directions on the eye-tracking experiment procedure and told all the participants they would answer 10 questions to check for their understanding of the story after reading. Depending on their randomly assigned group, participants received different instructions on time restrictions and vocabulary tests. Participants in Group 1 (Timed Incidental) were asked to read each slide in 30 seconds without any announcement of vocabulary tests at the end. Instead, they were asked to comprehend the reading in order to be able to answer comprehension check questions. On the other hand, participants in Group 2 (Untimed Intentional) were instructed to read at their own pace and encouraged to learn some unknown words because vocabulary tests would follow; however, instructions did not specify on which words they would be tested. Participants in Group 3 (Timed Intentional) were also forewarned of vocabulary tests after reading, but they were told to finish reading each slide in 30 seconds. I also told Group 1 and 3 that slides were programmed to transition to the next slide automatically after 30 seconds although, in reality, the slide was to be presented until participants hit the space bar to advance to the next page. I adopted the perceived time pressure rather than the real time limit to avoid losing data from slow readers. Considering that everyone has a different speed of reading, it would not be enough especially for the individuals with low reading speed to complete the reading task. Pilot test results indicated an average time of 28.83 seconds was spent on each slide, indicating that 30 seconds was sufficient time. Participants were calibrated with a standard nine-point grid for the right eye. Once the calibration was successful, they were instructed to fixate on a dot in the upper left corner of the monitor to start reading. If the eye tracker identified a fixation on the fixation spot, the reading text appeared. This procedure is called drift correction and it took place at the beginning of each screen. After reading five slides for practice, they read 17 slides for the reading passage on the screen while their right eye was being tracked. The break page was inserted after seven slides of the main text. Because vocabulary research indicates participants generally need multiple exposures to words to learn them, I had the participants re-read the main text. Again, they had a short break after the seventh slide. The reading experiment took a maximum of 30 minutes.

2.4.4 Posttests.

Immediately after participants completed the eye-tracking reading task, they took the comprehension tests first and the three vocabulary tests afterward. To minimize the transfer effect from preceding vocabulary tests, the participants took the vocabulary tests in the following order: the form recognition test, the meaning recall test, and the meaning recognition test. It took an average of 10 to 15 minutes for participants to finish the posttests. Participants received $25 for attending both session 1 and 2. Table 6 illustrates the procedure of the study.

Table 6

Procedure of the Study

Session 1	Session 2
1. Consent form 2. Background questionnaire 3. New Vocabulary Levels Test (NVLT) 4. Reading proficiency test	1. Reading 2. Re-reading 3. Comprehension test 4. Three vocabulary tests

2.4.5 Word predictability ratings.

Fifty-one native English speakers who did not participate in the main experiment performed a cloze predictability task to assess the degree of difficulty in guessing the meanings of the targets from context. They were provided with the reading passage 'Smart Cars, Intelligent Highways' with the target words deleted. On a separate sheet of paper, the raters were then asked to supply as many words as possible to fill in each blank and rate each case on a 5-point scale ranging from very easy to guess (1) to very difficult to guess (5). All raters were undergraduate students at Michigan State University.

I calculated the percentage of correct answers to each item and used the percentage as a continuous variable in the analysis. If one of the supplied answers was correct, the answer was counted as correct. As an example of the target word 'incessant', 'never-ending' and 'constant' were graded as correct, but 'busy' and 'terrible' were graded as incorrect. Semantically, syntactically, and contextually appropriate words were regarded as correct answers and spelling mistakes were ignored. The mean rating for each target word was also calculated, but I excluded the rating variable from the analysis. Therefore, to avoid any issues of collinearity in the model, I excluded the rating variable from the analysis. The predictability task took about 30 to 40 minutes to complete and raters received $10 for their participation.

2.5 Analysis

2.5.1 Definition of variables.

Term (Acronym)	Definition
Test Announcement (TA)	Whether the participant received an announcement of vocabulary posttests prior to reading
Time Limit (TL)	Whether the participants were told that a time limit was set for reading
Word Length (WL)	The number of letters in a word
Predictability (PD)	The correct answers expressed as a proportion of the total number of responses for an item on the cloze predictability task
Part of Speech (PoS)	Whether the word is a verb or not. 1 for verb, and 0 for non-verbs (i.e., nouns and adjectives)
Total Reading Time (TRT)	Summation of the duration across all fixations on the target word and across the two reading sessions
Summed Total Reading Time (STRT)	Aggregated Total Reading Time (TRT) by participants
ΔOE	The difference between observed total reading time and expected eye-fixation durations on the target word
Fixation count	The number of overall fixations
Form Recognition (FoReco)	Whether the subject correctly recognized the form of words
Meaning Recognition (MeReco)	Whether the subject correctly recognized the meaning of words
Meaning Recall (MeReca)	Whether the subject correctly recalled the meaning of words

2.5.2 Eye-tracking data preparation.

The purpose of cleaning data was to scan for unusual events, and deal with those events in an appropriate manner. I first filtered out fixations shorter than 120 milliseconds as these fixations are less likely to be associated with readers' cognitive processes (Ashby, Rayner, & Clifton, 2005; Reichle, Rayner, & Pollatsek, 2003). It is also common practice to exclude fixations longer than 800 milliseconds. However, considering that participants were English language learners and some of them were aware of the following vocabulary tests after reading, I did not remove the long fixations as I acknowledge that longer fixations could have been made intentionally. I also manually reviewed and inspected each trial fixation by fixation, looking for inconsistences in the data. For example, when fixations were off the line of text, I moved the fixations either up or down depending on which line a fixation was intended. A total of 80 data files were collected and 61 data files were used in the main analysis with 19 data files excluded for various reasons (see section 2.3).

There are many eye-movement measures including first fixation duration, first-pass reading time, regression path duration, and total reading time. In the current study, I used two eye-tracking measurements: total reading time (TRT) and the difference between observed and expected fixation duration (ΔOE). First, total reading time, the sum of all fixations on the target word, indicates how much total time a reader spent at the region during the entire course of reading. Total reading time is considered a late measure that reflects late cognitive processes such as text comprehension and information reanalysis (Roberts & Siyanova-Chanturia, 2013). Considering the predictor variable of the current study, Test Announcement, would affect on primarily late eye-movement measure, and the aim of the study is to uncover the associations with vocabulary learning, I concluded that total reading time is suitable to represent the amount of attention paid to the targets in the current study. Second, the difference between observed and expected fixation duration (ΔOE) was calculated following Indrarathne and Kormos (2016, 2017). The procedure of getting the ΔOE value is as follows:

1) Extract the total reading time for the whole page for each participant by summing up all fixation durations on all words within the page
2) Calculate the expected fixation durations based on the proportion of the number of syllables that the target word has in relation to the number of syllables on the whole page where the particular target word occurs.

   Expected fixation duration of a target word for a participant =

$$\frac{no.\,of\ syllables\ of\ target\ word\ \times\ total\ reading\ time\ of\ the\ whole\ page}{no.\,of\ syllables\ of\ the\ whole\ page}$$

3) Subtract the expected fixation duration from the observed total reading time for each target word for each participant

The difference between the observed and expected fixation durations (ΔOE) is regarded as instances of noticing because it measures "extra attentional processing load" (Indrarathne & Kormos, 2016, p.6) of target words.

2.5.3. Data structure.

The data have a hierarchically clustered structure whereby target words were nested within subjects and each subject provided multiple observations. That is, repeated observations were made on the same individual. Specifically, 61 participants reported eye-fixation data for 12 target words each and provided three types of vocabulary test results for each word. An important feature of the data is that predictors reside at different levels of the data structure. Word length, Predictability, and Part of Speech are measured at Level 1 because they are the characteristics of target words while Test Announcement and Time Limit are measured at Level 2 because they are the treatments given to subjects. Therefore, the number of data points at Level 1 is 732 (61 participants × 12 target words) and the sample size at Level 2 is 61. Figure 1 shows an example of the data structure of the current study.

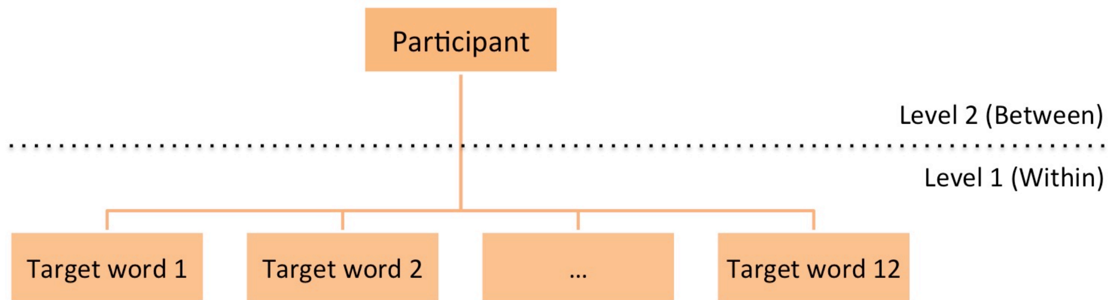

Figure 1 Data structure.

2.5.4 Multivariate Multilevel Mediation Analysis (MMMA).

2.5.4.1 Path analysis.

For the current study, I adopted a path analysis, which represents a special case of structural equation modeling (SEM) (Marcoulides & Schumacker, 1996). Both path analysis and SEM are extensions of multiple regression to estimate the relationships among the variables that can accommodate nested data with predictors at different levels and multiple outcomes. In addition, both analyses are useful ways to examine how the effect of an independent variable (X) on an outcome (Y) is mediated through an intervening variable (M). However, path analysis describes the relationships between *observed* or directly measured variables, while SEM deals with *latent* or unobserved constructs. An observed variable is measurable, such as a test score and time on task, whereas a latent variable is a variable that cannot be measured directly, for instance, motivation or attitude. Thus, latent variables are inferred indirectly from the variances and covariances in a set of observed variables. Although a general case of a mediation analysis with multiple outcomes and/or multiple mediators is commonly undertaken within an SEM framework, the statistical approach used in the current study is more closely related to path analysis for the reason that it does not involve a latent variable measure model.

2.5.4.2. Mediation analysis.

As diagrammed in Figure 2, mediation occurs through an added variable that affects the causal relationship of X to Y, describing the mechanism of how the predictor (X) causes the mediator (M), and the mediator (M) causes the outcome (Y). Rectangles indicate observed variables, and each straight line represents a causal relation with an arrowhead at one end, pointing from predictor to outcome. Mediation analysis distinguishes three types of effects: direct, indirect, and total effect. The direct effect refers to the influence the predictor variable has directly on the outcome variable, the indirect effect refers to the pathway from the predictor to the outcome through the mediator, and total effects denotes the aggregated effect of the direct and indirect effects on the outcome. In Figure 2, the paths a and b represent the indirect effect of X on Y, the path c' represents the direct effect of X on Y, and c = ab+ c' is the total effect of X on Y. The diagram on the top represents the total effect of X on Y and the diagram on the bottom shows the mediated effect of X on Y through M.

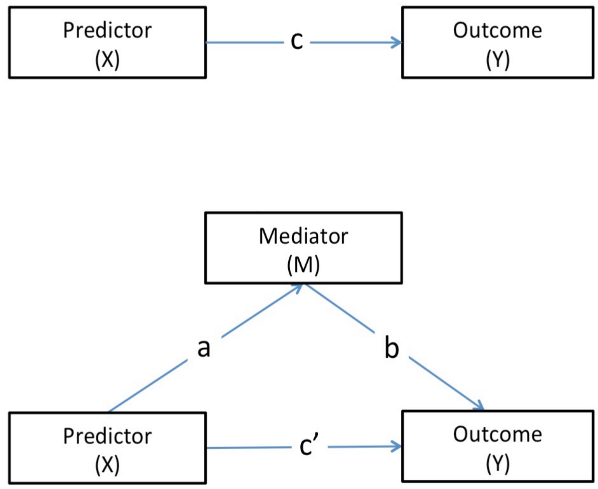

Figure 2 Path diagram for a basic single-mediator model.

The regression equations of these two diagrams are as follows:

$$Y = i_1 + cX + e_1$$

$$M = i_2 + aX + e_2$$

$$Y = i_3 + c'X + bM + e_3$$

Coefficient *c* denotes the total effect of X on Y, coefficient *a* represents the effect of X on M, coefficient *b* quantifies the effect of M on Y adjusted for X, and coefficient *c'* is the direct effect of X on Y that is not transmitted through M. $i_1, i_2,$ and i_3 are the intercepts and $e_1, e_{2,}$ and e_3 are the residuals.

Mediation can either be partial or complete. Figure 3 shows the diagrams for a partial and a full mediation model. Partial mediation is the case in which an independent variable has both direct and indirect effects on a dependent variable. If Test Announcement has a direct significant impact on the test scores and it also has a significant impact on Total Reading Time, which has a significant impact on the test scores, this is known as a case of partial mediation. Complete or full mediation is the case in which the total effect of an independent variable on a dependent variable is transmitted through mediators. If Test Announcement does not have a direct impact on test scores, but it has a significant effect on Total Reading Time, which also has a significant impact on test scores, this is known as a case of full mediation. In the case of a full mediation, the mediator fully explains the association between the predictor and the outcome.

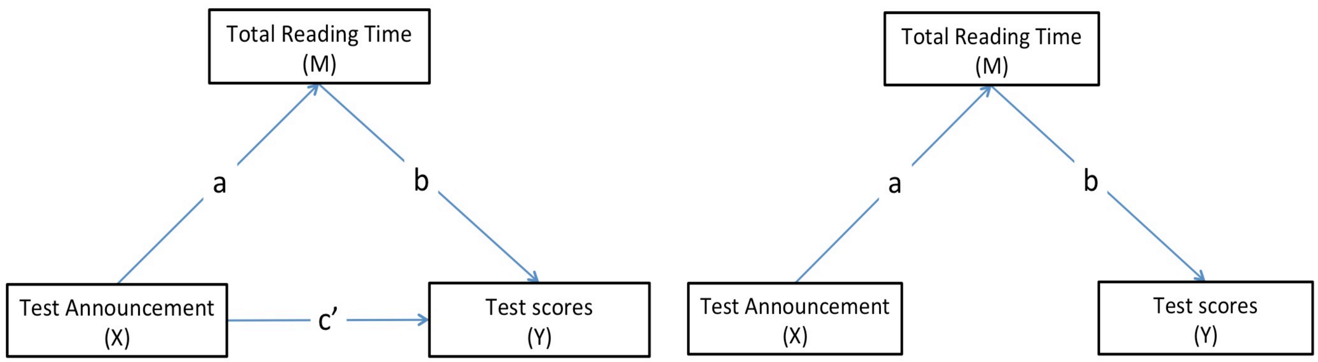

Figure 3 Path diagrams for a partial and full mediation model.

A meditation analysis can also be done through ordinary least squares (OLS) and logistic regression. However, the benefits of path analysis using the SEM framework for testing questions of mediation, Bryan, Schmiege, and Broaddus (2007, p.366) summarized and compared to the OLS and logistic regression approach.

(1) testing of direct, indirect, and total effects simultaneously

(2) testing of complicated mediation models with multiple mediators and/or dependent variables

(3) testing of particular indirect effects within the mediation models

(4) the ease of correcting for missing data and non-normality in data

Considering that the current study includes five independent variables, three dependent outcomes, multilevel data with some missing values, and a non-normal distribution for several variables, the use of path analysis via SEM framework was implemented to answer the questions related to mediation.

2.5.4.2.1. Multilevel Mediation Analysis.

In multilevel modeling, different types of models exist depending on the data structure. Regarding the types of models, when all variables are measured at Level 1, the model is referred to as a 1-1-1 mediation model (Krull & MacKinnon, 1999, 2001). Models in which the predictor and the mediator are assessed at Level 2 and outcome variables are assessed at Level 1 are called 2-2-1 mediation models. For example, one could hypothesize that the relationship between classmates' language skill (a Level 2 predictor) and individuals' language skill (a Level 1 outcome) is mediated by the effect of classroom quality (Level 2 mediator). In the 2-1-1 mediation model, a Level 2 predictor influences a Level-1 mediator, which then affects a Level-1 outcome. An example of this mediation would be that the instructional practices (Level 2 predictor) impact on individual's motivation (Level 1 mediator), which in turn affects learning outcomes (Level 1 outcome).

In the current study, total reading time denotes a mediator, learning gains from three vocabulary tests serve as outcome variables, and Test Announcement, Time Limit, Word Length, Predictability, and Part of Speech are predictors. That is, it is hypothesized that Announcement, Time Limit, and three other lexical factors (X) predict processing time (M), which in turn affects learning of unknown words (Y). One of the complications of the study is that predictors lie at different levels as described in the previous part. More specifically, Test Announcement and Time limit are measured at Level 2 while Word Length, Predictability, and Part of Speech are measured at Level 1. Therefore, I merged the 2-1-1 and 1-1-1 mediation models to accommodate all predictors at two levels in the present study. In addition, the existence of the 1-1 linkage in the model invites special attention because the between-subjects effect and the within-subjects effect of the mediator (Total Reading Time and ΔOE) on the outcome variable needs to be examined separately, according to Preacher, Zyphur, and Zhang (2010). In the current study, for instance, Test Announcement is the Level-2 independent variable, with one group receiving the announcement and another not receiving the announcement. Total Reading Time and ΔOE are the Level-1 mediators and learning outcomes are the Level-1 dependent variables. In this case, Test Announcement only varies between groups, whereas both Total Reading Time (or ΔOE) and test scores vary both within and between groups. That is, each target word differs from each other within the person in its eye fixations and learning outcomes (within-subjects effect), and there are differences between the person in eye fixations and learning outcomes (between-subject effects). When one estimates the influence of Test Announcement on eye fixations, Test Announcement influences individual target words but does so for the person, making the effect of Test Announcement a between-subjects effect. Because Test Announcement was provided to the person without differential application across target words within the person, it cannot account for within-person differences of any kind. This is not to say that Test Announcement has no impact on Level-1 vocabulary learning. It does, but only because each target word belongs to participants that either did or did not receive the announcement of the upcoming posttests. For the same reason, Test Announcement can impact on vocabulary gains only at the level of person. Test Announcement cannot account for individual differences within a participant in only reading patterns and vocabulary learning, because Test Announcement was applied equally to each participant. Therefore, the indirect effect of Test Announcement on vocabulary learning through Total Reading Time may function only through the between-group variance in the mediator (Total Reading Time) and the dependent variables (learning gains). The idea is supported by Preacher, Zyphur, & Zhang (2010), stating that in a mediation model for 2-1-1 data, "when the *b* effect (the effect of the mediator to the outcome variable) estimate conflates the Within and Between effects, the indirect effect that necessarily operates between groups in confounded with the within-group portion of the conflated *b* effect (p.211)."

In order to disentangle between-subjects and within-subjects effect, I adopted a strategy called *unconflated multilevel modeling* (UMM) (Hedeker & Gibbons, 2006; MacKinnon, 2008; Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010). Following UMM, I replaced the Level 1 mediator with two mediators: a group-mean centered total reading time (i.e., deviations from group means) at the within-subjects level, and the group mean of total reading time at the between-subjects level. Here, group-mean centering subtracts the individual's group mean of total reading time from an individual's total reading time for each target word. I followed the same procedure for the ΔOE value. In this way, the within- and between-subjects effects of the model are no longer conflated, because they are not combined into a single estimate (Preacher, Zyphur, Zhang, 2010). However, the main drawback of this approach is that using the group mean as a proxy for the between-subjects effect introduces bias of the between-subjects effect for the predictor, which in turn also contributes to biased indirect effects at the between-level. To solve this problem, Preacher et al. (2010) suggest using a *multilevel structural equation modeling* (MSEM) approach in investigating multilevel mediation. This approach allows for separate estimation of the within- and between-subjects components of the model, so the direct and indirect effects at each level can be examined separately. Although MSEM is the most advanced approach for mediation in nested data, more future studies need to empirically prove that the MSEM method is superior in decreasing bias in indirect effects and estimating those effects in an absolute sense (Preacher et al., 2011).

I first attempted to run the multilevel mediation model using the MSEM approach, but the model did not converge correctly because between-subjects variances were too small to support the MSEM model. So, I chose the UMM to assess the multilevel phenomenon of the present study, considering that UMM also allows for decomposition of Level 1 and Level 2 effects, and it is known to be more valid than traditional multilevel statistical methods including multilevel modeling and multiple linear regression (Bauer, Preacher, & Gil, 2006).

2.5.4.2.2. Multilevel Mediation Analysis with Multiple Outcomes.

Following other recent work on incidental vocabulary learning, the study includes multiple dependent variables: Form Recognition, Meaning Recognition, and Meaning Recall. Instead of running the multilevel model three times for each outcome variable, data were analyzed using the multivariate multilevel model, the extended version of the multilevel model, to accommodate multiple outcomes (Baldwin, Imel, Braithwaite, & Atkins, 2014). Snijders and Bosker (2012) explained that the multivariate approach is more rigorous than the univariate approach, especially if a correlation between dependent variables exists. This approach decreases the probability of Type I error, which will otherwise be inflated when carrying out separate tests for multiple dependent variables.

2.5.5. Statistical analysis.

Multivariate multilevel mediation model was conducted using M*plus* 8 (Muthén & Muthén, 2017) to evaluate the possible relationships among clustered data with multiple dependent variables simultaneously. As the data set included 1.89% of missing values, I handled the missing data using the full-information maximum likelihood (FIML; Enders, 2010) estimator implemented in Mplus. Because the outcome variables are binary (either correct or incorrect), a logistic regression model was fitted by using the robust maximum likelihood (MLR) estimator with the LINK option. The MLR estimator has the benefit of accounting for the non-normality in the measures (Muthén & Muthén, 2017). To include random intercepts, random slopes, and random variances in the multilevel analyses, TWO LEVEL RANDOM option was selected for the type of analysis. However, I group-mean-centered the mediator (Total reading time and ΔOE), I specifically fixed the intercept within participants to zero to take it out of the model. Also, the residual variances of Meaning Recognition and Meaning Recall were fixed at 0 because they were close to 0 (.002 and .009, respectively).

M*plus* uses a binary logistic regression for all multilevel analyses with a categorical outcome variable, thus the estimates for paths from predictors to dependent variable are logit regression coefficients (*b*). For ease of interpretation, I also report the exponentiation of the B coefficient (*exp*(*b*)) for the final model (model 1d), which is an odds ratio. An odds ratio greater than 1 implies a positive relationship. Putting it differently, a positive coefficient indicates that the probability of the categorical dependent variables occuring (the probability of getting a correct answer on the vocabulary tests) increases when the predictor values increases. In contrast, when the odds ratio is smaller than 1, it implies a negative relationship. An odds ratio smaller than 1 indicates that when the predictor values decreases, the likelihood of a correct test answer increases. In order to compare the relative strength of the effect of each individual independent variable to the dependent variable, standardized coefficients (β) were additionally calibrated using Bayesian estimation because M*plus* does not provide standardized estimates for the multilevel analyses with the current MLR estimator. Standardized beta coefficients (β) are reported along with unstandardized coefficients (b) in the path diagrams.

CHAPTER 3: RESULTS

The results presented in this chapter are organized by research questions.

1. What is the effect of test announcement and time restrictions on the acquisition of receptive and productive knowledge of word form and meaning, as measured in vocabulary posttests?
2. What is the effect of test announcement and time restrictions on eye-fixation times on novel words?
3. What are the interrelationships between intentionality (test announcement), time pressure, attention (eye-fixation duration), and vocabulary learning (test scores)?

I first look at the comparability of the groups at the pretest before examining the vocabulary posttests and processing time by group. Then, I examine whether the effect of test announcement and time limit on vocabulary learning was mediated by eye-fixations. It was hypothesized that the intentional learning condition would produce longer processing times, and longer fixations on target words would enhance the initial stages of the acquisition process, such as recognizing the word form or inferring the meaning of the word; that is, I hypothesized total time as a mediation variable. For the statistical analysis, the alpha level was set at .05 ($\alpha = .05$).

3.1 Pretests

Table 7 displays the means and standard deviations derived from participants' performance on the reading proficiency test and the New Vocabulary Levels Test (NVLT). The average score from the reading proficiency test was greatest for the Timed Intentional group (Group 3) ($M = 8.60$, $SD = 1.00$), followed by the Timed Incidental group (Group 1) ($M = 8.15$, $SD = 1.23$), and lastly the Untimed Intentional group (Group 2) ($M = 7.90$, $SD = 1.18$). The average score on the first five parts of New Vocabulary Levels Test (NVLT) was 81.85 out of 120 (*SD* = 13.29) for the Timed Incidental group (Group 1), 81.80 (*SD* = 10.28) for the Timed Intentional group (Group 3), and finally 80.24 (*SD* = 12.87) for the Untimed Intentional group (Group 2).

Table 7

Average Scores on Two Pretests by Group

	Reading Proficiency Test Mean (SD)	New Vocabulary Levels Test (NVLT) Mean (SD)
Group 1 (Timed Incidental, *n* = 16)	8.31 (1.20)	83.00 (12.55)
Group 2 (Untimed Intentional *n* = 14)	7.64 (1.15)	80.50 (13.07)
Group 3 (Timed Intentional *n* = 14)	8.64 (1.15)	82.29 (11.70)

Note. Summed scores from first five parts of the NVLT were reported and analyzed. The maximum score is 12 for the reading proficiency test and 100 for the NVLT.

To ensure that three groups were comparable in terms of their English proficiency level, a series of one-way analyses of variance (ANOVA) were conducted on the reading proficiency test and the New Vocabulary Levels Test (NVLT). Participants with prior knowledge of three or more lexical items were excluded in the final analysis, so I did not perform a statistical test on the prior knowledge of the targeted lexical item across groups. First, a one-way ANOVA was conducted with Group as the independent variable and Reading test scores as the dependent variable. The results showed that there were no significant differences between groups, F (2, 43) = 2.676, p = .081, indicating participants started out at similar reading proficiency level. Second, a one-way analysis of variance (ANOVA) was run with Group as the independent variable and total scores from the five parts on the New Vocabulary Levels Test (NVLT) as the dependent variable. Results confirmed that the three groups did not differ statistically with respect to their vocabulary level (F (2, 43) = .157, p = .856. Taken together, these two tests established the comparability of the three groups in the study.

3.2 Time on the reading task by group

To measure the accurate time on reading, I calculated the average of all fixations on the text for each session. Table 8 indicates that the Untimed Intentional group (Group 2) yielded the longest the average reading time (M = 515.00, SD = 79.38) followed by the Timed Intentional group (Group 3) (M = 394.06, SD = 79.88), and lastly the Timed Intentional group (Group 1) (M = 313.60, SD = 94.66).

Table 8

Average Time on Reading Task by Group

	Mean	SD
Group 1 (Timed Incidental, n = 16)	313.60	94.66
Group 2 (Untimed Intentional, n = 14)	515.00	79.38
Group 3 (Timed Intentional, n = 14)	394.06	79.88

Note. Times are given in seconds.

An initial inspection of the eye-tracking data revealed that the assumptions of using a parametric test including normality, homogeneity of variance, and the independence of observations were met (Field, 2009). Therefore, one-way ANOVA was conducted to compare the time taken to complete reading among the groups. The comparison of the three showed a statistically significant difference between groups (F (2, 43) = 20.867, p < .001). Tukey's post hoc tests revealed that the Timed Incidental group elicited statistically significantly shorter time to complete reading than the Untimed Intentional group (mean difference = 201.40, 95% CI [123.36 279.45], $p < .001$). Also, the Timed Incidental group and the Timed Intentional group showed a statistically significant difference (mean difference = 80.46, 95% CI [2.42, 158.51], $p = .041$). Likewise, there was statistically significant difference between the Untimed and Timed Intentional groups (mean difference = 120.94, 95% CI [40.33, 201.54], $p = .002$).

3.3 Vocabulary test results by group

To compare the groups with respect to their test scores, I used the summed scores over the 12 target words by participants for the analysis. Descriptive statistics for the participants' performance on three vocabulary tests are presented in Table 9. Overall, regardless of which group they belonged to, participants earned highest scores on the form recognition test (46.75% accuracy on average) and lowest scores on the meaning recall test (10.08% accuracy on average). Comparing the groups, the untimed intentional group (Group 2) recorded the highest gains in all three tests, followed by the timed intentional group (Group 3) and finally the timed incidental group (Group 1). Figure 4 illustrates the results on vocabulary tests by group.

Table 9

Average Scores on Three Vocabulary Post-test Measures by Group

	Form Recognition Mean (SD)	Meaning Recognition Mean (SD)	Meaning Recall Mean (SD)
Group 1 (Timed Incidental, $n = 16$)	5.25 (3.07)	2.94 (1.18)	.97 (1.26)
Group 2 (Untimed Intentional, $n = 14$)	6.29 (2.92)	4.29 (1.98)	1.48 (1.80)
Group 3 (Timed Intentional, $n = 14$)	5.29 (2.87)	3.50 (1.02)	1.18 (1.40)

Note. The maximum score of each test is 12.

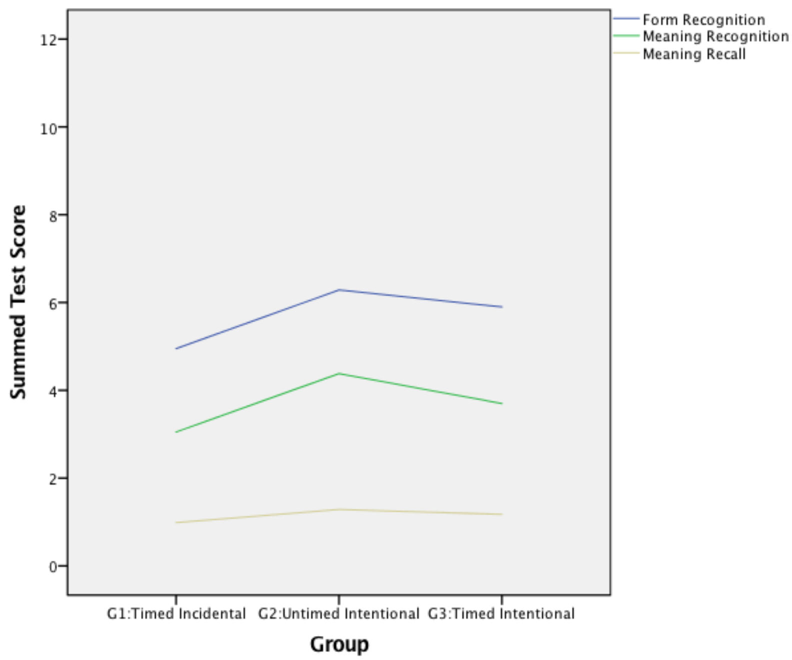

Figure 4 Performance on the three vocabulary post-test measures.

To explore potential group differences in vocabulary test performance, a 3 (Vocabulary test) x 3 (Group) mixed-design ANOVA was performed. Before conducting the statistical test, I initially inspected the learning data to check for statistical assumptions following Larson-Hall's guidelines (2011). Scores on the form recognition test were normally distributed and variances were largely equal, whereas the meaning recognition and meaning recall test scores were found to violate both the normality and equal variances assumptions across the groups. Therefore, a log transformation was performed on each set of test scores. The results showed that there was no significant interaction between Vocabulary Test and Group, F (4, 64) = 1.118, p = .356, η_p^2 = .065. The main effect of Group was not significant F (2, 32) = .189, p = .829, η_p^2 = .012, indicating that the learning gains of participants were not significantly different from each other across the different types of vocabulary tests. The main effect of Test was significant F (2, 64) =

66.281, $p < 0.001$, $\eta_p^2 = .674$, indicating test scores differed strongly and significantly between Vocabulary Tests. Post hoc tests using the Bonferroni correction revealed that Form Recognition was significantly different from Meaning Recognition ($p = .021$) and Meaning Recall ($p < .001$) and Meaning Recognition is significantly different from Meaning Recall ($p < .001$). These findings indicate that the participants performed in a parallel manner regardless of the groups they belong to.

Table 10
Mean Fixation Count, Mean Total Reading Time, and Mean ΔOE *for the First and Second Session*

Target words		1st reading session			2nd reading session		
		Fixation count	Mean TTR (SD)	Mean ΔOE (SD)	Fixation count	Mean TTR (SD)	Mean ΔOE (SD)
1	apprise	2.98 (2.51)	745 (703.92)	182 (367.91)	3.26 (3.46)	840 (1091.92)	302 (369.62)
2	bewildering	4.13 (2.77)	799 (654.51)	16 (413.92)	3.23 (2.80)	786 (735.04)	24 (599.42)
3	calamity	3.7 (2.71)	1003 (774.21)	-44 (462.98)	2.9 (2.30)	763 (624.98)	-17 (454.43)
4	chauffeur	3.8 (2.53)	880 (529.55)	352 (401.53)	2.85 (1.89)	654 (444.7)	262 (367.92)
5	decipher	3.26 (2.26)	772 (600.39)	36 (388.18)	2.54 (2.03)	609 (485.33)	-8 (396.17)
6	fatalities	3.54 (2.65)	868 (705.04)	-41 (503.57)	2.87 (2.61)	733 (682.56)	40 (534.83)
7	gizmos	2.77 (2.30)	635 (534.88)	-4 (358.28)	2.48 (1.76)	645 (536.12)	172 (384.10)
8	incessant	3.61 (2.36)	880 (653.68)	95 (474.89)	2.62 (2.17)	652 (696.68)	31 (373.86)

Table 10 (cont'd)

9	perilous	3.9 (3.00)	917 (730.64)	50 (402.37)	3.41 (2.09)	835 (584.99)	162 (400.11)
10	sipped	4.02 (2.36)	954 (612.32)	605 (441.00)	2.33 (1.51)	559 (420.00)	360 (324.41)
11	staggering	4.39 (2.49)	1019 (594.64)	174 (452.67)	3.97 (3.14)	946 (820.49)	278 (521.84)
12	succumb	3.36 (2.57)	802 (570.08)	295 (414.38)	2.38 (2.18)	569 (546.05)	190 (461.23)

Note. TRT = Total Reading Time in millisecond (the sum of all fixations on the target word)

3.4 Eye fixations by group

3.4.1 Comparison between Session 1 and Session 2.

Since participants read the same text twice for the purpose of comprehension, total reading time on targets from first and second reading are reported. Table 10 shows that in both reading sessions, the target word "staggering" elicited the longest mean total reading time whereas the target word "gizmo" in the first session and "sipped" in the session elicited the shortest mean total fixations. Expectedly, total reading time in the second session was shorter than those in the first session. As shown in Figure 5 and 6, reading-time patterns for each target word were generally similar between first and second reading sessions although these patterns appeared to be different across the three groups.

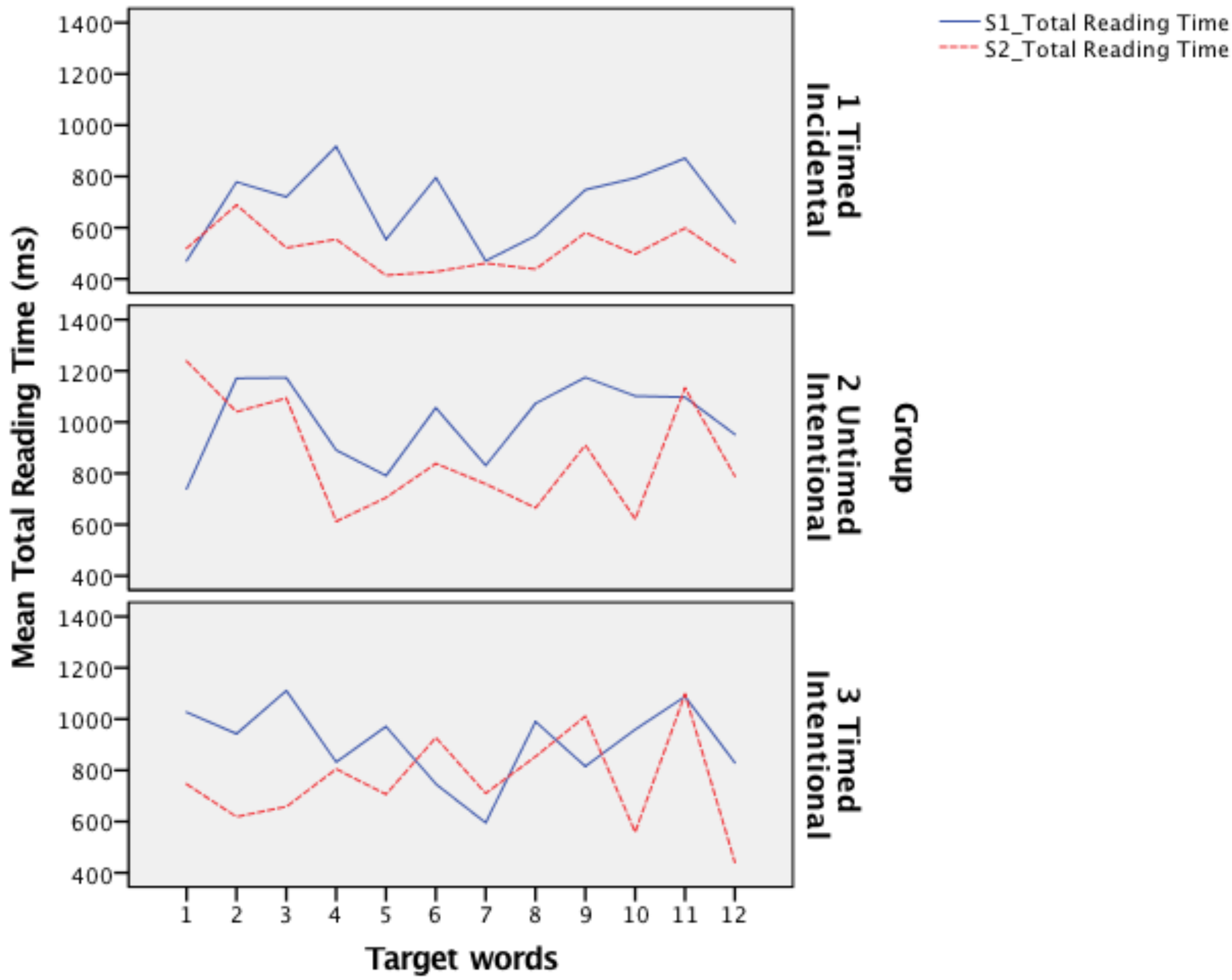

Figure 5 Mean total reading time by target words and groups.

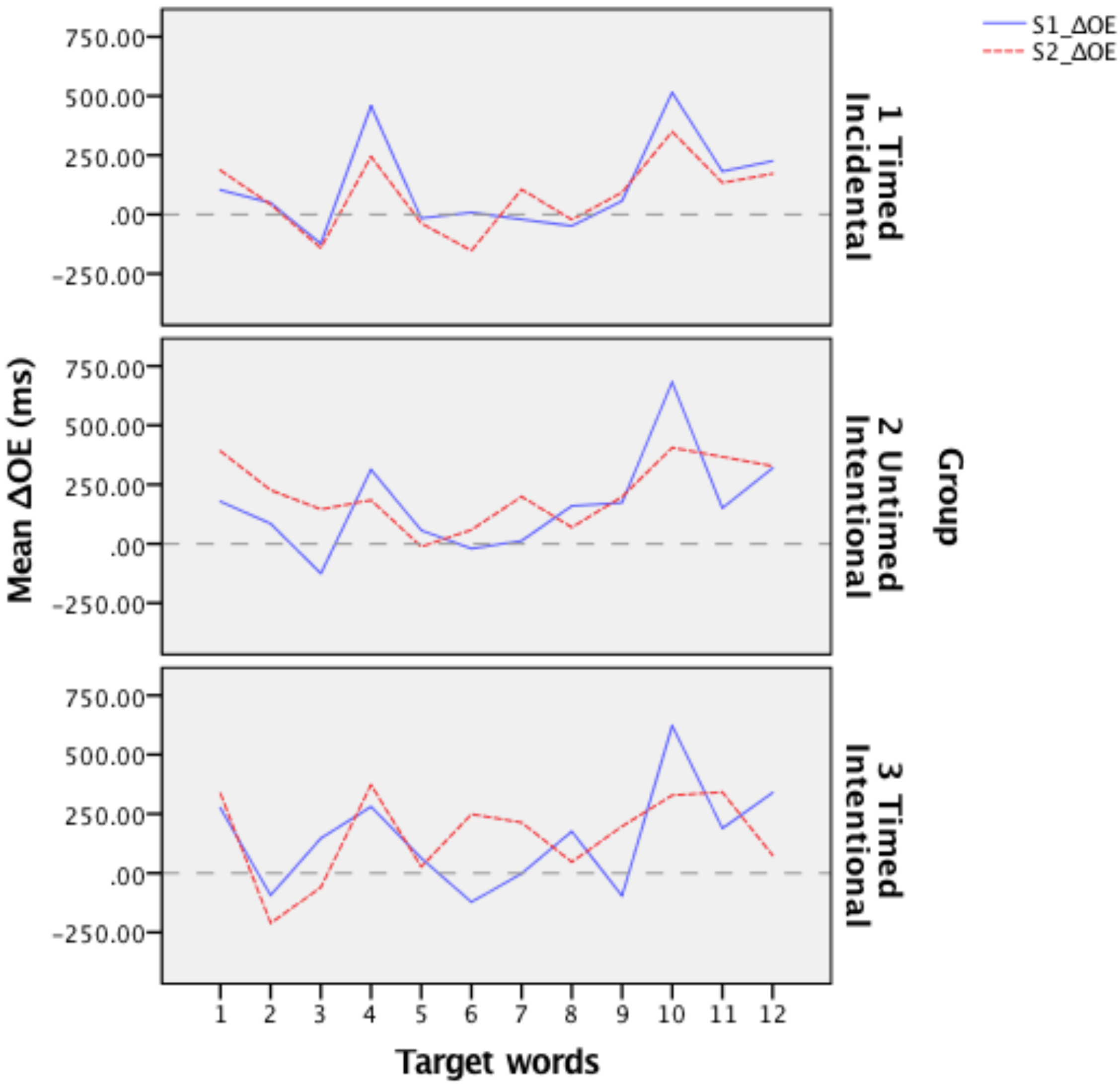

Figure 6. Mean ΔOE by target words and groups.

3.4.2. Summed Total Reading Time and Summed ΔOE

In the second research question, I asked whether the attention that participants paid to the target words differed by group. First, mean scores for both variables, summed total reading Time and the summed difference between the observed and expected TFD (ΔOE), were computed. The descriptive statistics presented in Table 11 show that Untimed Intentional group had the highest TRT (M = 1871, SD = 599.52) and ΔOE values (M = 220, SD = 544.77). This group is followed by the Timed Incidental group (M = 1669, SD = 609.2 for TRT; M = 220, SD = 544.77 for ΔOE) while the lowest values were recorded for the participants in the Timed Incidental group (M = 1206, SD = 405.03 for TRT; M = 157, SD = 465.75 for ΔOE).

Table 11

Average Summed Total Reading Time and Summed ΔOE *by Group*

	Mean Summed Total Reading Time (SD)	Mean Summed ΔOE (SD)
Group 1 (Timed Incidental, n = 16)	1162 (634.40)	138 (493.03)
Group 2 (Untimed Intentional, n = 14)	1871 (599.52)	369 (801.27)
Group 3 (Timed Intentional, n = 14)	1352 (658.06)	153 (601.61)

Note. Times are given in milliseconds.

As both eye-tracking measures, summed total reading Time and the summed difference between the observed and expected TFD (ΔOE), were positively skewed, a log transformation was performed (Larson-Hall, 2010) and 25 outliers were excluded for normality of the data. The transformed data satisfied for parametric analysis.

The one-way ANOVA results revealed that there was a statistically significant difference on summed total reading time between groups as determined by (F (2, 707) = 25.804, $p < .001$, η_p^2 = .06). Comparisons using Tukey's contrast revealed a statistical difference between the Timed Incidental group and the Untimed Intentional group (mean difference = .148, 95% CI [.09, .20], $p < .001$, d = .42) and between the Timed Incidental group and the Timed Intentional group (mean difference = .106, 95% CI [.05, .26], $p < .001$, d = .60). However, there was no statistically significant difference between the Untimed Intentional group and the Timed

Intentional group (mean difference = .042, 95% CI [-.01, .10], p = .144, d = .17). In summary, these results suggest that participants in the intentional-learning mode spent a comparable amount of time on target words regardless of time restrictions. Participants in the incidental-learning mode spent significantly less time on target words than ones in the intentional-learning mode. The same analyses were conducted on the ΔOE data, the results showed that there was no significant effect on ΔOE, F (2, 683) = 1.557, p =.211, η_p^2 = .04).

3.5 Multivariate Multilevel Mediation Model Results

As the primary purpose of analyzing the ΔOE data is to compare the effect of Test Announcement on total reading time and ΔOE value. Therefore, I first tested alternative models and found the best-fitting model to describe the relation among predictors, vocabulary knowledge, and total reading time. Next, using the same model, I estimated the relation among predictors, vocabulary knowledge, and ΔOE value to investigate the differential effects on variables.

3.5.1 Model comparisons.

To identify the most parsimonious and well-fitting model, I removed one path at a time and evaluated changes in fit across models. This approach is known as a theoretically driven model testing, and is consistent with common practice (Kline, 2016). For this, I began by testing the full model in Figure 7 and continued testing theoretically plausible alternatives. For the sake of clarity and succinctness, I report the four representative models including the first full model and the final model in this paper. Table 12 presents fit statistics of all models that were considered to determine the most appropriate model. The models presented from Figure 7 to Figure 10 illustrate the sequential process implemented to examine the direct and indirect contributions of intention and time limit to vocabulary learning via total reading time. For convenience and presentation clarity, the path diagrams are presented separately by dependent variable despite the fact that all the estimates were examined simultaneously within a single multivariate model. That is, Model 1a, Model 1b, and Model 1c are *not* three separate univariate analyses, but three parts of one multivariate analysis. The final path model (Model 4) is presented once again with all the dependent variables included (see Figure 11). From Figure 7 to 11, standardized coefficients are presented on the left and unstandardized coefficients are presented on the right. Solid and dashed lines represent significant and nonsignificant effects, respectively.

A series of alternative models were tested to find the best-fitting model to describe the relationship among pedagogical interventions, attention, and learning outcomes. The first is a partial mediation model where the relationship of all five independent variables (i.e., Test announcement, Time limit, Word length, Predictability, and Part of speech) to vocabulary learning is partially mediated by total reading time (Model 1). In the second model, total reading time completely mediates the relation between the independent variables on Level 1 (i.e., Word length, Predictability, Part of speech) and vocabulary learning whereas relations of the other variables on Level 2 (i.e., Test announcement, Time limit) and vocabulary learning are partially mediated by total reading time (Model 2). In the third model, total reading time partially mediates the relationship between the independent variables on Level 1 (i.e., Word length, Predictability, Part of speech) and vocabulary learning, whereas relationships of the other variables on Level 2 (i.e., Test announcement, Time limit) and vocabulary learning are completely mediated by total reading time (Model 3). The fourth model is a fully mediated model in which only total reading time is hypothesized to have a direct relationship with vocabulary learning, completely mediating the relationship of each independent variable (i.e., Test announcement, Time limit, Word length, Predictability, Part of speech) to vocabulary learning (Model 4).

Alternative models were evaluated using the relative fit statistics of Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted Bayesian Information Criteria (aBIC) (see Table 12). As M*plus* does not produce the degrees of freedom of multilevel models in which dependent variables are categorical, the other fit indices (such as Chi-square statistics, comparative fit index, and the Tucker-Lewis index) cannot be considered. For AIC, BIC, and aBIC, models with lower values are preferred. According to Raftery's (1995) guidelines, a BIC difference of over 10 implies "very strong" evidence in favor of the model with the smaller BIC; a difference of 6 – 10 is "strong;" 2 – 6 is "positive," and 0 – 2 is "weak" evidence. As a counterpart, Burnham and Anderson (2002) declared some rules of thumb for AIC differences, $\Delta AIC_i = AIC_i - AIC_{min}$, in which AIC_{min} is the minimum AIC value (i.e., the best model) over all models considered, which are especially useful for nested models. The larger difference in AIC indicates strong evidence against the best model in the set of models of

Table 12
Model Fit Comparisons

	AIC	*BIC*	*a.BIC*
Model 1	5213.684	4379.132	4264.820
Model 2	4211.410	4335.496	4249.496
Model 3	4206.940	4344.814	4249.554
Model 4	**4204.646**	**4301.158**	**4234.476**

Note. AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; a.BIC, sample-size adjusted Bayesian Information Criteria interest. The evidence associated with a difference of greater than 10 is "essentially none," 4-7 is "considerably less," and 0-2 is "substantial" for supporting that the model is the best model given the data.

Based on the results, the preferred model for the data at hand is the fourth one (Model 4), which is the fully mediated model. This model has the lowest values on both AIC, BIC, and aBIC. The absolute value of difference in BICs between Model 4 and the next best fitting model (Model 3) is 43.656 (= 4344.814 - 4301.158), providing very strong evidence that Model 4 is favored. However, the difference in AICs for Model 3 is 2.294 (= 4206.940 - 4204.646), providing substantial evidence for continuing to consider the alternative model. The results are summarized in Table 12 above.

3.5.2 Final Multivariate Multilevel Mediation Model

The final model is a fully mediated model in which only total reading time has direct relations to vocabulary learning, completely mediating the relation of each independent variable (i.e., Test Announcement, Time Limit, Word Length, Predictability, Part of Speech) to vocabulary learning (Model 4). I will focus on the final model by analyzing and reporting each path between variables. As I have group mean centered total reading time, I fixed intercepts at within-subject level to zero to take it out of the model. Intercepts of total reading time at between-subjects level were estimated as $b = 1.409$, $SE = .204$, $p < .001$, $exp(b) = 2.384$. Equations corresponding to this model are as follows:

i = target words (Level 1)
j = participants (Level 2)

$$M_{ij} = \beta_{0M_j} + \beta_{1Mj}WL_{ij} + \beta_{2Mj}PoS_{ij} + \beta_{3Mj}PD_{ij} + e_{M_{ij}}$$
$$\beta_{0M_j} = \gamma_{00M} + \gamma_{01M}TA_j + \gamma_{02M}TL_j + u_{0M_j}$$

$$\beta_{1M_j} = \gamma_{10M}$$
$$\beta_{2M_j} = \gamma_{20M}$$
$$\beta_{3M_j} = \gamma_{30M}$$

$$\mathrm{Y}_{ij} = \beta_{0_j} + \beta_{1j}(M_{ij} - M_{.j})$$
$$\beta_{0_j} = \gamma_{00} + \gamma_{01}\mathrm{M}_{.j} + \mathrm{u}_{0_j}$$
$$\beta_{1_j} = \gamma_{10} + \mathrm{u}_{1_j}$$

Note. WL = Word Length; PoS = Part of Speech (Verb or not); PD = Predictability; TA = Test Announcement; TL = Time Limit.

3.5.3 Effect of Test Announcement and Time Limit

In this section, I investigate the direct and indirect effects of Test Announcement and Time Limit (Level 2 predictors) on eye-tracking measures (i.e., Total Reading Time and ΔOE) and vocabulary learning.

3.5.3.1 Effect on eye-tracking measures.

As Test Announcement and Time Limit are Level 2 (subjects-level) predictors, the effect of the predictors on the mediator is fixed, given that only specific comparisons (test announcement, no test announcement, and time limit, no time limit) are of interest. To illustrate with an example, a subject read all target words in a timed condition, indicating no variation between target words. The results showed that intentional learning mode (receiving test announcement) was associated with longer processing time on average (b = .463, SE = . 180, p < .05, $exp(b)$ = 1.589) across all 12 target words whereas time restriction was not a significant predictor of total reading time (b = -.202, SE = .180, p = .274, $exp(b)$ = .817).

On the other hand, the ΔOE was not related to Test Announcement (*b* = -.213, *SE* = .738, *p* = .100, *exp(b)* = 1.237) or Time Limit *(b* = -.013 *SE* = .941, *p* = .282, *exp(b)* = .987). The results suggest that vocabulary test announcement and time pressure did not have an effect on additional attention on target words.

3.5.3.2 Indirect effect on learning.

In order to calculate the average indirect effect, I multiplied the effect of the mediator (Summed Total Reading Time or summed ΔOE) on learning by the effect of the predictor on the mediator (Summed Total Reading Time or summed ΔOE) (Preacher, Zyphur, and Zhang, 2010). The results of a formal test of the indirect effects are summarized in Table 16. The results revealed that the indirect effect of Test Announcement on Form Recognition, through summed total reading time, was significant (*b* = .258, *SE* = .131*, p* = .050*, exp(b)* = 1.294), and the indirect effect on Meaning Recognition approached the borderline of significance (*b* = .163, *SE* = 095*, p* = .086*, exp(b)* = 1.177). However, the indirect effect on Meaning Recall (*b* = .200*, SE* = .141*, p* = .156*, exp(b)* = 1.221) were found non-significant. The results indicate that the total reading time fully mediated the relationship of Test Announcement and Form and Meaning Recognition. Putting it differently, participants forewarned of an upcoming vocabulary test are likely to recognize the forms and meanings of target words, but only through paying attention to the words while reading.

A test of the indirect effect revealed no evidence of an indirect effect of Time Limit on vocabulary learning through summed total reading time: Form Recognition (*b* = -.112*, SE* = .112*, p* = .315*, exp(b)* = .894), Meaning Recognition (*b* = -.071*, SE* = .072*, p* = .324*, exp(b)* = .931), and Meaning Recall (*b* = -.087*, SE* = .097*, p* = .368*, exp(b)* = .917).

Moreover, the indirect effect of Test Announcement and Time Limit on vocabulary learning through the ΔOE was not significant: Form Recognition (*b* = .364, *SE* = .168, *p* = .130, *exp(b)* = 1.439 for TA; *b* = -.158, *SE* = .160, *p* = .322, *exp(b)* = .854 for TL), Meaning Recognition (*b* = .213, *SE* = .111, *p* = .255, *exp(b)* = 1.237 *for TA, b* = -.093, *SE* = .094, *p* = .324, *exp(b)* = .911 for TL), and Meaning Recall (*b* = .284, *SE* = .160, *p* = .176, *exp(b)* = 1.328 for TA, *b* = -.124, *SE* = .129, *p* = .339, *exp(b)* = .883 for TL).

3.5.4 Effect of Word Length, Predictability, Part of Speech

In this section, I investigate the direct and indirect effects of Level 1 (word-level) predictors including Word Length, Predictability, and Part of Speech on eye-tracking measures and vocabulary learning. The effects of Level 1 predictors are fixed in the final model, suggesting that there is no variation between participants.

3.5.4.1 Effect on eye-tracking measures.

As can be seen from Table 13, the effect of Word Length on Total Reading Time was not significant (*b* = -.021, *SE* = .014, *p* = .136, *exp(b)* = .979) but Predictability and Part of Speech were significantly associated with Total Reading Time (*b* = .004, *SE* = .002, *p* = .020, *exp(b)* = 1.004 and *b* = -.349, *SE* = .094, *p* = .004, *exp(b)* = .705, respectively). This is an indication that longer processing time was elicited by words with higher predictability than ones with lower predictability, and by nouns and adjectives rather than verbs.

The results also revealed that Word Length and Part of Speech significantly predicted the ΔOE (*b* = 135, *SE* = .180, *p* = .003, *exp(b)* = 1.140 *and b* = -.628, *SE* = 653, *p* < .001, *exp(b)* = .534, respectively), but there was no significant relationship between Predictability and ΔOE (*b*

$= .015$, $SE = .013$, $p = .267$, $exp(b) = 1.015$). The results suggest that longer words and nouns and adjectives (rather than verbs) invited longer additional attention on target words.

3.5.4.2 Indirect on learning.

Indirect effects were calculated by multiplying the Level 1 effect of the mediator (Total Reading Time or ΔOE) on learning and the effect of the Level 1 predictors on the mediator (Preacher, Zyphur, and Zhang, 2010). A formal test of the indirect effect resulted in that there is no significant indirect effect of Word Length on any vocabulary tests through Total Reading Time; Form Recognition ($b = -.008$, $SE = .006$, $p = .182$, $exp(b) = .992$), Meaning Recognition ($b = .001$, $SE = .001$, $p = .137$, $exp(b) = 1.001$), and Meaning Recall ($b = .001$, $SE = .000$, $p = .128$, $exp(b) = 1.001$) (see Table 15).

The indirect effect of predictability, via total reading time, on Meaning Recall was statistically significant ($b = .001$, $SE = .000$, $p = .018$, $exp(b) = 1.001$) and the effect on the Form Recognition was at the margin of statistical significant ($b = .002$, $SE = .001$, $p = .071$, $exp(b) = 1.002$). There was no evidence of a significant indirect effect of predictability, through total reading time, on Meaning Recognition ($b = .001$, $SE = .001$, $p = .137$, $exp(b) = 1.001$). The results of a formal test showed a significant indirect effect of Part of Speech on Form Recognition ($b = -.138$, $SE = .062$, $p = .027$, $exp(b) = .871$) and Meaning recall ($b = -.073$, $SE = .020$, $p < .001$, $exp(b) = .930$) through total reading time, but not on Meaning Recognition ($b = -.076$, $SE = -.049$, $p = .121$, $exp(b) = .927$).

Furthermore, I did not find any significant indirect of predictor variables on vocabulary learning through the ΔOE. The results are presented in Table 16. The insignificant indirect relations between the predictors and test scores suggest that the additional attention on target words (ΔOE) do not mediate those relationships.

In sum, total reading time significantly and fully mediated the relationship between predictability and meaning recall and form recognition. Likewise, the effect of part of speech on form recognition and meaning recall was fully mediated by total reading time. Taken together, the total reading time learners spent on words fully explains the association between the predictability and word learning (form recognition and meaning recall) and between the part of speech and word learning (form recognition and meaning recall) .

3.5.5. Effect of summed total reading time and ΔOE

The effect of eye-fixation measures on learning success has two components, within subjects (Level 1) and between subjects (Level 2). Table 14 provides more detailed information on the effects of eye-fixation durations on vocabulary learning in this study.

First, the within-subjects effect of eye-fixation durations on vocabulary is estimated as a random effect, allowing that the effect varies between participants. In investigating the within-subjects effect of total reading time on learning, the results showed that total reading time on target words statistically predicted learning success in all vocabulary measures, particularly in recognizing the form and recalling the meaning of the words (b = .395, SE = .139, p = .004, $exp(b)$ = 1.484; b = .209, SE = .030, p < .001, $exp(b)$ = 1.232, respectively), whereas the effect of total reading time on Meaning Recognition was borderline significant (b = .219, SE = .115, p = .058, $exp(b)$ = 1.245).

Regarding the within-subjects effect of the ΔOE on learning, the results showed that the effect of the ΔOE was statistically significant on Form Recognition (b = .149, SE

= .055, p = .006, *exp(b)* = 1.161) and borderline significant on Meaning Recognition (b = .079, SE = .040, p = .050, *exp(b)* = 1.051). However, the ΔOE was not significantly related to Meaning Recall (b = .100, SE = .059, p = .091, *exp(b)* = 1.095).

Second, focusing on the between-subjects effect, the summed total reading time significantly increased average scores on Form Recognition (b = .556, SE = .207, p = .007, *exp(b)* = 1.744), and Meaning Recognition (b = .352, SE = .165, p = .033, *exp(b)* = 1.422), but not on Meaning Recall (b = .431, SE = .271, p = .111, *exp(b)* = 1.539).

In investigating the between-subjects effect of the summed ΔOE on learning, all three types of vocabulary learning was significantly predicted by ΔOE: b = .028, SE = .013, p = .030, *exp(b)* = 1.028 for Form Recognition, b = .020, SE = .013, p = .043, *exp(b)* = 1.145 for Meaning Recognition, b = .021, SE = .011, p = .041, *exp(b)* = 1.052 for Meaning Recall.

3.5.6 Comparative strength

Unstandardized estimates represent the amount of change in the outcome variable as a function of a single unit change in the predictor variable, while the standardized coefficients indicate the amount of change in an outcome variable per standard-deviation-unit increase in a predictor variable. Thus, in order to compare the relative strengths of relations across observed variables that are measured on different scales, the standardized coefficients (β) were additionally estimated using Bayesian analyses. The standardized coefficients (β) for the paths are presented on the left side in the diagrams that follow.

At the student level (Level 2), Test Announcement was found to be a stronger predictor of Summed Total Reading Time than Time Limit. Test Announcement showed a moderate effect (β = .26) on total reading time, which in turn had a moderate effect on Form Recognition (β =.31), a strong effect on Meaning Recognition (β =.54), and a small effect on Meaning Recall (β =.21). The entirety of the indirect effect on Form Recognition (Test Announcement → Summed Total Reading Time → Form Recognition) corresponded to a significant indirect effect (see *3.5.3.2*. section for detailed values). This indirect pathway reveals a fully mediated relationship whereby Test Announcement contributed to participants' recognizing word forms at test through their increased allocation of attentional resources to words during the reading task.

Test Announcement and Time Limit did not predict the ΔOE. Nevertheless, the ΔOE predicted Form Recognition (β =.09) stronger than Meaning Recognition (β =.05), although the effects are very small.

At the word level (Level 1), Part of Speech was the most strongly associated with Total Reading Time (β = -.21) followed by Predictability (β = .12) and Word Length (β =-.04). Part of Speech and Predictability had a moderate and small effect on Total Reading Time respectively, which in turn had a small effect on each of the vocabulary tests: Form Recognition (β =.19), Meaning Recognition (β =.12), and Meaning Recall (β =.09). The indirect effect of Part of Speech on Form Recognition and Meaning Recall (Part of speech → Total Reading Time → Form Recognition/ Meaning Recall) was consistent with a significant indirect effect (see *3.5.4.2*. section for detailed values). This indirect pathway suggests that the contribution of Part of Speech (i.e., nouns and adjectives rather than verbs) to learning forms and remembering the meanings was fully mediated by total fixation times on target words. The indirect effect of Predictability on Meaning Recall also dovetailed with a significant indirect effect, which serves as evidence that total reading times fully mediate the relations between Predictability and Meaning Recall.

In addition, Part of Speech (β =.09) had a stronger effect on the ΔOE than Word Length (β =.19) had, which in turn had a small effect on Form Recognition. The indirect path (Part of Speech → Summed ΔOE → Form Recognition) was also confirmed by the results from a formal test of the indirect effect of Part of Speech on Form Recognition through the ΔOE (see *3.5.4.2*. section for detailed values).

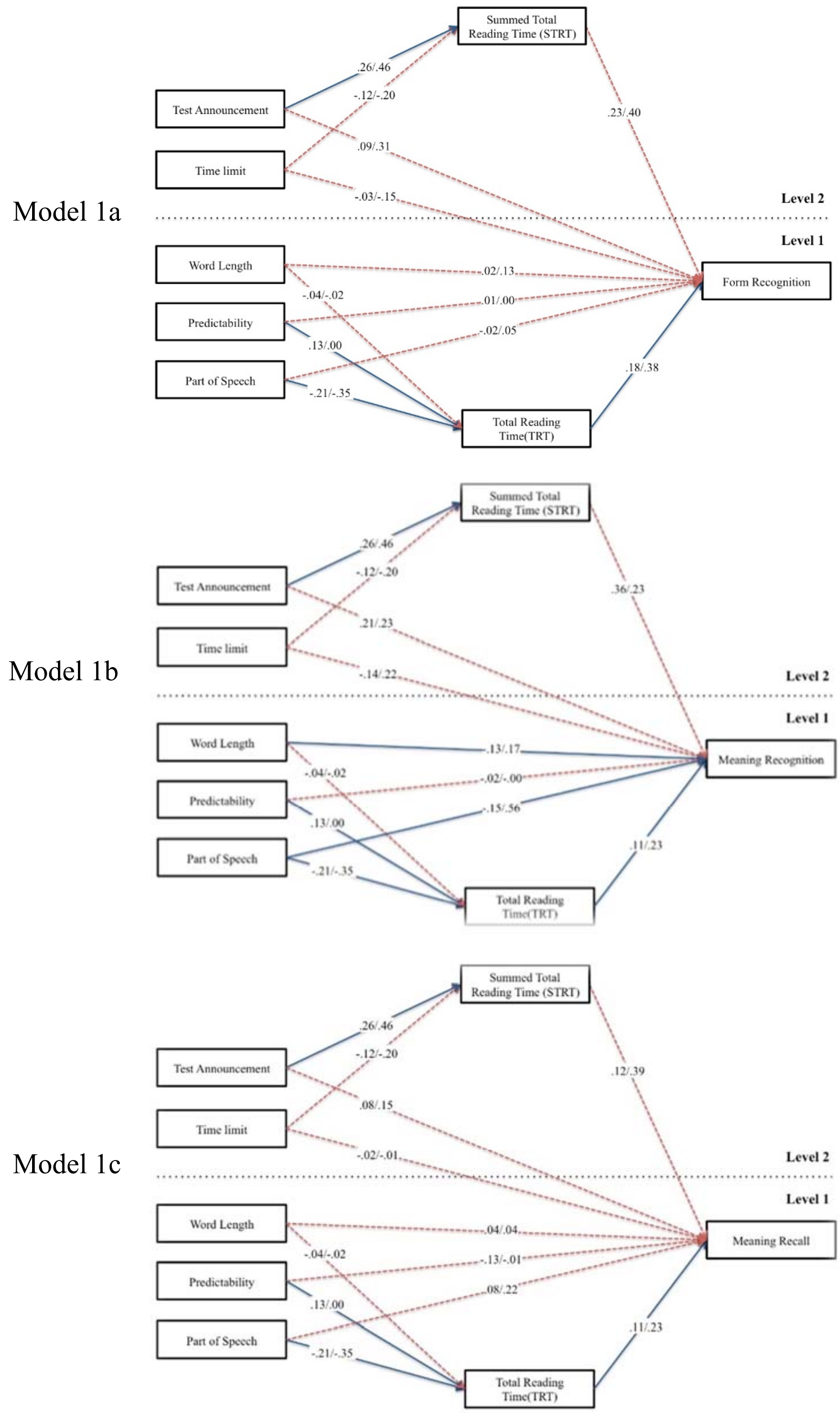

Figure 7 Alternative path model 1.

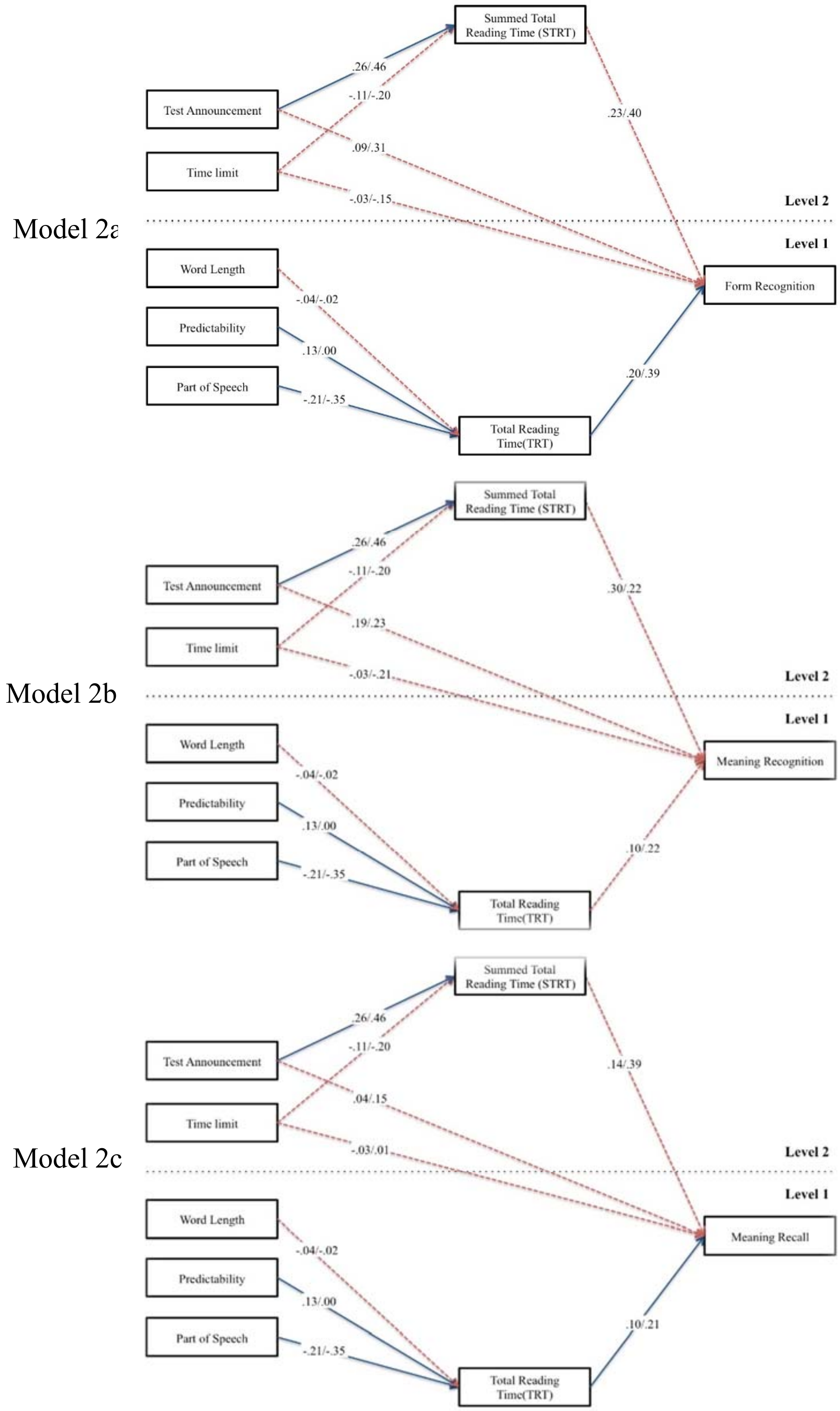

Figure 8 Alternative path model 2.

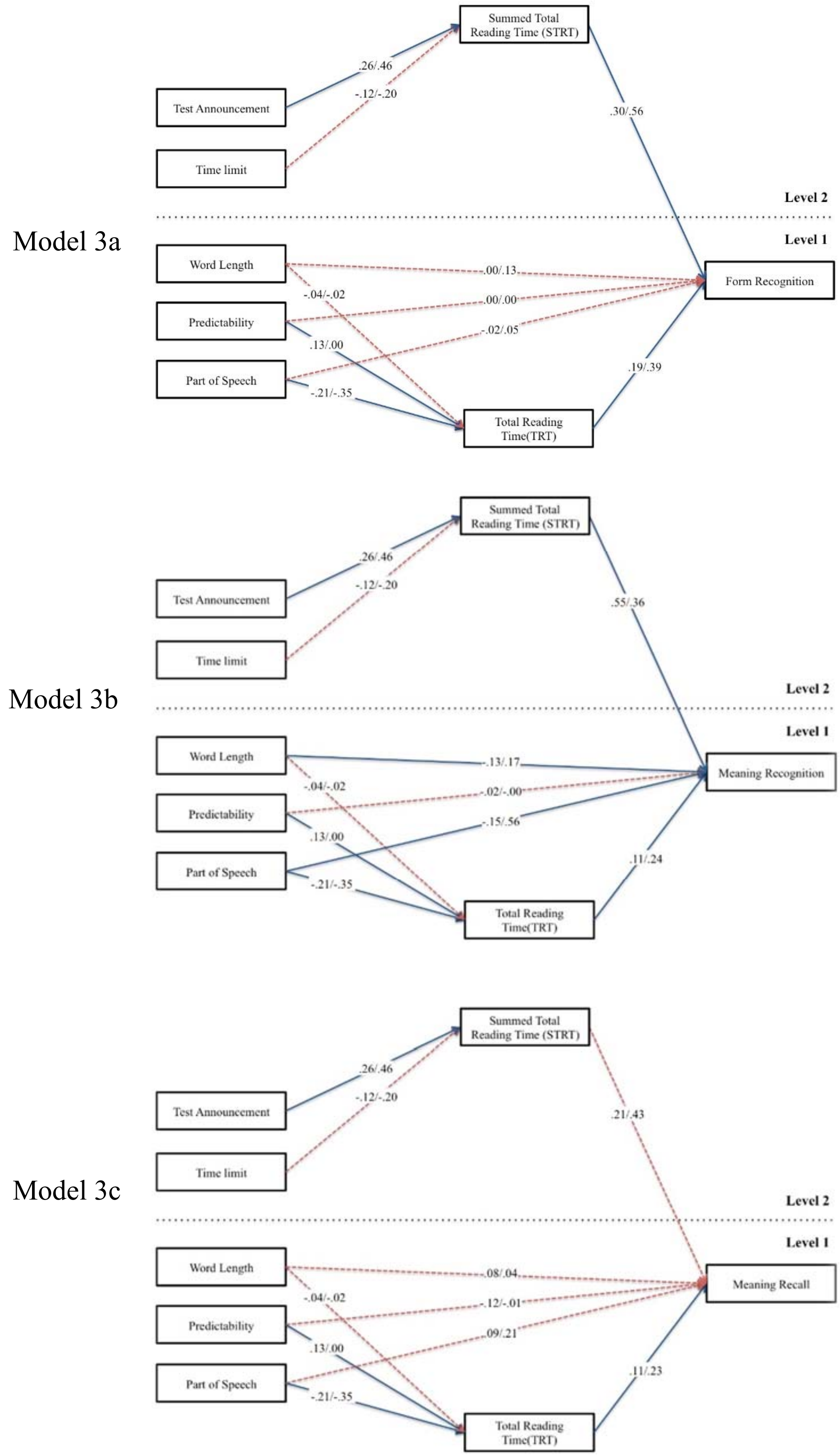

Figure 9 Alternative path model 3.

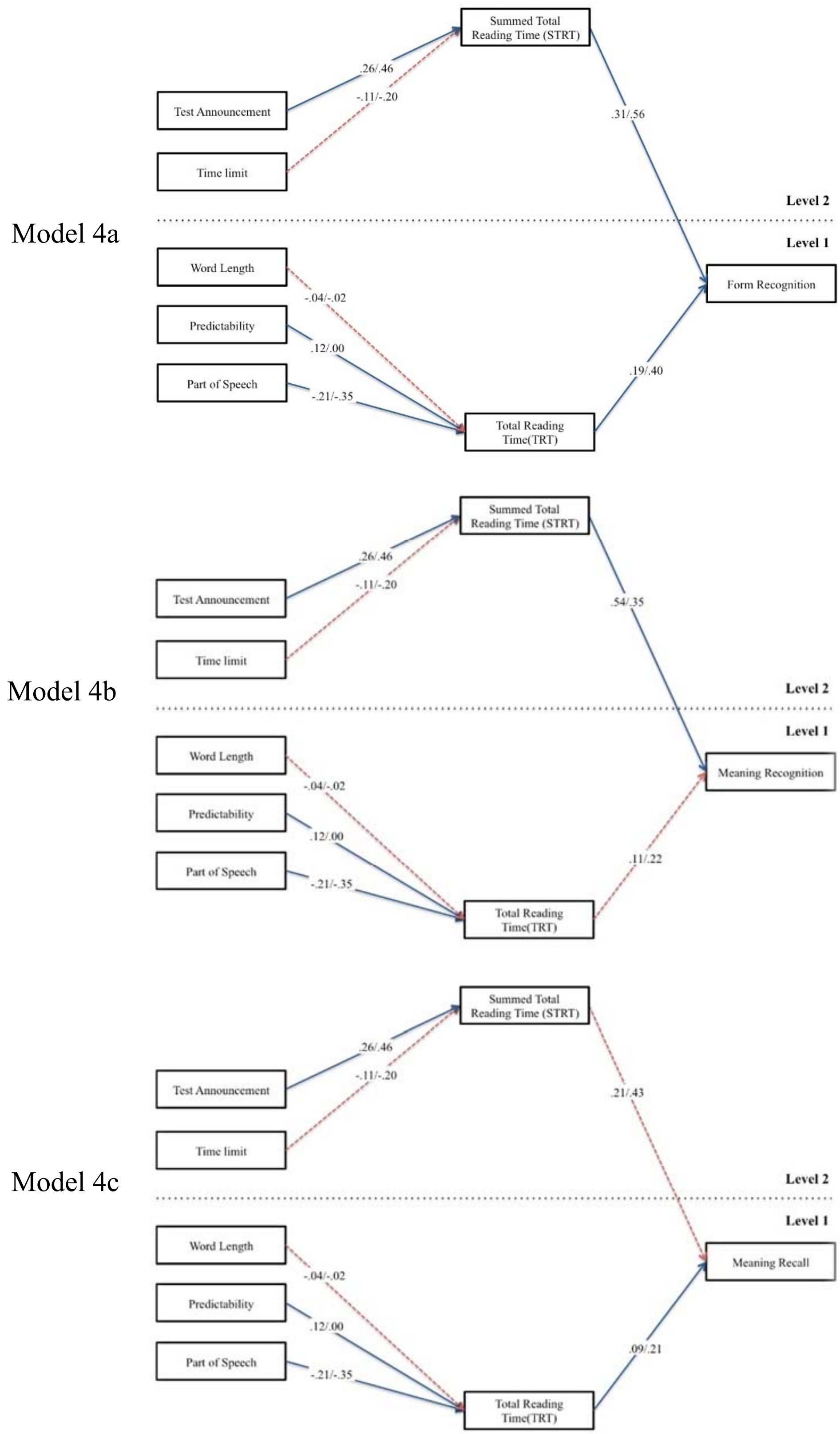

Figure 10 Final path model 4.

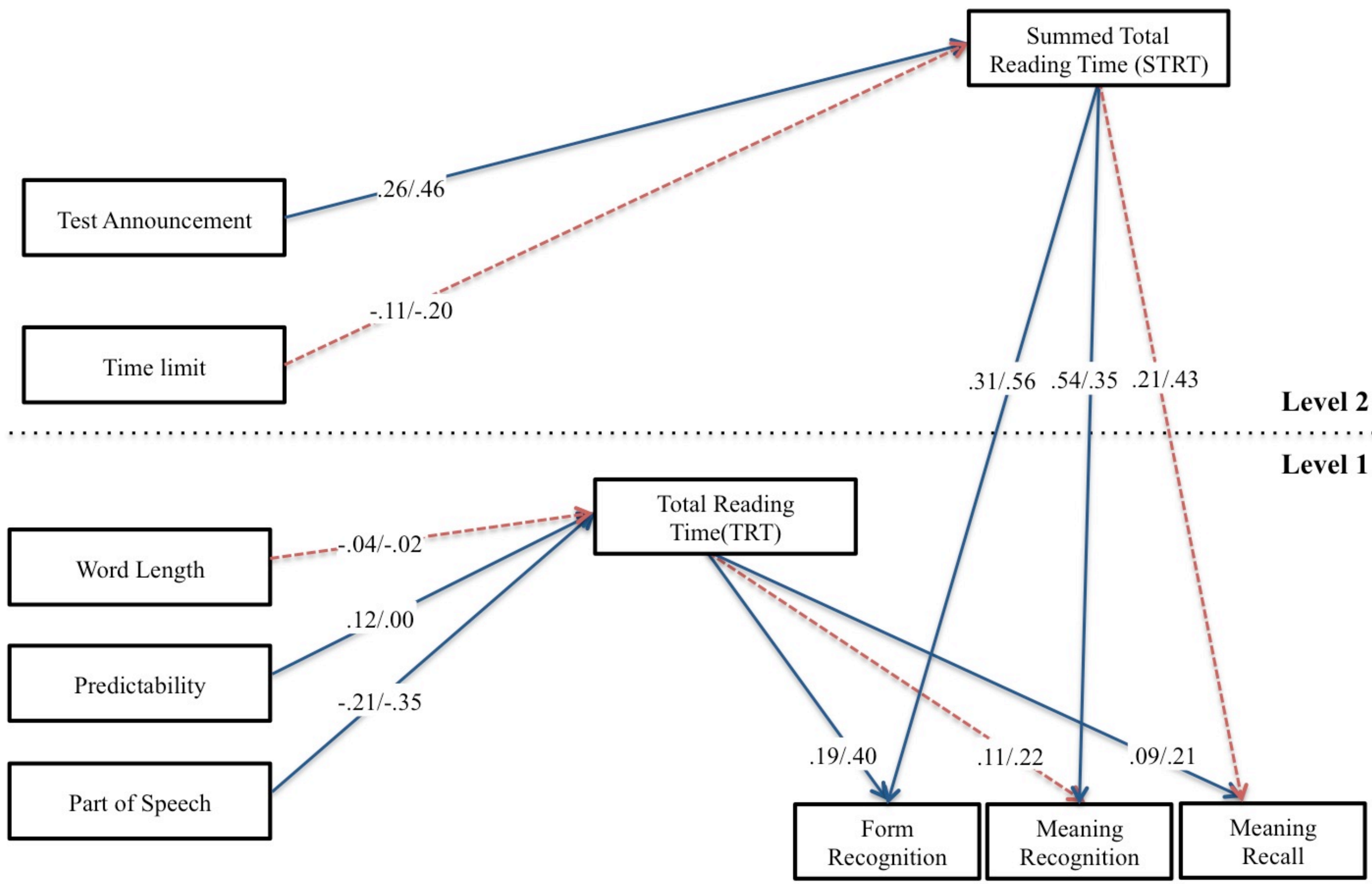

Figure 11 Final path model 4 with all dependent variables included.

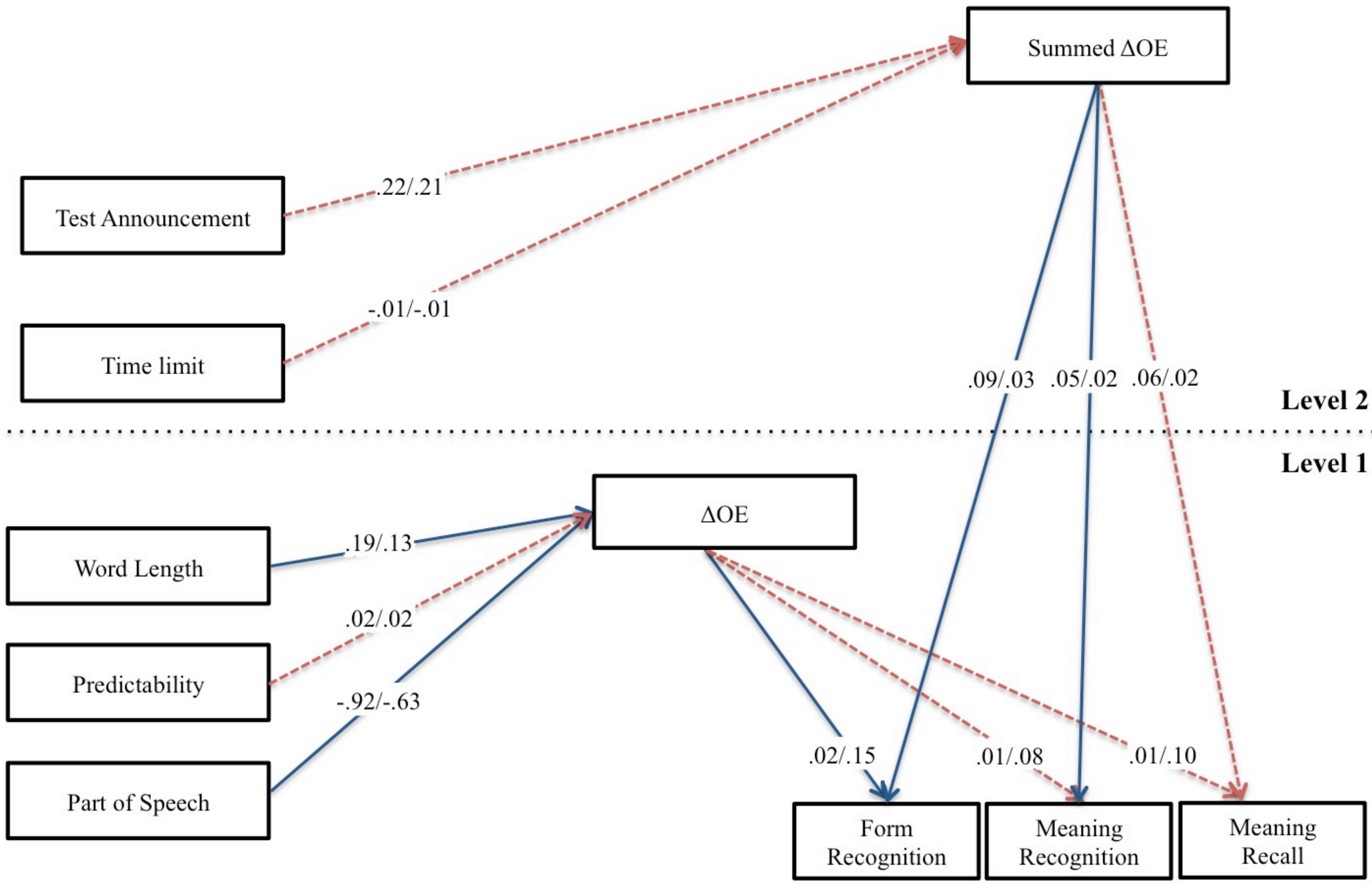

Figure 12 Path model 5 for ΔOE.

Table 13

Effects of Predictors on Total Reading Time and ΔOE

	X → TRT				X → ΔOE			
	b	*SE*	*p*	*exp(b)*	*b*	*SE*	*p*	*exp(b)*
Test Announcement	.463	.160	.004	1.589	.213	.738	.100	1.237
Time Limit	-.202	.180	.274	.817	-.013	.941	.282	0.987
Word Length	-.021	.014	.136	.979	.135	.180	.003	1.140
Predictability	.004	.002	.020	1.004	.015	.013	.267	1.015
Part of Speech	-.349	.094	<.001	.705	-.628	.653	<.001	0.534

Table 14

Effects of Total Reading Time and ΔOE *on Vocabulary Learning*

	Form Recognition				Meaning Recognition				Meaning Recall			
	b	*SE*	*p*	*exp(b)*	*b*	*SE*	*p*	*exp(b)*	*b*	*SE*	*p*	*exp(b)*
Level 1: Total Reading Time	.395	.139	.004	1.484	.219	.115	.058	1.245	.209	.030	<.001	1.232
Level 2: Summed Total Reading Time	.556	.207	.007	1.744	.352	.165	.033	1.422	.431	.271	.111	1.539
Level 1: ΔOE	.149	.055	.006	1.161	.079	.040	.050	1.051	.100	.059	.091	1.095
Level 2: Summed ΔOE	.028	.013	.030	1.028	.020	.013	.043	1.145	.021	.011	.051	1.052

Table 15

Indirect Effects of Predictors on Vocabulary Learning

	X →(S)TRT→Form Recognition				X→(S)TRT→ Meaning Recognition				X→(S)TRT→Meaning Recall			
	b	*SE*	*p*	*exp(b)*	*b*	*SE*	*p*	*exp(b)*	*b*	*SE*	*p*	*exp(b)*
Test Announcement	.258	.131	.050	1.294	.163	.095	.086	1.177	.200	.141	.156	1.221
Time Limit	-.112	.112	.315	.894	-.071	.072	.324	.931	-.087	.097	.368	.917
Word Length	-.008	.006	.182	.992	-.005	.004	.216	.995	-.004	.003	.128	.996
Predictability	.002	.001	.071	1.002	.001	.001	.137	1.001	.001	.000	.018	1.001
Part of Speech	-.138	.062	.027	.871	-.076	-.049	.121	.927	-.073	.020	<.001	.930
	X →(S) ΔOE →Form Recognition				X→(S) ΔOE → Meaning Recognition				X→(S) ΔOE →Meaning Recall			
Test Announcement	0.364	0.168	0.130	1.439	0.213	0.111	0.255	1.237	0.284	0.160	0.176	1.328
Time Limit	-0.158	0.160	0.322	0.854	-0.093	0.094	0.324	0.911	-0.124	0.129	0.339	0.883
Word Length	0.008	0.006	0.191	1.008	0.005	0.004	0.212	1.005	0.005	0.003	0.133	1.005
Predictability	0.002	0.001	0.081	1.002	0.001	0.001	0.129	1.001	0.001	0.000	0.100	1.001
Part of Speech	-0.134	0.063	0.035	0.875	-0.082	0.051	0.107	0.921	-0.080	0.022	0.092	0.923

4. DISCUSSION AND CONCLUSION

In This study, I adopted an eye-tracking method to investigate the cognitive processes of vocabulary learning under incidental and intentional conditions. In addition to disparate theoretical viewpoints concerning incidental and intentional L2 vocabulary learning, significant gaps in the research remain concerning the efficacy of incidental and intentional learning for L2 vocabulary acquisition. Through this study, I was able to address some of their gaps; in particular, I provided further evidence for the interrelationships between intention (incidental vs. intentional), attention (total reading time and the difference between the expected and observed reading time), and vocabulary learning (test scores) through a multivariate multilevel mediation model.

4.1. Offline vocabulary post-test measures

Consistent with previous research (Godfroid et al., 2017; Laufer & Goldstein, 2004; Mohamed, 2017; Pellicer-Sánchez & Schmitt, 2010; Schmitt, 2008, 2010), the results of the current study confirmed the incremental nature of vocabulary acquisition. Not surprisingly, form recognition yielded the highest scores (46.75% on average), followed by meaning recognition (29.81% on average), and finally meaning recall (10.08 % on average). These results imply that the ability to recall a word's meaning is the most advanced knowledge, whereas recognizing a word's form and meaning can be considered a less advanced degree of word knowledge. The results also confirm that receptive knowledge develops before productive knowledge as Schmitt (2010) claimed. This hierarchical difficulty of vocabulary acquisition can be explained in terms of the different level of cognitive demands on the learner's memory (Mohamed, 2017). Retrieving a word's meaning requires inferencing processes combined with relevant contextual information, which leads to the development of a form-meaning association. This task is considered deeper, more effortful, and more demanding than the meaning recognition task, in which learners are presented with retrieval cues to aid recall of a stimulus that had been encountered while reading. In contrast, a form recognition test can be said to measure learners' knowledge under the lowest cognitive demands because the test only creates the need to retrieve orthographical information from memory.

Comparing the three experimental conditions, I found that the Untimed Intentional group had the highest scores, followed by the Timed Intentional, and then the Timed Incidental groups in all vocabulary tests. However, the observed differences were not found to be statistically significant when analyzed using ANOVA. The results support for the Involvement Load Hypothesis (Laufer and Hulstijn, 2001) by demonstrating that the intentional condition inducing higher involvement load statistically differed from the incidental condition in vocabulary learning measured by three postreading vocabulary tests. Another finding is that participants in the Timed Incidental group, although showing the lowest vocabulary gains out of the three groups, demonstrated gains nonetheless. The results seem to suggest that incidental vocabulary learning is worth students while in that participants in the Timed Incidental group were able to recognize 43.75% of the forms and 24.50% of the meanings of target words, despite their relatively low levels of effort, although they recalled the meanings of only 8.08% of the items (all figures given as averages). Consistent findings regarding the effect of incidental vocabulary learning have also been reported in previous studies (e.g., Godfroid et al., 2017; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003).

However, it is important to note that these methodological procedures or pedagogical interventions per se do not directly contribute to vocabulary learning. It is rather the *amount* and *quality* of cognitive processes that determines, predicts, and explains the actual learning of vocabulary. The involvement load hypothesis (Hulstijin & Laufer, 2001), depth of processing (Craik & Lockhard, 1972), and the notion of engagement (Schmitt, 2008) all represent attempts to theorize the role of the cognitive processes in lexical learning. Consistent with these various theoretical positions, the full mediation model (Model 4) supports a mediating role of attention in vocabulary learning. The lack of significant direct effects of both Test Announcement and Time Restriction on Form Recognition, Meaning Recognition, and Meaning Recall provided strong evidence for the mediation role of cognition.

4.2. Online eye-tracking measures

To help support vocabulary learning, participants were asked to read the same text twice. I compared the summed fixations on target words from the first and second sessions, revealing that participants gave relatively less attention to the target words when they encountered them for the second time as compared to the first time. The decrease in the amount of attention was also observed in lower fixation counts and in the difference between observed and expected eye-fixation durations during the second session compared to the first session (See Table 10). One possible explanation is that repeated encounters of target words increased learners' familiarity with them, which led to faster processing of the lexical forms. This is the argument often offered to explain decreased repeated reading times in second language vocabulary learning (Godfroid et al., 2017; Joseph et al., 2014; Mohamed, 2017; Pellicer-Sánchez, 2015).

However, it is important to note that repetition in the current study differed slightly from previous literature in that the repeated exposure took place through rereading the text. That is, learners in the current study encountered target items twice in an identical context. The psycholinguistic view would explain this as repeated readings of the same lexical items or the same text expediting the retrieval process by expanding the number of related memories available to be retrieved. When participants read the text for the first time, they may have created a memory trace of operations they performed t access the meanings of unknown words. These memories of the first reading could then facilitate the second reading by reproducing the same contextual clues, comprehension strategies, and extralinguistic information. Two earlier studies, Hyönä and Niemi (1990) and Raney and Rayner (1995), presented data relevant to this point. They also demonstrated that participants read faster during the second or third iteration over the first one, which reflects the reduced processing demand.

Another interesting observation is that despite the increased familiarity with the target words, participants in the intentional groups (Group 2 and Group 3) displayed longer processing times on the second encounter of several target words than the first encounter (apprise and staggering for Group 2 and fatality, perilous, and staggering for Group 3; see Figure 5). Given their strong departure from the overall decreasing trend, I speculate that the additional time represents the particular cases that learners paid deliberate attention to the lexical items during the second reading. The findings also highlight the effect of explicit warning of the forthcoming tests, which creates an intentional learning environment and might control the degree of learners' attention on targets. This experimental manipulation, supported by ΔOE data, is especially important for researchers who would like to look at the effect of intentional and incidental task types.

4.3. Looking at many variables combined: The multilevel multivariate mediation model

I tested a multilevel multivariate mediation model of vocabulary learning, which postulates that the attention learners have paid to target words provides a critical link between intentionality and learning gains for adult language learners. A series of alternative models were tested to find the most parsimonious model that was able to accurately represent the observed data. The results confirmed the Total Reading Time and the difference between observed and expected eye-fixation durations as mediators, fully mediating the relationship between Test Announcement, Time Limit, Word Length, Predictability, and Part of Speech, and vocabulary learning as measured by three vocabulary tests. This full mediation implies that the cumulative reading time and additional attentional processing time fully explains the association between treatments and textual factors and the learning of new vocabulary. To put it differently, full mediation can be considered an indication of the importance of attention in explaining the total effect.

Another noteworthy finding was that Test Announcement had a significant effect on Total Reading Time, indicating that participants who were aware of the post-reading vocabulary tests spent a significantly longer time on target words than those who were not aware. I replicated a former finding from eye movement studies (e.g., Godfroid et al., 2013; Godfroid et al., 2017) by demonstrating that longer processing time on target words was a significant predictor for all the vocabulary outcomes except for meaning recall, but to somewhat different degrees. This causal pathway of vocabulary test announcement on total reading time, which in turn aids learning of new lexical items, is also evidenced by the significant indirect effect of Test Announcement on the vocabulary tests.

Interestingly, Test Announcement did not significantly affect the difference between observed and expected eye-fixation durations, which nevertheless served as a significant predictor of learning of form and meaning. Considering that the difference between observed and expected eye-fixation duration indicates additional attentional processing (Indrarathne &

Kormos, 2016), informing participants of a vocabulary test might not lead them to disproportionately allocate their attention to target words. This finding potentially reveals that superior learning of the intentional group is due to the fact they spend more time on task overall and therefore also more time on the target words, but their processing did not change qualitatively. To put it differently, the distinction between intentional and incidental learning may reflect a quantitative difference in overall reading time rather than a qualitative difference in the learners' orientation to the target forms.

4.4. Methodological contribution to SLA

One of the major contributions of this study is its attempt to adopt a multilevel multivariate mediation model to gain a multi-faceted picture of incidental and intentional vocabulary learning. In the field of the SLA, many studies involve multiple outcomes, including immediate or delayed posttests, and pendent variables, including type of instruction × learner characteristics. Eye-movement measures also come in many different forms. Especially in second language vocabulary studies, researchers are in agreement that learners' vocabulary knowledge has to be assessed by multiple measures (e.g., Nation, 2001, Schmitt, 2008, 2010, Webb, 2005, 2007) to have a more accurate assessment of the degree and type of knowledge that has been learned. However, many researchers do not test for differential treatment effects across outcomes directly, but often go on to interpret their results as if they had indeed done so very good. For example, Pellicer-Sánchez (2017) conducted Kruskal-Wallis tests separately on each eye-movement measures to find the effect of repetition on her four eye-tracking measures. She compared the statistical significance for different learning measures by stating that "For targets, the effect of repetition was significant in all measures expect for first fixation duration whereas for controls it was only significant in two measures (p. 113)." However, an explicit test is required to test whether the magnitude of the repetition effect differs depending on the types of eye-fixation measures (e.g., fixation counts but not first fixation duration). In other words, a statistical test of the difference is required to compare the treatment effects. Using a multivariate multilevel model, which can accommodate two or more outcomes in a single analysis, can alleviate this concern because it explicitly treats the outcomes as discrete, co-varying variables. In the current study, in fact, I did find the differential effect for recall and recognition and for total reading time and ΔOE (albeit not in the same model).

The multilevel multivariate mediation model is also useful to investigate hierarchically clustered data. SLA researchers are increasingly using data organized in two or more hierarchical levels, such as students nested within groups, classes nested within schools, or words (repeated measures) nested within individuals. Methodologically, however, the clustered structure leads to correlations among the observations. For example, in the eye-tracking study, eye-fixation data from one participant are likely to be correlated relative to data from another participant. This correlation compromises the assumption of independence for many statistical tests, which in turn contribute to incorrect *p*-values, confidence intervals, and effect sizes (Baldwin, Murray, & Shadish, 2005). One benefit of multilevel models is that they allow researchers to model between-cluster variability using random effects and/or unconflated multilevel modeling (UMM) to accommodate the correlation of the clustered data. Additionally, the multilevel multivariate mediation model is easier to correct for missing data and non-normally distributed data problems.

Last but not least, the multilevel multivariate mediation model is specifically helpful for eye-tracking studies in the field of SLA because attention, as a mediating variable, is foundational to many SLA theories. For instance, the Noticing Hypothesis (Schmitt, 1990) claims that intake is the subset of input that is attended and noticed. The involvement load hypothesis (Hulstijn and Laufer, 2001) states that tasks with a higher involvement load cause more engagement, which then results in better learning outcomes. It is no longer satisfying to find out whether some pedagogical treatments have an effect on learning in a specific context. Researchers seek to further understand how such effects come to be and what the causal pathways are through which treatments exert their effects. For example, Winke (2013) examined separately the effect of input enhancement on learners' attention and on learners' form learning and meaning comprehension. Mediation analysis is one way that a researcher can explain a causal chain of relationships such as this one, where input enhancement treatments promotes learners' attention to the targeted form, which then facilitates form learning. Therefore, the current study is meaningful in that the effect of intention, attention, and learning outcomes were examined simultaneously in one model as hypothesized based on SLA theory.

4.5. Limitations and future research

The results of the current study should be interpreted in the context of several limitations, one being the study's reliance on eye-fixation measures. Although eye-tracking data from readers reveal how much attention the readers are giving to each target, the quality of processing (i.e., the type and nature of processing) cannot be explored through eye-tracking data. Godfroid and Schmidtke (2013) similarly highlighted the need for triangulating eye-movement data, verbal reports, and vocabulary learning scores by stating that "eye-tracking data can only tell us what participants looked at but not what their internal thought processes were" (Godfroid & Schmidtke, 2013, p.185, also Leow, Grey, Marijuan, & Moorman, 2014, Winke, 2013).

Therefore, it is recommended the results of the eye-tracking analysis be complemented by participants' retrospective verbal reports such as stimulated recall in future studies.

Another limitation is related to the types of eye-fixation measures examined. The study is limited to the examination of two eye-tracking measurements: the sum of all fixations within an interest area and the difference between the expected and observed fixation durations. Total fixation duration was favored because it is regarded as one of the main measures in the field of SLA (Godfroid & Uggen, 2013; Issa et al., 2015, Winke, Gass, & Sydorenko, 2013). Further, it is known to index the later stages of processing, such as information reanalysis and text comprehension (Clifton, Staub, & Rayner, 2007), which I thought would be the most relevant to the current study. Considering that different measures tap into different stages of learning, future research could benefit from including various measures to obtain richer accounts of learners' online language processing. Additionally, a large portion of participants' data had to be excluded from the analyses. Out of 80 participants, a total of 36 were eliminated. Thirty-one of the 36 were removed because of the technical issues related to the eye-tracking method such as inaccurate calibrations and a large amount of track loss.

Perhaps an additional weakness of this study is that participants reported floor effects on the meaning recall test, with only 1.21 out of 12 target words being correctly recalled. Same as in other studies, as discussed in the methods section, this is likely the reason why the MSEM model did not converge appropriately. This also may be the reason why I could not find a significant effect of Total Reading Time on Meaning Recall at the between-subjects level. In fact, it is not surprising that participants remembered the meaning of only about 10% of the target words, given that the target words were read merely twice. In this respect, although the answers to how many encounters are needed to learn unknown words is inconclusive, it is clearly more than two. Many researchers have suggested at least eight to ten encounters, as many researchers have suggested (e.g., Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2007) should be provided in order to obtain immediate, measurable learning gains. Raising the number of exposures would be more likely to induce meaningful variances on meaning recall tests.

Certain limitations of the study are related to the methodological issues regarding the nature of the instruments and tasks in the present study. First, due to the lack of delayed-post tests, I cannot infer whether learning gains were durable. Second, laboratory experiments, although allowing for the precise control of variables, may not fully capture the potential effects of incidental and intentional reading on vocabulary learning. Specifically, reading the relatively short text, reading the same text twice in succession, not being able to go back to the previous page, and using the head mount and a chin rest while reading did not resemble real-word incidental learning, all came at the expense of ecological validity. In some ways, the study procedure departed from real word incidental learning.

Considering the nature of incidental vocabulary learning, the use of long and authentic text with more extended reading time is necessary. Ideally, longitudinal studies with more prolonged vocabulary instruction should be conducted in an effort to produce robust and generalizable findings.

4.6. Pedagogical implication

The main implication that can be taken from the findings in this study is the role of attention in language learning, which has been underscored by many researchers (e.g., Godfroid et al., 2013; Leow, 2015; Robinson, Mackey, Gass & Schmidt, 2012; Schmidt, 1990, 2001). The results reveals that the pre-learning instructions led learners to allocate attentional resources to unfamiliar words in the reading, which, in turn, facilitated their processing of those words, and thus promoted vocabulary learning. Taking into consideration vocabulary pedagogy, the present study may offer some evidence that implicit and explicit techniques to make target vocabulary more salient can be effective tools to direct learners' attention to the targeted lexical items, increasing the time spent in reading targeted vocabulary. For instance, language instructors could use input enhancement techniques such as underlining, capitalizing, boldfacing, and color-coding to increase the perceptual salience of target forms in the input. However, the enhanced input does not essentially guarantee that learners will notice and pay attention to the targets. Thus, it is important for instructors to explicitly increase the salience of target linguistic forms through metalinguistic explanations, consciousness-raising tasks, corrective feedback, or the announcement of vocabulary tests ahead of time as employed in the current study. This is in line with Schmitt (2008), suggesting that supplementing extensive reading with explicit teaching activities boosts students' engagement and maximizes their learning gains. Furthermore, the study sheds light on the need to strengthen students' motivation or intentionality to learn unknown lexical items while reading. Although I fostered the extrinsic/instrumental motivation through the vocabulary test announcement, it would be more ideal for instructors to create activities that hold intrinsic interests for learners (for a fuller explanation of intrinsic and extrinsic motivation, see Ryan & Deci, 2000).

4.7. Conclusion

This eye-tracking study is the first of its kind to directly compare incidental and intentional learning with the mediating effect of attention in a single analysis. Corroborating many previous studies (e.g., Schmidt, 2011), the study confirmed that learning gains are strongly associated with the attention participants pay to the input. It moreover showed that vocabulary test announcement significantly predicted longer fixation durations across the whole text and not the target word specifically. The results indicate that the intentional and incidental learners may differ in terms of their overall reading times (i.e. time on task) rather than particularly increased attention paid to the target words. Lastly, I attempted to accelerate the adoption of the modern method of multilevel multivariate mediation analysis in the field of SLA by challenging conventional statistical methods and offering new alternatives. Given the inherently multivariate nature of SLA, I hope that other researcher will follow this lead and explore causal pathways between instruction, cognition and learning outcomes.

APPENDICES

Appendix A

Experimental text

Smart Cars, Intelligent Highways

Cars today are smart. No, they may not be smart enough to change their own oil or find the lost coins in their seats, but they are smart and getting smarter. The average car today has more computing power than the 1969 Apollo 11 spacecraft that carried the first astronauts to the moon. Every car produced today has at least one computer for monitoring fuel consumption and pollution controls. The average car uses twelve computerized devices, and high-end cars have many more, controlling everything from the sunroof to the braking system. In the near future, cars may be virtually stuffed with computer chips from front fender to taillight. That's because motorists enjoy computerized gizmos, and providing these little devices is cheaper for automakers than building a better engine or making other engineering changes that might actually be more important.

Many of the smart features we are seeing today are safety-related. Some are systems to avoid collisions. These may use sonar, radar, lasers, computers, or video cameras, or some combination of these. These systems beep or warn drivers with a voice signal if the vehicle gets too close to an object or another vehicle or if it strays out of its lane. The system can suggest actions to the driver or even temporarily take control to avoid calamity. Another safety device is a smart airbag system. To deploy airbags with the minimum necessary force, sensors determine an occupant's weight and size and the severity of impact. This system should reduce the number of children hurt by airbags that open too vigorously. Another system can automatically notify emergency services that an accident has happened and, using a Global Positioning System (GPS), can pinpoint the location of the vehicle for police and rescue units. This system can save precious minutes and many lives.

One of the most convenient aspects of smart cars is their ability to navigate. Drivers tell them where they want to go and then, by means of a GPS navigation device and computerized maps, smart cars can figure out the best ways to reach the drivers' destinations. The cars can show the information on a map or give drivers voice directions. They can even correct drivers if they make a mistake. Using communication devices connected to the Internet, cars can apprise drivers of problems ahead – construction work, traffic jams, and accidents – and then suggest different routes to the drivers' offices, favorite pizza places, or closest shopping malls.

Smart cars create problems as well, however. One problem is how to control all this automotive technology. More buttons take more of the drivers' attention. Even voice controls are bewildering for drivers. A recent study showed that drivers talking on handheld cell phones were four times more likely to be involved in accidents as drivers who were not. In fact, drivers using cell phones were almost as likely to be involved in accidents as those who were legally intoxicated. Using voice controls, even a hands-free system, might prove to be as perilous as chatting on the phone. Nevertheless, the auto industry's answer to the control problem so far has been voice control. When it comes to simple tasks – changing channels on the radio or opening the trunk – voice controls work well enough. But it is probably not the best method for directing more difficult operations such as navigating the Internet or controlling the car itself. Engine noise, highway noise, and the music on the stereo tend to garble instructions, and voice recognition systems often cannot decipher strong accents.

No matter how smart cars become, they cannot solve all the problems facing a "car-crazy" world by themselves. Anyone who has traveled by car in or around almost any city in the world knows that the problem of traffic congestion is becoming worse every year. Cars, buses, and trucks caught up in the incessant traffic jams in the cities waste vast amounts of fuel and pour pollution into the atmosphere. Then there are the terrible statistics for highway fatalities. In the United States alone, over 40,000 people die a year. Around the world, it is believed that between 800,000 and 1.15 million succumb in automobile accidents annually. Some transportation planners believe that better mass transportation is the answer – more monorails, subways, and bullet trains. Other analysts believe that there will always be a demand for the convenience and independence of private automobiles. The traditional solution has been to simply build more roads. However, another solution is self-driving vehicles operating on automated "intelligent" roadways.

What is an "intelligent" roadway? It is one type of automated highway that features one or more lanes on which vehicles with special sensors and communications systems can travel completely under computer control. The vehicles follow each other at closely spaced intervals in groups called "platoons". (Some lanes would also have to be open to conventional cars). Vehicles in platoons traveling on the automated lanes would be temporarily linked into communications networks. These vehicles could then constantly exchange information about speed, acceleration, braking, and so on. To keep vehicles in their lanes and control their speed and direction, special devices in cars might be used to sense magnets buried in the roadbed. One expert has said that the typical highway lane today can handle 2,000 vehicles per hour but estimated that an intelligent highway lane could accommodate up to 6,000 vehicles, depending on the number of entrances and exits.

The technology required to operate an automated highway already exists and has been tested. On a stretch of San Diego Expressway, a platoon of seven smart cars traveled on a lane of intelligent highway. The cars followed one another about 5 meters apart at around 105 kilometers per hour. The drivers sat back and sipped their lattes. They said that traveling that fast and that close together with no control was exciting and a little frightening at first, but that, it became rather humdrum in a short time.

But don't plan to have your car chauffeur you to work any time soon. For one thing, the cost would be staggering. Even equipping one lane of traffic on the busiest urban expressways with the necessary technology would be too expensive to do in the near future. Installing the required equipment on cars would also add thousands of dollars to the cost of new cars. Besides, many people would not trust self-driven cars. Much of the public has a warped sense of risk. Some people hesitate to fly even though studies show that flying is safer than driving. That's because every plane crash is highly publicized, while individual automobile accidents are not. Similarly, although automated cars would certainly be safer than standard cars, when an accident occurred it would probably involve hundreds of deaths and injuries. Even a few such accidents would probably cause the public to call for the closing of automated roads.

Appendix B

Language background questionnaire

General Information

1. Gender: □ Female □ Male

2. Age: ________________________

3. What best describes your eyesight?
   □ I can see well without glasses or contact lenses.
   □ I can see well when I wear my glasses.
   □ I can see well when I wear my contact lenses.
   □ I have a vision impairment. → Please explain (optional):
   ____

4. Do you have any reading or learning disabilities (e.g., dyslexia)?
   □ No
   □ Yes → Please explain (optional):
   ____
   □ I prefer not to say

5. Year in college: □ Freshman □ Sophomore □ Junior □ senior □ MA/Ph.D. □ n/a
   If you are enrolled in ELC, please indicate your level: _____________________

6. Major field of study: ________________________

7. What is the main language you speak at home? ________________________

8. What other languages do you speak at home?
   ____

9. How long have you been living in the U.S.? ________(years) _________(months)

10. Have you lived or travelled in other countries (excluding the U.S.) for more than three months?

    □Yes □ No

    If **YES**, fill in the following table. If you answered **NO** to question 10, go to question 11.

Country	Length of stay	Language used in the country

11. Have you taken a standardized English proficiency test (e.g., iBT TOEFL, IELTS, TOEIC, OPIc)?

□Yes □No

If **YES**, fill in the following table. Please list your reading test score(s) for test(s) that contains the reading test section (e.g., iBT TOEFL, IELTS, TOEIC, etc). If you answered **NO**, go to question 12.

Test	Total score	Reading test score

Language Learning Background

12. How long have you been studying English? ___________(years)

13. In which contexts/situations did you study English? Check all that apply.
    - ☐ At home (from parents, caregivers)
    - ☐ At school (Primary, secondary, high school)
    - ☐ At private institutions
    - ☐ After immigrating to an English-speaking country
    - ☐ At language courses during my study abroad in an English-speaking country
    - ☐ Other (specify): ____

14. What percentage of your time each day do you speak English (as opposed to other languages)? ____________%

15. Please rate on a scale of 1-6 your current ability on English reading, writing, and listening (circle the number below). (1= *Beginner*; 2= *Pre-intermediate*; 3= *Intermediate*; 4= *Upper-intermediate*; 5= *Advanced*; 6= *Native-like*)

Reading	Writing	Speaking	Listening
1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6

16. Please rate on a scale of 1-6 your interest in studying English (circle the number below).

Strongly interested Not interested

1	2	3	4	5	6

17. Have you studied other foreign language(s) (besides English)?

□ Yes □ No

If **YES**, fill in the following table. Rank the language(s) from you know best to the one you are least familiar with.

Rank	Language	Age first learned the language	Number of years spent learning

Appendix C

Sample of prescreening vocabulary test

NVLT Part 6 – Academic Word List

1. concept: This is a difficult **concept.**
a. legal agreement
b. idea about what something is
c. way of doing things
d. a written explanation of a law

2. similar: These articles are **similar.**
a. about a certain thing
b. of great quality
c. easy to understand
d. close to the same

3. item: The next **item** is very important.
a. thing on a list
b. question sheet
c. meeting of people
d. way something looks

4. component: Each **component** is very important.
a. set of ideas which support something
b. flat part that sits on top of another
c. small part of something bigger
d. the person you work with

5. compensate: The government should **compensate** the farmers.
a. give something good to balance something bad
b. stop them from joining a group
c. find where they are
d. bring them together

6. professional: She wants to be a **professional** musician.
a. someone who stays at home
b. someone who gets paid to play
c. someone on a list
d. someone known by many people

7. external: They worried about the **external** damage.
a. not known
b. outside
c. based on facts
d. following

8. clause: Please fix that **clause.**
a. part of a sentence
b. something you are trying to do
c. large picture
d. small object

9. migrate: The animals began to **migrate.**
a. work together
b. move together to a different place
c. come together as a group
d. change together

10. priority: That is our **priority.**
a. deal between two people
b. most important thing
c. something that has been printed
d. person who comes next

11. reverse: Try it in **reverse.**
a. the other direction
b. the way things are arranged
c. with the correct sound
d. at the correct time

12. humdrum: I kept a journal about **humdrum** life in England.
a. lacking variety
b. causing excitement
c. involving risks
d. having good health

13. arbitrary: Her decision was **arbitrary.**
a. not chosen for a reason
b. necessary for success
c. not able to be changed
d. good enough for a purpose

14. warped: He has a **warped** sense of reality.
a. positive
b. improved
c. little
d. strange

15. mutual: The feeling was **mutual.**
a. easy to understand
b. fully developed
c. the same between two people
d. kept under control

16. decipher: It is difficult to **decipher.**
a. succeed in understanding the meaning
b. find the exact size and amount
c. speak in an unclear way
d. go in front of somebody

17. alternative: Is there an **alternative**?
a. another choice
b. thing to do
c. something to say
d. activity with many people

18. wreck: I have never been in a **wreck.**
a. accident
b. relationship
c. fight
d. film

19. colleague: That is my **colleague.**
a. something that people talk about
b. plan of things to do
c. person you work with
d. piece of writing

20. perilous: Their situation was **perilous.**
a. very dangerous
b. totally different
c. somewhat exciting
d. completely unfair

21. legal: Is this meeting place **legal**?
a. based on the law
b. free to be used
c. easy to see
d. important to someone

22. congestion: There is **congestion** on their route.
a. the state of being crowded
b. the act of constructing buildings
c. the condition of snow-covered surface
d. the situation of over-speeding

23. site: He looked for a better **site.**
a. basic part of something
b. opinion about the price
c. place where something is
d. something brought from another country

24. severity: It shows the **severity** of the problem.
a. root
b. solution
c. seriousness
d. context

25. institute: We must **institute** new changes.
a. get with effort
b. control with laws
c. begin or create
d. search for

26. apprise: He was **apprised** of the situation.
a. relieved
b. advised
c. ashamed
d. informed

27. retain: How will the club **retain** its members?
a. mix them together
b. help them develop
c. help them work together
d. keep them

28. chauffeur: He **chauffeured** me to the stadium.
a. sent
b. invited
c. drove
d. follow

Appendix D

Sample of reading proficiency test

Reading Reading Reading

This passage is about animal behavior.

In order to deal with their environment, animals depend not only on instinct and individual learning but also on something called *social learning*. One famous example of social learning was observed among a group of macaque monkeys in Japan. In the 1950s, researchers noticed a young monkey washing the sand off a sweet potato in a stream before eating it. Over time, this behavior spread to other monkeys of the group, and today, potato washing among the macaques is common.

Biologists report another example of social learning on an island in the U.S. state of Virginia. They noticed herring gulls using hard paved roads to crack open clamshells in order to get at the clam meat. The gulls take clams out of a river, fly two hundred meters to a road and then drop the clams onto the pavement.

For over three years, the biologists collected and measured thousands of these broken clamshells. The gulls seem to prefer medium-sized clams, about three inches wide. The researchers think that a shell smaller than three inches isn't worth the energy needed to drop it because it doesn't contain much meat. A large clam has more meat but is too heavy to carry. As the birds grow older, they seem to get better at calculating the right clam size and the most efficient dropping height.

There are five species of gulls on the island, but only the herring gulls drop clams. The biologists are not certain how the herring gulls first learned to do this, but think that herring gulls may be able to learn from one another. The other gull species on the island, however, do not appear to be capable of social learning.

126. What is the main purpose of the passage?
    a. to explain the differences between social learning, individual learning, and instinct
    b. to compare macaque monkeys and herring gulls
    c. to show how researchers observe wild animals
    d. to show that some animals use social learning

127. How did the group of monkeys mentioned in paragraph 1 learn to wash potatoes?
    a. by a process called individual learning
    b. by using a natural instinct
    c. by watching researchers
    d. by watching other monkeys

128. According to the passage, what did the herring gulls learn to do?
    a. break clamshells
    b. carry heavy clams
    c. wash their food
    d. find clams in the river

129. How did the biologists conduct their study of the gulls?
    a. They looked for clams in the river.
    b. They collected broken clamshells.
    c. They practiced dropping clamshells on the road.
    d. They taught the gulls how to find clams.

130. What conclusion did the biologists come to about the other four species of gulls on the island?
    a. They aren't able to learn from other gulls.
    b. They are not as social as the herring gulls.
    c. They depend too much on social learning.
    d. They depend too much on instinct.

Appendix E

Comprehension test

Read the following statements and circle **"True"**, **"False"**, or **"I don't know"**.

1. The main purpose of the passage is to advertise smart cars.

True False I don't know

2. The author predicts that there would be more computer chips in cars in the near future.

True False I don't know

3. Many of the smart features in cars are related to weather.

True False I don't know

4. In case of emergency, smart cars can make a beeping sound.

True False I don't know

5. In the event of an accident, a smart airbag can determine the driver's weight.

True False I don't know

6. Sometimes, the voice recognition system makes mistakes because of noise.

True False I don't know

7. On an intelligent roadway, cars are not allowed to communicate with other cars.

True False I don't know

8. Intelligent highways will increase the number of traffic accidents.

True False I don't know

9. An intelligent highway and smart cars have already been tested in San Diego.

True False I don't know

10. The author recommends that people use self-driving cars very soon.

True False I don't know

Appendix F

Form recognition test

Direction: Circle words you saw in the reading. If you do not know the answer, DO NOT GUESS. There is a penalty for wrong answers.

1	banter	potent	jest	calamity	I don't know
2	chisel	adorn	apprise	montage	I don't know
3	entice	perilous	reparation	iterate	I don't know
4	succumb	marquee	carnage	unjust	I don't know
5	redress	scour	gizmo	bait	I don't know
6	forsake	stampede	tremor	chauffeur	I don't know
7	staggering	immaculate	colossal	rubble	I don't know
8	agnostic	plunder	incessant	ferret	I don't know
9	wrangle	fatality	bucolic	atoll	I don't know
10	bawdy	bewildering	gulp	adroit	I don't know
11	entice	wean	decipher	heinous	I don't know
12	veer	distend	cardinal	sip	I don't know

Appendix G

Meaning recognition test

Direction: Circle the correct meaning of each given word. If you do not know the meaning, please circle "I don't know".

	Word	Meaning				
1	**bewildering**	confusing	surprising	amusing	boring	I don't know
2	**calamity**	conflict	accident	mistake	risk	I don't know
3	**decipher**	outline	repair	understand	respect	I don't know
4	**chauffeur**	develop	drive	generate	match	I don't know
5	**staggering**	extremely hot	very high	quite normal	more accurate	I don't know
6	**fatality**	capacity	control	death	difficulty	I don't know
7	**incessant**	diverse	powerful	massive	constant	I don't know
8	**succumb**	die	donate	realize	maintain	I don't know
9	**apprise of**	inform of	think of	consist of	make of	I don't know
10	**gizmo**	case	gift	device	bag	I don't know
11	**sip**	brew	order	serve	drink	I don't know
12	**perilous**	related	impossible	dangerous	continued	I don't know

Appendix H

Meaning recall test

Direction: For each word, write down anything you can remember about its meaning.

	Word	Meaning
1	gizmo	
2	fatality	
3	calamity	
4	apprise	
5	decipher	
6	succumb	
7	chauffeur	
8	sip	
9	perilous	
10	incessant	
11	bewildering	
12	staggering	

Appendix I

Instruction by condition

Group 1: Incidental Timed Group

You are about to read two passages. Be sure to read for meaning. You will be tested afterward on your **OVERALL** comprehension. **You have 30 seconds to read each screen.** There are 30 screens in total, 17 for each reading passage. You will have a break between the reading passages. When you are finished with one screen, press OK to advance to the next screen.

Group 2: Intentional Untimed Group

You are about to read two passages. You will be tested afterward on your OVERALL comprehension and vocabulary. So, it is IMPORTANT that you read for meaning and understand EVERY SINGLE word at the same time. **There is NO time limit**. There are 30 screens in total, 17 for each reading passage. You will have a break between the reading passages. When you are finished with one screen, press OK to advance to the next screen.

Group 3: Intentional Timed Group

You are about to read two passages. You will be tested afterward on your OVERALL comprehension and vocabulary. So, it is IMPORTANT that you read for meaning and understand EVERY SINGLE word at the same time. **You have 30 seconds to read each screen.** There are 30 screens in total, 17 for each reading passage. You will have a break between the reading passages. When you are finished with one screen, press OK to advance to the next screen.

REFERENCES

REFERENCES

Ashby, J., Rayner, K., and Clifton, C., Jr. (2005) Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58*(A), 1065-1086.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing multiple outcomes in clinical research using multivariate multilevel models. *Journal of Consulting and Clinical Psychology, 82*, 920 –930.

Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals, 37*(2), 200-208.

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly, 43*(1), 79-103.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11,* 142–163.

Bowles, M. (2010). *The think-aloud controversy in second language research*. New York, NY: Routledge.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 33*(2), 247-271.

Bruton, A., Garcia Lopez, M., & Esquiliche Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable Term? *TESOL Quarterly*, *45*(4), 759-768.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16*, 157-168.

Bryan, A., Schmiege, S. J., & Broaddus, M. R. (2007). Mediational analysis in HIV/AIDS research: Estimating multivariate path analytic models in a structural equation modeling framework. *AIDS and Behavior*, 11, 365–383.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye-movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53-78). Oxford: Oxford University Press.

Carr, T. & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition, 16*, 205-230.

Calvo, M., & Meseguer, E. (2002). Eye-movements and processing stages in reading: Relative contributions of visual, lexical, and contextual factors. *The Spanish Journal of Psychology, 5*, 66-77.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341– 372). Oxford, UK: Elsevier.

Cobb, T. (2013). The Compleat Web VP! (in progress) Available online from www.lextutor.ca/vp/bnc/

Coll, J. (2002). Richness of semantic encoding in a hypermedia-assisted instructional environment for ESP: Effects on incidental vocabulary retention among learners with low ability in the target language. *ReCALL, 14*, 263-284.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom and second language acquisition*. Cambridge: Cambridge University Press.

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology*, *16,* 79-103.

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

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365–414.

Ellis, N. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1-32). London: Academic Press.

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning, 5*(1), 1-32.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of research. *Studies in Second Language Acquisition, 24*, 223-236.

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition, 29*(1), 119-126.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273−293.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344.

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.

Halla, J. W. (1988). A psychological study of psychometric differences in Graduate Record Examinations General test scores between learning disabled and nonlearning disabled adults. Unpublished doctoral dissertation. Texas Tech University, Lubbock.

Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333.

Gass, S., Svetics, I., & Lemelin, S. (2003). Differential affects of attention. *Language Learning, 53*, 497-545.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., …Yoon, H. (2017). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition,* 1-22.

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.

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: John Benjamins.

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge. *Studies in Second Language Acquisition, 37*, 269–297.

Godfroid, A., & Schmidtke, J. (2013).What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J.

M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawaiʻi, National Foreign Language Resource Center.

Godfroid, A., & Spino, L.A. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, *65*(*4*), 896–928.

Godfroid, A., & Uggen, M. S. (2013). Attention to irregular verbs by beginning learners of German. *Studies in Second Language Studies, 35*(*2*), 291–322.

Grabe, W., & Stoller, F. L. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 98-122). Cambridge: Cambridge University Press.

Hill, G. A. (1984). Learning disabled college students: The assessment of academic aptitude. Unpublished doctoral dissertation. Texas Tech University, Lubbock.

Halla, J. W. (1988). A psychological study of psychometric differences in Graduate Record Examinations General test scores between learning disabled and nonlearning disabled adults. Unpublished doctoral dissertation. Texas Tech University, Lubbock.

Hama, M., & Leow, R. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*(3), 465-491.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, *11*, 459–479.

Hedeker, D., & Gibbons, R. D. (2006). Longitudinal data analysis. Hoboken, NJ: Wiley.

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, H. C. M. (2013). The Effects of Word Frequency and Contextual Types on Vocabulary Acquisition from Extensive Reading: A Case Study. *Journal of Language Teaching and Research, 4*(3), 487-495.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*(2), 181-193.

Hulstijn, J. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H. Dechert & M. Raupach (Eds.), *Interlingual processing* (pp. 49-73). Tübingen: Gunter Narr.

Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: Macmillan.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.

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

Indrarathne, B., & Kormos (2016). Attentional processing of input in different input conditions: an eye- tracking study. *Studies in Second Language Acquisition. 39*(3), 401-430.

Indrarathne, B., & Kormos (2017). The role working memory in processing L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition, 21*(2), 355-374.

Issa, B., Morgan-Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. In L. Roberts, K. McManus, N. Vanek, & D. Trenkic (Eds.), *EUROSLA Yearbook 2015* (pp. 114–142). Amsterdam, The Netherlands: John Benjamin Publishing.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing in second language acquisition? *TESOL Quarterly*, *34*, 239–278.

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.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 262-284.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23*, 418–444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249–277.

Laufer, B. (1997). The lexical plight in second language reading Words you don't know, words you think you know and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press.

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo & J. Cenoz (Eds.), *EUROSLA Yearbook 5* (pp. 223–250). Amsterdam: Benjamins.

Laufer, B. (2006). Comparing focus on form and focus on formS in second-language vocabulary learning. *Canadian Modern Language Review, 63*, 149-166.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*, 399–436.

Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing, 19*(1), 21-48.

Leow, R. P. (2015b). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 47-65). Amsterdam: John Benjamins.

Leow, R. P., Grey, S., Marijuan, S., & Moorman, C. (2014). Concurrent data elicitation procedures, processes, and the early stages of L2 learning: A critical overview. *Second Language Research, 30*(2), 111-127

Leow, R., Hsieh, H., & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning, 58*, 665-695.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis , S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.) *Implicit and explicit knowledge in second language learning, testing and teaching*, (pp. 94 – 112). Bristol, UK: Multilingual Matters.

Loewen, S. (2011). The role of feedback. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 24—40). New York: Roudedge.

Long, M. (1996). The role of linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition*. San Diego: Academic Press.

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

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marcoulides, G. A., & Schumacker, R. E. (1996). Introduction. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. (pp. 1–6). Mahwah, NJ: Erlbaum.

McLaughlin, B. (1965). 'Intentional' and 'incidental' learning in human subjects: The role of instructions to learn and motivation. *Psychological Bulletin, 63*, 359-376.

Mohamed, A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. Studies in Second Language Acquisition. doi: 10.1017/S0272263117000092

Morgan, G., Hodge, K., Wells, K., & Watkins, M. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*, 2–20.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 667 -682.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Nation, I. S. P. (2005). British national corpus. Retrieved from http://www.natcorp.ox.ac.uk/

Neumann, O. (1996). Theories of attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of perception and action*, Volume Three: Attention (pp. 389-446). London: Academic Press.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(*1*), 97–130.

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

Peters, E. (2006). L2 vocabulary acquisition and reading comprehension: The influence of task complexity. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 178-198). Clevedon: Multilingual Matters.

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*(1), 113-151.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18,* 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.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233.

Preacher, K., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. Structural Equation Modeling: *A Multidisciplinary Journal, 18*, 161-182.

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.

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., Ashby, J., Pollatsek, A., & Reichle, E. (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.

Rayner, K., Sereno, S., & Raney, G. (1996). Eye-movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188-1200.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain sciences, 26*(4), 445-476.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in non-reading tasks: A unified framework for understanding the eye-mind link. *Psychological Review*, 119, 155–185.

Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *VIEWS, 12*, 24-39.

Roberts, L. & A. Siyanova-Chanturia. 2013. Using eye-tracking to investigate topics in L2 acquisition and L2 sentence and discourse processing. [Special issue]. *Studies in Second Language Acquisition 35* (2). 213–235.

Robinson, P. (1995). Attention, memory, and the "Noticing" Hypothesis. *Language Learning, 45*, 283-331,

Robinson, P., Mackey, A., Gass, S., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *Routledge handbook of second language acquisition* (pp. 247-267). New York: Routledge.

Rogers, B. (2005). World class readings 3 student book: A reading skills text. New York, NY: McGraw-Hill.

Rosa, E., & Leow, R. (2004). Awareness, different learning conditions, and L2 development. *Applied Psycholinguistics, 25*, 269-292.

Runyan, M. K. (1991a). The effect of extra time on reading comprehension scores for university students with and without learning disabilities. *Journal of Learning Disabilities, 24*, 104–108.

Ryan, M.R., & Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.

Sachs, R., & Polio, C. (2007). Learners\ uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language Acquisition, 29*(67-100).

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

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*, 1270–1281.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11-26.

Schmidt, R. (1995). Consciousness and foreign language learning: a tutorial on the role of attention and awareness in learning *R. Schmidt (Ed.), Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language*

Teaching Research, 12(*3*), 329–363.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istatnto, M. Nagami, J. W. Sew, T. Suthiwan & I. Walker (Eds.), *Proceedings of CLaSIC*. National University of Singapore: Center for Language Studies.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese *R. Day (Ed.), Talking to learn: Conversation in second language acquisition* (pp. 237-322). Rowley, MA: Newbury House.

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

Schwanenflugel, P. J., Stahl, S. A., & Mcfalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*(4), 531-553.

Snijders, T. & Bosker, R. 2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: Sage publications.

Sok, S. (2014). Deconstructing the concept of 'incidental' L2 vocabulary learning. *Teacher's College, Columbia University Working Papers in TESOL & Applied Linguistics, 14*(2), p.21-37

Stahl, S. (1999). Vocabulary development. Cambridge, MA: Brookline Books.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language,* 64, 88–107.

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

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge*. Studies in Second Language Acquisition, 27*(01), 33-52.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(*1*), 46–65.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269-304.

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

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A Modified Replication of Lee (2007) with Eye-Movement Data. *Studies in Second Language Acquisition, 35*(2), 323-352.

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.

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

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effect of frequency and contextual richness. *Canadian Modern Language Review, 57*(*3*), 541–572.

Zhang, R. (2015). Measuring university-level l2 learners‘ implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37*(3), 457-486.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the Maximum Potential Thesis. Applied Measurement in Education, 13(1), 99–117.