Looking at words: An eye-tracking investigation of incidental vocabulary learning

Abstract

In this study I test the claim that the number of repeated exposures and the amount of attention to novel vocabulary words in a foreign language predict readers’ incidental acquisition of multiple aspects of knowledge of the words. Thirty English-speaking learners of Arabic read 100 medium-length sentences with 12 novel (pseudo-words) arranged in three exposure conditions: four, seven or ten encounters, while their eye movements were recorded. Participants were tested on measures of word production, form recognition, meaning recognition, and meaning recall.

Eye-movement results showed that learners fixated more on initial encounters with target words and that their fixation times gradually decreased from first to last exposure. The longer they looked at novel words, the more learning gains they reported, particularly in meaning recognition and recall of these words. A GEE regression model reported the number of exposures as a stronger predictor of vocabulary gains than online processing times. Results from paper-and-pencil vocabulary knowledge tests confirmed a significant effect of exposure condition on all aspects of word knowledge. Learning gains were the highest in form recognition and lowest in meaning recall, which points to a potential cognitive trajectory of incidental lexical development. The study provides practical implications and further directions for investigating the cognitive aspects of reading comprehension and incidental vocabulary acquisition.

Keywords: Incidental vocabulary acquisition, repeated exposure, eye tracking, lexical knowledge, Arabic as a Foreign Language.
Introduction

In second language vocabulary research, a widely held view is that lexical development can occur incidentally, particularly through reading (Fraser, 1999; Kweon & Kim, 2008; Matsouka & Hirsh, 2010; Nation, 2001; Paribakht & Wesche, 1999; Pellicer-Sánchez & Schmidt 2010; Pulido, 2007; Rott, 1999; Watanabe, 1997; Webb, 2008). One factor that has received considerable attention in reading studies is repeated exposure, which reportedly accounts for substantial differences in learning gains on different aspects of word knowledge (Horst, Cobb & Meara, 1998; Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003; Webb, 2007).

Lexical gains have also been found to depend on the cognitive load of reading tasks (Hulstijn & Laufer, 2001; Kim, 2008; Keating, 2008). However, these studies were all paper-and-pencil based ones with learning measured post-hoc, through testing. They captured the offline result of lexical learning, but provided no insights into the online cognitive processes of individuals while they were exposed to vocabulary or engaged in instructional tasks.

Cognitive approaches to vocabulary learning basically relied on Schmidt’s (1990) noticing hypothesis, in that learners are expected to notice the target linguistic items while focusing on meaning in reading tasks (see Laufer, 2005), and that the amount of attention directed to a specific lexical item would facilitate subsequent initial intake of that item after exposure (Godfroid, Boers & Housen, 2013). A large challenge, however, was how to capture these intricate cognitive processes as they occur in real time. Studies that used think-aloud protocols or interviews might have been the first to probe into the cognitive processes underlying lexical acquisition (e.g. Paribakht and Wesche, 1999; Fraser, 1999). With the recent advent of eye tracking technology, psychology-based research has set the scene for more sophisticated language acquisition inquiries. Based on the ‘eye-mind link’ assumption that learners’ eye
movements accurately reflect real-time processing (Rayner, 2009), Godfroid, Housen, and Boers (2010) and Godfroid et al. (2013) used eye tracking as an innovative technique to measure the amount of attention L2 readers paid to novel vocabulary in reading tasks. The authors found evidence of an association between attention and word recognition.

The initial hypothesis that motivates the current study is that frequent exposure to novel lexical items invites more attention to form and meaning, which maybe reflected in more time for processing and more opportunities for incidental intake and retention. Because the previous study measured attention on single exposures to lexical items, the question arises whether accumulated attention to multiple exposures would account for variable gains in word knowledge. Although the factors of exposure and attention can understandably be correlated (Schmitt, 2008), that is, more exposure opportunities available for the reader would likely trigger more attention to the given targets, it is worthwhile to test this hypothesis empirically so as to disentangle the effects of repeated exposure and accumulated attention over target vocabulary during reading.

In the present study I implement eye-tracking methodology to investigate both online and offline aspects of incidental vocabulary learning from reading Arabic-like pseudo words embedded into short sentences. The main goal of the study is to track the cognitive effect of repeated exposure to vocabulary on the patterns of reading exhibited by English-speaking learners of Arabic, and whether the learners’ eye-movement reading measures reflect the development of different components of their vocabulary knowledge, including form and meaning recognition and meaning recall.
Review of Literature

Earlier hypotheses on incidental vocabulary learning have generally characterized the distinction between incidental and intentional learning in terms of the goal of reading. Wesche and Paribakht (1999) defined incidental learning of vocabulary as what happens when learners are focusing on understanding meaning rather than on the explicit goal of learning new words. However, they also claimed that learning through reading is in some fundamental sense not ‘incidental,’ at least from the learner’s perspective. Bruton, Garcia Lopez and Esquiliche Mesa (2011) took a similar position, pointing out that even if incidental learning can be confined to naturalistic (outside-of-the-classroom or un-instructed) contexts, the learning may include aspects of intentional learning strategies on the part of learners. In this sense, Bruton and colleagues partially supported a methodological distinction derived from psychology that incidental learning holds when participants are not told of an upcoming test after a given treatment (Hulstijn, 2001, 2003). At the same time, they also suggested a more objective operationalization in terms of what they called induced vocabulary salience. The present study combines both the psychological and SLA views to account for incidental learning from reading as learners in this study do not anticipate being tested after reading, and the target items were made intrinsically salient through repetition.

Incidental Vocabulary Acquisition from Reading

Reviews of vocabulary studies usually indicate that incidental vocabulary learning is often slower than intentional learning (Horst, 2005; Hulstijn, 2001; Laufer, 2005; Macaro, 2003; Read, 2004; Schmitt, 2008). However, researchers and teachers tend to agree that both modes complement each other. This is because it is impossible to cover all vocabulary needs in the classroom context and that creating opportunities for incidental exposure can positively support
partial lexical knowledge and aid explicit teaching.

Previous studies on incidental learning from reading focused on how to promote engagement in reading tasks either by manipulating word presentation in text or administering different tasks with varying degrees of complexity (e.g., Hulstijn, 1992; Hulstijn, Hollander, & Greidanus, 1996; Hulstijn & Trompeter, 1998; Paribakht & Wesche, 1997; Peters, Hulstijn, Sercu & Lutjeharms, 2009; Watanabe, 1997). The results in these studies were mostly in favor of inducing more cognitive effort while learning new lexical items. Building on Craik and Lockhart’s (1972) Depth of Processing Hypothesis, Laufer and Hulstijn (2001) introduced the involvement load hypothesis to account for the pattern of results observed in previous literature. The hypothesis was based on an analysis of the cognitive and motivational involvement imposed by any given L2-vocabulary task. Involvement, a cognitive-motivational construct, was defined as the combined effects of need, search and evaluation. Tasks that induce higher involvement were hypothesized to produce higher vocabulary gains. The hypothesis received empirical support from several studies (e.g. Keating, 2008; Kim, 2008). In the same line, Schmitt (2008) referred to engagement with lexical items as a key factor in vocabulary learning. Engagement, in his view, can be fostered by several other interventions beyond what the involvement hypothesis claimed, including, but not confined to, frequency of exposure, increased attention to target words, and increased time spent on the target items.

**Exposure Frequency in Vocabulary Learning**

Several studies demonstrated the effect of repeated exposure on vocabulary learning. Rott (1999) found that six exposures led to better word retention than two or four exposures in reading comprehension. Horst, Cobb and Meara (1998) found that words that appeared eight or more times in a long novel had a better chance of being learned by most of the participants than words
that appeared less often. Waring and Takaki (2003) found that about eight exposures to target words in extensive reading led to a 50% chance of participants’ recognizing form and meaning in a delayed multiple-choice test three months later, while less than five repetitions was not enough to translate a word correctly after three months.

Repetition was also found to play different roles in the development of other aspects of word knowledge besides word meaning. Webb (2007) investigated multiple components of word knowledge including spelling, associations, syntax, grammatical functions, and form-meaning mapping. He found that the group that encountered the target words more than 10 times showed a better grasp of different aspects of word knowledge than other groups who received fewer exposures. Similarly, studies on extensive reading found that reliable learning of vocabulary started after 10 exposures (Pellicer-Sanchez & Schmitt, 2010; Pigada & Schmitt, 2006). Schmitt (2008) thus concluded from previous research that an average of 8 to 10 repetitions could be appropriate for the development of receptive knowledge of vocabulary while leading only to relatively low gains in productive knowledge.

**Exposure Frequency and Attention**

The present study is motivated by the fact that many factors that Schmitt (2008) identified as determinants of engagement (frequency of exposure, increased attention to target words, and increased time spent on the target items) are interrelated. This hypothesized interrelatedness was not directly addressed in earlier studies on incidental learning. As shown above, there were studies that investigated engagement as operationalized through the involvement load hypothesis and other studies that provided evidence for the positive effect of exposure per se. In line with the latter strand of research, Folse (2006) reported that vocabulary retention can result from repeated retrieval of lexical items through multiple tasks, rather than the
mere involvement load of the tasks themselves. Because the involvement load is can be seen as a manifestation of attention, a valid comparison between repetition and attention should rather attempt to disentangle their effects empirically and account for their interrelatedness at the same time. What Folse (2006) stated invites investigation into whether these two variables (repetition and attention) are necessarily correlated and whether they produce strongly correlated or independent effects on learning gains.

In terms of measuring attention in real time processing, Godfroid et al. (2013) reviewed several methods that have been used in psychology studies including note taking, underlining and think-aloud protocols and eye tracking. They concluded that, with all the precision and completeness it has to offer, can provide a more sensitive measure of the amount and locus of attention to linguistic input. Reviews of eye tracking research have argued that eye movements reflect an accurate representation of the cognitive processes in the reader’s mind. This assumption was coined the ‘eye-mind’ link hypothesis, which proposes a connection between overt and covert attention (Rayner, 1998, 2009). While recording eye movement patterns in natural reading, researchers were able to manipulate text features and word properties, such as frequency, predictability, familiarity and other context variables in order to examine their effects on reading behavior.

**Eye Tracking in Vocabulary Studies**

Eye tracking is defined as the online recording of learners’ eye-movement behavior, which is described in terms of fixation times (how long readers look at something) and saccades (the movement of the eyes from one point to the next) (Godfroid, 2012). Eye movement measures tend to be categorized as either early measures of processing (e.g.; first fixation and first pass time) or late processing measures (e.g.; second pass time and total time). First fixation duration
captures the time of the first look at the target area (for example, a novel vocabulary word) when encountered for the first time during forward reading. First pass time, which is also termed gaze duration, combines first fixation duration along with any other fixation made on the target area at the initial visit before the eyes move forward or backward to the next target area. Second pass time captures the rereading episodes when the eyes reenter an interest area after they have left it. Total reading time is the sum of all fixations on the target area (see Winke, Godfroid, & Gass, 2013).

Eye movement research on reading has pointed to several factors that could affect eye-fixation times including word frequency, context predictability, word length (Kliegl, Nuthmann, & Engbert, 2006), part of speech, familiarity, and concreteness (Liversedge & Findlay, 2000; Rayner, 1998, 2009; Starr & Rayner, 2001). In psychology, one line of research was to disentangle the effects of printed word frequency, repetition and subjective familiarity ratings on fixation times. Chaffin, Morris, and Seely (2001) reported that the familiarity of target words and context quality determined the amount of time readers spent on the targets. Raney and Rayner (1995) investigated the effects of word frequency on native-English speaker’s rereading performance (reading the same text two times). They found that individuals had shorter reading times, made fewer fixations, and had longer saccades during the second reading. Moreover, shorter fixation durations were associated with high frequency words, suggesting independent effects of frequency and repetition on reading times. Hyönä and Niemi (1990) tracked Finnish readers’ performance on three text repetitions. Readers’ fixation times decreased consistently from first to third encounter with target sentences, their progressive fixations and number of regressions also decreased. Rayner, Raney, and Pollatsek (1995) found similar results regarding the effect of three repetitions of lexical items in a given text, and they also found frequency
effects on the first two repetitions, but no further differences occurred after that.

With regard to the association between online processing and learning, Williams and Morris (2004) examined the effect of word familiarity in reading comprehension and word recognition. They found that readers spent more processing time on novel words than familiar words, and that there was a systematic relationship between online processing patterns, i.e.; reading times, and retention of new word meanings. Brusnighan and Folk (2012) conducted a self-paced, L1-reading study on incidental vocabulary learning. They found that readers spent more time processing sentences that contained novel compound words, and that they were able to retain new word meanings from a single exposure.

One recent study that specifically targeted vocabulary in second language reading was conducted by Godfroid et al. (2013). They operationalized attention to novel pseudo words as a quantitative variable reflected in the participants’ eye fixation times during reading. Twenty-eight advanced EFL learners read 12 paragraphs in English with target areas that consisted of known words, pseudo words or a combination of both. Results showed that readers fixated longer on pseudo words than on known words, regardless of whether these pseudo words were combined with appositive cues. There was a strong association between the total fixation time on pseudo words and subsequent recognition of these words in a surprise posttest.

**The Current Study**

Adopting Godfroid et al.’s (2013) measure of attention, I aim to investigate vocabulary development through repeated encounters by collecting eye-movement records. I will also look at different components of word knowledge, including form and meaning recognition and meaning recall, with the purpose of capturing a multi-faceted picture of the trajectory of lexical acquisition. In other words, I question the relative speed with which the different aspects of
vocabulary knowledge accrue in the process of incidental learning, assuming that faster processes will occur earlier in the cognitive trajectory of incidental lexical development.

One unique feature of the proposed study is that I investigate this issue with a population less commonly represented in literature; i.e., English-speaking learners of Arabic. This may entail certain practical challenges in an eye-movement investigation, not in the least because of differences in script and the direction of reading. Rayner (1998, 2009) reviewed issues in eye tracking and maintains that different text direction entails differences in the perceptual span which is, in the case of Arabic text, asymmetric and larger to the left of fixation. On the readability and accuracy of reading, Al-Wabil and George (2010) reported that L1 readers of Arabic would be more accurate with 14-points and 16-point of simplified fixed Arabic rather than traditional Arabic font type. In general, no previous empirical studies, to the best of my knowledge, have been conducted on Arabic foreign language reading. Experimenting with Arabic text using the eye tracking method is likely to inform further research on the perceived challenges and implications for the learners’ online interaction with a non-roman script during reading.

In the light of the previous discussion, the present research seeks to answer the following questions:

1. How do English speaking learners of Arabic process novel lexical items in Arabic written input? And how do their eye movements change from the first encounter to subsequent encounters with new vocabulary?

2. Do L2 Arabic learners’ eye fixation durations on novel words predict their learning gains in the vocabulary knowledge posttests?

3. What is the effect of repeated exposure to novel target words in L2 Arabic text on the
incidental acquisition of receptive and productive knowledge of form and meaning of these words in vocabulary posttests?

**Method**

**Participants**
Thirty English-speaking learners of Arabic (22 females and 8 males) participated in the study. Fourteen were enrolled in second-year Arabic, and 16 were in their third-year of Arabic, all at an American university. Their ages ranged from 18 to 26 with a mean age of 21 years. Their level of proficiency in Arabic ranged between low and high intermediate as defined by the ACTFL oral proficiency guidelines. This information was obtained from recent OPI exam results, if available, or their teachers’ evaluations of their current proficiency levels.

**Materials**

*Target words*
To control for prior word knowledge, I created a list of 12 pseudo words resembling real Arabic words for the study. The list consisted of 6 nouns, 3 verbs, and 3 adjectives. Word length ranged from 4 to 6 letters and followed the patterns and morphological rules of actual modern standard Arabic words. The pseudo words basically denoted concrete and familiar concepts that lead themselves to being inferred from sentence context. No cognates or pseudo homophones with English were used. To control for the effects of intrinsic word properties, the 12 target words were organized into three counterbalanced word lists. In each word list, four words appeared 4 times, four words appeared 7 times, and the remaining four words appeared 10 times in the reading task. To minimize the effect of visual factors on eye fixations, word lists were matched for the number of letters and part of speech within and across exposure conditions. Table 1 includes the list of pseudo words, given in Arabic script and transliteration, and their respective meanings in the reading context.
Table 1.

*Target pseudo words (in transliteration) for the reading task*

<table>
<thead>
<tr>
<th>Target word</th>
<th>Meaning (in English)</th>
<th>Target word</th>
<th>Meaning (in English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>qasaba</td>
<td>city</td>
<td>sakheem</td>
<td>sad</td>
</tr>
<tr>
<td>shinkeet</td>
<td>poor</td>
<td>buzook</td>
<td>money</td>
</tr>
<tr>
<td>kurduus</td>
<td>castle</td>
<td>baquur</td>
<td>hospital</td>
</tr>
<tr>
<td>eftakasa</td>
<td>to wear</td>
<td>gurana</td>
<td>job</td>
</tr>
<tr>
<td>zaana</td>
<td>to travel</td>
<td>ghasheem</td>
<td>new</td>
</tr>
<tr>
<td>estabdaa</td>
<td>to buy</td>
<td>mafsah</td>
<td>market</td>
</tr>
</tbody>
</table>

*Reading material*

I created simple sentences to integrate the target words for the experiment. The sentences were adapted mainly from the textbooks that students use in the second and third years. To ensure a good level of comprehensibility, I passed the sentences to two teachers of Arabic in the program, and they were asked to revise the structure and vocabulary to make them suitable for their students. I also consulted the teachers (as additional native speakers) on the plausibility of the pseudo words and their suitability in different contexts which represented everyday situations.

The revised reading material consisted of 100 sentences. Seventy-five percent of the sentences were followed by a yes/no question to check for comprehension and direct learners to focus on meaning. The reading set included 6 practice sentences, 84 critical sentences and 10 fillers. The texts were randomized for all participants so that encounters with a given target word were distributed randomly across the reading set. Each sentence ranged in length between 15 and 24 words. I created three counterbalanced versions of the reading material so that each set of four
words occurred with a different frequency across conditions. Table 2 shows an outline of how
target vocabulary items were distributed across subjects and conditions. Samples of the reading
material are available in Appendix A.

Table 2

*Presentation of target vocabulary in sentence sets*

<table>
<thead>
<tr>
<th>Exposure</th>
<th>4 exposures</th>
<th>7 exposures</th>
<th>10 exposures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>kasaba</td>
<td>shinkeet</td>
<td>sakheem</td>
</tr>
<tr>
<td>S1</td>
<td>ghashim</td>
<td>mafsah</td>
<td>kardoos</td>
</tr>
<tr>
<td>S4</td>
<td>zana</td>
<td>iftakasa</td>
<td>gurana</td>
</tr>
<tr>
<td>…</td>
<td>istabdhaa</td>
<td>bakuur</td>
<td>buzook</td>
</tr>
<tr>
<td>S28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>shinkeet</td>
<td>bakheem</td>
<td>kasaba</td>
</tr>
<tr>
<td>S2</td>
<td>mafsah</td>
<td>kardoos</td>
<td>ghasheem</td>
</tr>
<tr>
<td>S5</td>
<td>iftakasa</td>
<td>gurana</td>
<td>zaana</td>
</tr>
<tr>
<td>…</td>
<td>bakuur</td>
<td>buzook</td>
<td>istabdhaa</td>
</tr>
<tr>
<td>S29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group III</td>
<td>sakheem</td>
<td>kasaba</td>
<td>shinkeet</td>
</tr>
<tr>
<td>S3</td>
<td>kardoos</td>
<td>ghasheem</td>
<td>mafsah</td>
</tr>
<tr>
<td>S6</td>
<td>gurana</td>
<td>zaana</td>
<td>iftakasa</td>
</tr>
<tr>
<td>…</td>
<td>buzook</td>
<td>istabdhaa</td>
<td>bakuur</td>
</tr>
<tr>
<td>S30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Procedures**

I invited students from second and third year Arabic classes to take part in a study about reading
Arabic on a computer screen. Based on the availability of students and the lab schedule, I met
with each participant individually in an eye tracking lab managed by the second language studies
program in the university. Before a participant began the session, he received instructions on the
procedures of the experiment and directions on how to use the hand controller. Learners were
instructed to read for meaning. They were not made aware that this was a study about vocabulary
acquisition or the fact that they would encounter pseudo words in context and receive a test
afterwards. After filling in a background questionnaire and signing a consent form, each
participant was randomly assigned to read only one version of the reading material.
Apparatus

Eye movements were recorded with a desk-mounted EyeLink 1000, an eye-tracker manufactured by SR Research (http://www.sr-research.com/). The texts were presented in simplified Arabic font, size 18, on a 19-inch computer monitor set up 55 cm from the participants’ eyes. Participants placed their heads on a chin and forehead rest during the experiment to minimize head movements. The experiment included two breaks, once after a block of 40 sentences and next after sentence 70. Eye calibration was performed at the beginning of the experiment and twice after the breaks. Participants moved across screens using a button on the right side of a hand-held controller, and answered questions by pressing an upper left button to respond with ‘yes’ and an upper right button to respond with ‘no.’ Drift correction was performed before each sentence. Because Arabic text goes from right to left, the experiment options were adjusted to assign interest areas from right to left and to define eye-movement measures accordingly. The drift correction was also placed at the right of the screen.

Testing

The testing session started immediately after the eye tracking experiment with the purpose of obtaining paper-and-pencil data on the quantity and quality of vocabulary learning from reading. To obtain a multi-faceted picture of lexical knowledge development, it was important to include multiple measures of vocabulary knowledge. All the vocabulary tests consisted of the target 12 items of the study in addition to distractor items. Details of the tests are given below in the order they were conducted. Tests were specifically given in this order to avoid any testing-transfer effects from one vocabulary measure to the other.

Cued production test. This aspect of word knowledge entails that the learner can use a target word productively in context. To test this ability, a fill-in-the-gap test was designed
in which participants were required to provide one missing word within each of the provided sentences. Full sentences were taken from the online reading material but with the target word deleted from context. The initial two letters of the missing word in each sentence were given as a cue to help learners recall the whole lexical item. Samples of the sentences are given in Appendix B.

**Form recognition test.** Form recognition is assumed to be the first aspect to develop in learners’ knowledge of novel lexical items. It entails a participant’s ability to recognize the form of the word and determine if he or she has seen it before. Participants were given 12 sets of words, each containing four non-words, one of which was the target word. The task was to identify the target lexical item that occurred in the reading material among three other distractors. Students were instructed to circle only the word they recognized in each set. A participant received 1 point for a correct choice and zero for an incorrect response. Examples from the test are given in Appendix C.

**Meaning recall test.** Meaning recall entails learners’ ability to have some representation of what a new word means and be able to retrieve that meaning when prompted. To examine this, a simple Arabic-English translation test was designed in which participants were required to either provide a meaning or an explanation for the target word or just provide any clue that they thought could be relevant to its meaning. Participants received 1 point for a correct meaning, close synonym, or explanation and zero for a blank or irrelevant response. A sample form the test is given in Appendix D.

**Meaning recognition test.** This aspect of word knowledge requires less mental effort than meaning recall as learners are only required to recognize and match the meaning to the word. To test this ability, a multiple-choice task was designed in which each target item was
given along with four possible meaning options in English. The participants were required to circle only one meaning. Participants received 1 point for the correct choice and zero for an incorrect response. Examples from the test are given in Appendix E.

**Analyses**

Eye-movement analyses of the reading data included reports on first fixation durations, gaze durations, and total reading times of the target words on all exposures. I compared these three measures for all participants to capture the pattern of looking at a given word from the first to the last encounter. Several steps were taken to code and classify eye-fixation data. First, I compiled a list that included the fixation times by participant to each of the target words on every exposure. Second, I averaged fixation times across participants, which resulted in looking times for each word at every encounter. Then, data were averaged over exposures providing fixation times for each target encounter regardless of item. Mean eye-movement times were then compared through line graphs to detect significant rises and declines in reading times.

To associate the online data with vocabulary learning measures, I organized the eye movement data differently by summing all reading times (first fixations, gaze durations, total times) on all exposures of each target word for each participant. In this way, the records would reflect accumulated episodes of attention for specific target items and relate these to either failure or success in retrieving a given word in the different vocabulary tests. Because the data consisted of multiple observations per participant, I used a generalized estimation equation (GEE) to fit a repeated measures logistic regression, which is appropriate because the vocabulary posttests represent binary dependent variables. The purpose of the analysis is to test if the fixation durations and/or exposure condition predict the probability of form and meaning recognition on the posttest. The GEE output quantifies probability in terms of an odds ratio (exp [B]) which
represents the predicted change in the dependent measure as a function of a one unit change in a given predictor. Finally for the offline data, performance on the four vocabulary tests was compared holistically between tests and by exposure condition to explore the typical trajectory of incidental lexical knowledge acquisition.

Results

Online Reading Measures

*First fixation durations (FFD)*

Mean first fixation times were tracked across exposures, averaged over participants and vocabulary items. Fixation durations started at an average of 403 ms ($SD = 19$) for first encounter and ended with an average of 375 ms ($SD = 26$) by the last exposure. The line graph created for first fixation durations in Figure 1 shows plateau and increase in FFD values on initial exposures with a small rise at exposure 6 before it declines to reach its lowest by exposure 10.

Figure 1

*Mean first fixation durations (in milliseconds) on encounters with target words*
Gaze durations (GD)

Gaze durations started at an average of 2148 ms (SD= 495) for first encounter and ended with an average of 1164 ms (SD = 489) by the last exposure. Unlike the FFD graph, the gaze duration plot showed a steadily decreasing pattern with the longest time spent on exposure 1 and the least on exposure 10 (see Figure 2). Visual inspection of the graph shows that gaze durations plateaued from the fifth to the eighth exposure before declining again until the final target encounter.

Figure 2

Mean gaze durations (in milliseconds) on encounters with target words

Total reading time

The total reading times followed a gradually decreasing pattern, alike to gaze duration, with the longest time spent on the target word at first exposure and the shortest time spent at the last exposure (see Figure 3). Reading times started from an average of 3340 ms (SD= 958) and ended with an average of 1612 ms (SD= 545). Visually, a plateau in total times can be observed from the sixth to the ninth exposure before declining at the last target encounter.
Considerable individual variations were noted as shown in Figure 4 where most readers seem to cluster in one area with some rises and declines of times at the exposures from 5 to 7 and particularly sharp increases at exposures 8 and 9, which again reflect some of the individual variations and contribute to the overall plateau shown in the graph of average reading times.
Vocabulary Learning Measures

To analyze the results of the paper-and-pencil tests, I looked at the chances of acquisition after a given number of exposures; i.e. what percentage of words individual readers could learn from having seen the word 4, 7 or 10 times. Individuals considerably varied in their word gains, particularly in meaning recall, as some could not recall any words and some others were able to recall accurate meanings of almost up to 90% of the target words. An inspection of the percentage of learning gains for the different paper-and-pencil vocabulary tests showed consistently increased chances of learning as target word encounters increased (see Table 3). For example, in meaning recall there was a 9% chance of correct recall when the word was repeated four times, and this chance increased to 25% after the seventh exposure, and reached 33% when the target items were seen ten times. On the other hand, learners could correctly recognize the form of words 74% of the time after they had encountered them 10 times in reading. Table 3
summarizes the percentages of word gains over exposure conditions and how they differed across different vocabulary tests.

Table 3

*Percentages of word gains (%) with standard deviations by exposure condition*

<table>
<thead>
<tr>
<th>Exposure Condition</th>
<th>Form recognition</th>
<th>Meaning recognition</th>
<th>Cued production</th>
<th>Meaning recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 exposures</td>
<td>50 (27)</td>
<td>41 (36)</td>
<td>13 (21)</td>
<td>9 (13)</td>
</tr>
<tr>
<td>7 exposures</td>
<td>69 (28)</td>
<td>52 (27)</td>
<td>28 (33)</td>
<td>25 (21)</td>
</tr>
<tr>
<td>10 exposures</td>
<td>74 (26)</td>
<td>67 (26)</td>
<td>39 (28)</td>
<td>33 (23)</td>
</tr>
<tr>
<td>Average by test</td>
<td>64</td>
<td>53</td>
<td>26</td>
<td>22</td>
</tr>
</tbody>
</table>

From Table 3, it is clear that learning gains developed at a high level for form and meaning recognition, but to a smaller extent for meaning recall and form production, which means that receptive knowledge, unsurprisingly, preceded productive knowledge development. Some large standard deviations reflect individual variation in vocabulary acquisition. The retrieval success ranged from a minimum of 0 % to a maximum 50 % under 4 exposures in all measures except form recognition, where some participants retrieved all word forms after 4 exposures. Under 7 and 10 exposures, retrieval ranged between 0 % and 100 % in all measures. Figure 5 summarizes learning gains across vocabulary tests and by exposure condition.
Relationship between Online Reading and Vocabulary Learning

The major question in regards to the eye-movement data was to see if the probability of word learning would increase as learners spent more total time processing (looking at) the words and whether the predictive power of online reading times would be upheld after factoring in exposures. To this end, I ran a GEE logistic regression for each of the vocabulary tests as outcome variables with subjects as a random factor and target items as a repeated within-subject variable. Frequency of exposure was entered as a factor with two dummy variables for 7 and 10 exposures respectively. Each reading measure was entered separately as a covariate to compare their effect against exposure frequency. In each step, two models were compared: a model with exposure conditions only and a model with exposure plus one reading time covariate. The two models were compared in terms of goodness of fit (-2LL). A decrease in the values of (-2LL) would indicate that a model of exposure frequency plus reading times can explain more variance.
than a model with only exposure frequency. The difference values between each pair of models were analyzed against a chi-square and degrees of freedom distribution. The effect of a reading time was only reported when (-2LL) differences were at least at .05 significance.

In the light of model fit analysis, online reading measures showed mild positive effects to variable extents. First fixation duration significantly predicted form recognition scores, $\text{Wald } \chi^2(1) = 19.3, p < .001, \exp (B) = 1.45$, and meaning recall scores, $\text{Wald } \chi^2(1) = 16.52, p < .001, \exp (B) = 1.4$. Gaze duration predicted meaning recognition scores: $\text{Wald } \chi^2(1) = 12.79, p < .001, \exp (B) = 1.7$, and meaning recall, $\text{Wald } \chi^2(1) = 4.17, p = .041, \exp (B) = 1.04$. The total time predicted cued production, $\text{Wald } \chi^2(1) = 18.026, p = .002, \exp (B) = 1.08$, meaning recognition, $\text{Wald } \chi^2(1) = 16.59, p < .001, \exp (B) = 1.58$, and meaning recall, $\text{Wald } \chi^2(1) = 3.98, p = .046, \exp (B) = 1.24$. The odds ratios ($\exp [B]$) in the model output showed that exposure frequency were a stronger predictor than summed total reading times on all the outcome variables. For example, meaning recall was moderately enhanced by seeing a word 7 times ($\text{Wald } \chi^2(1) = 7.029, p = .007, \exp (B) = 2.65$, and 10 times ($\text{Wald } \chi^2(1) = 22.17, p < .001, \exp (B) = 6.95$) rather than 4 times. In other words, increasing the number of exposures to 7 increased the chances of meaning recall by 2.65 times (medium effect) and seeing the word 10 times increased successful meaning recall by almost 7 times (large effect) compared to 4 exposures. Table 4 models all vocabulary outcomes as a function of online reading measures and exposure conditions and listing significant Wald Chi-square values and odds ratios termed $\exp (B)$. 


Table 4

Vocabulary knowledge predictors represented in chi-square values and odds ratios

<table>
<thead>
<tr>
<th></th>
<th>Exposure 7</th>
<th>Exposure 10</th>
<th>First fixation</th>
<th>Gaze</th>
<th>Total time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$ (exp B)</td>
<td>$\chi^2$ (exp B)</td>
<td>$\chi^2$ (exp B)</td>
<td>$\chi^2$ (exp B)</td>
<td>$\chi^2$ (exp B)</td>
</tr>
<tr>
<td>Form recognition</td>
<td>7.28 (2.27)</td>
<td>8.87 (3.11)</td>
<td>19.3 (1.45)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Meaning recognition</td>
<td>NS</td>
<td>4.72 (2.19)</td>
<td>NS</td>
<td>12.79 (1.76)</td>
<td>16.16 (1.58)</td>
</tr>
<tr>
<td>Meaning recall</td>
<td>7.029 (2.65)</td>
<td>22.17 (6.95)</td>
<td>NS</td>
<td>4.17 (1.04)</td>
<td>3.98 (1.24)</td>
</tr>
<tr>
<td>Cued production</td>
<td>21.33 (2.17)</td>
<td>23.33 (2.60)</td>
<td>NS</td>
<td>NS</td>
<td>18.26 (1.08)</td>
</tr>
</tbody>
</table>

Table 4, which reports only the significant predictors in the model, shows that some online reading measures had a small positive effect particularly on meaning recognition and recall, but the stronger predictor on all outcome variables was the number of exposures.

To summarize the results, online and offline measures confirmed the effect of repeated exposure on incidental learning gains for different vocabulary measures. Eye-tracking data showed that learners fixated longer on initial exposures of novel words and that gaze durations and total time gradually decreased, hitting a plateau in the middle before they declined once again to reach their lowest at the final exposure. First fixation times showed an increase from first to fifth exposure before they gradually declined until final encounters. A relationship between online processing patterns and learning gains was established as total reading times predicted meaning recognition, meaning recall and cued production; gaze duration predicted meaning recognition and recall while first fixation duration only predicted form recognition of target words. Eye movement results in general indicated that different fixation measures predicted different aspects of word knowledge. However, the frequency of exposure was a
stronger predictor than online processing times of target words. Offline vocabulary measures confirmed the advantage of repeated exposure, showing consistently higher chances of learning with more frequent exposures in all vocabulary tests. The overall picture of incidental vocabulary gains showed discrepancy across measures in that performance was highest in form recognition and lowest in meaning recall, pointing to a potential cognitive trajectory of second language lexical development.

**Discussion**

The aim of the present study was to disentangle the effects of two factors known to facilitate vocabulary learning: attention and repeated exposures. Evidence for the association between online processing and offline vocabulary learning was provided by Godfroid and colleagues (2013) for single exposures to pseudo words. Webb (2007) explored the effect of repetition on word knowledge. The current study builds on both directions of research by examining the joint effects of repetition and the amount of online processing and exploring their association with learning gains in different vocabulary knowledge measures.

The first question concerned the processing patterns of novel lexical items by L2 Arabic learners in sentence reading. The common observation about the fixation times is that learners recorded the longest fixations on first encounter and the shortest times on the last encounter of the target item in both gaze duration and total times. This result is in line with previous eye tracking studies that probed into the effect of frequency and repetition (e.g. Hyönä & Niemi, 1990; Raney & Rayner, 1995). Learners paid more attention to the pseudo words as low frequency items, but as they got repeated they may have gained frequency properties that mediated attention and made it easier for readers to retrieve them or reach a satisfactory level of word familiarity. At the level of individual participants, some readers showed a sharp rise in
fixation times by exposures 6 and 7 and then a gradual decline until the last exposure. Perhaps this could be explained as potential “noticing” of novel exemplars (Godfroid et al., 2013; Schmidt, 1990); it may mark a peak of curiosity on the part of the learner that could promote successful meaning acquisition, after which the familiarity with the word causes reading times to decline. However, as stated above, this trend was not generalizable across the whole sample. What can be generalized, though, is that the pattern of online reading across participants clearly showed increased reading fluency of the target words along exposures. Further research would benefit from tapping into individual differences in reading behavior through self-reports or interviews that could probe into learners’ subjective experiences while processing the target words (Godfroid & Schmidtke, 2013). Stimulated recall or self-reports can probably explain instances where individual readers’ fixation times peaked at a specific exposure or even point out instances where the participant was aware that he or she was paying attention to some item or that he or she has guessed a meaning or learnt another aspect of a new word. Interviews can also provide accounts of learners’ awareness of the existence of novel words and their repetition and the strategies they used in interacting with the text.

The second question sought to interpret learning gains in different vocabulary measures in terms of their potential relationship with the reading patterns of target words. Results from the present study largely supported the assumption tested by Godfroid et al. (2013) that more attention to target words, quantified as the summed total reading time spent on the lexical item, predicted how well learners demonstrated some knowledge of form and meaning of the item in the vocabulary posttest. The kind of associations found between fixation times and different types of vocabulary gain aligns with the claim that different eye movement measures reflect different cognitive processes. Within the framework of the E-Z reader model (Reichle,
Rayner, & Pollatsek, 2003; Pollatsek, Reichle, & Rayner, 2006), lexical processing has been posited to proceed in two stages: an early stage which the authors called ‘familiarity check’, and a later stage referred to as the completion of lexical access. The fact that only first fixation durations predicted form recognition conforms to this hypothesis that early lexical processing is largely form-focused. Gaze duration, as a somewhat later measure, predicted meaning recognition and recall, which may indicate that subsequent lexical processing of form-meaning mapping and encoding into memory becomes more important. The same principle would explain why total time, as a late measure, predicted productive recall as a type of higher-order lexical knowledge. The reading data in the present study provides a platform for further inquiry into how different eye movement measures actually point to different cognitive processes. By including a test battery in future studies, eye tracking researchers may be able to explore these associations in greater depth.

The third question in the study concerned the cognitive trajectory of lexical development; i.e. in what order the components of word knowledge emerge in the learners’ lexicon as a function of repeated exposure. In line with previous literature on vocabulary acquisition (Nation, 2001; Schmitt, 2008, 2010), receptive knowledge of form seemed to be the first component to develop followed by meaning recognition and finally cued production and meaning recall. These differential learning rates can be explained in terms of progression from lower-order to higher-order processes. They can also be related to the incremental nature of lexical development in that larger reading times initially contributed to building a sound knowledge of form and meaning while there was still more need for additional reading and exposures to establish full knowledge of a given target word. In form recognition, the learner only needs to map a word form onto the memory traces of the orthographic form of the target
word while in meaning recall the learner needed to have enough chances to infer meanings correctly and subsequently decontextualize words and retain them in memory, which is evidently a more complex task. In general, the overall picture of vocabulary knowledge development in the current study shows predictable patterns in line with previous research. However, further research could provide more insight into how these knowledge measures relate to the quality of context and L2 readers’ meaning inference abilities.

One interesting finding regarding the distinction between online and offline measures was that the number of exposures per se strongly predicted the learning outcomes above and beyond the online reading measures. One explanation to consider is the amount of variance that was shown across readers in their processing times as well as their vocabulary knowledge gains. Although it is understandable that more exposure opportunities entailed larger reading times, there still existed variations among participants in that some would see a target word for seven times but pay less attention to these encounters than those who saw the same words four times. This may invite investigation of individual differences in reading comprehension and cognitive processes in incidental learning. These individual variations can include features related to both the amount and quality of attention, which in turn could be attributable to different individual factors such as vocabulary size, working memory, motivation or overall language proficiency. Further eye movement research would benefit from including more participants and more reading material to uncover more robust associations.

Some methodological issues need to be discussed regarding the nature of tasks and participants in the present study. The first point concerns the use of pseudo words for the study. As learners were normally expected to know the real words for the target items (e.g. city, wear or sad), they may have concluded that the novel words they encountered in reading were less
frequent synonyms of the words they already knew, an impression that may have reduced the motivation or cognitive effort to incorporate the new lexical items. Moreover, the lab-controlled experiment condensed the number of exposures into one experimental session, which may not exactly match the typical incremental route that learners go through in incidental learning where repeated exposures are spaced over longer periods of time. Finally, the reading measures of learners of Arabic and the amount and quality of their learning gains from repeated exposure may have presented a language-specific picture or population-specific trends. I recommend that further studies experiment with other languages and other populations to see whether the present results are generalizable.

**Teaching implications**

The results of the study are important for teachers and practitioners in the field to help them make informed choices about their material presentation and implement effective methods to enhance learners’ exposure to vocabulary in different contexts. When implementing reading tasks in the classroom with a vocabulary learning goal, teachers should take into account the incremental nature of lexical development by manipulating incidental, yet planned, exposure opportunities to specific vocabulary items that have been explicitly covered elsewhere across different reading material over a considerable time span. On the other side of the continuum, teaching should also consider the potential of incidentally acquiring less frequent vocabulary as a by-product of recycling class activities.

The results showed that learners’ knowledge in different vocabulary aspects accrued as a function of successive encounters with lexical items. Teachers should respond to this need for recycling in their explicit vocabulary teaching practice and testing instruments. Furthermore, the present study looked at vocabulary exposure within written input only and under silent
reading conditions. Teachers should also consider reading as one type of input that can still be boosted through the other modalities that can afford additional exposure opportunities to vocabulary under variable contexts.

Vocabulary reviews have shown that much is involved in knowing a word and being able to retrieve the meaning of a given L2 word is just one aspect of this knowledge (Nation, 2001; Schmitt, 2008). Lexical gain results in the present study corroborate this principle and relate it to multiple exposures. Teachers should consider this fact in their testing material so as to accurately gauge different levels of their students’ lexical knowledge.

**Conclusion**

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

**References**


<table>
<thead>
<tr>
<th>Identifier</th>
<th>sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital1</td>
<td>قال الطبيب أن وليد أحسن الآن وليس مريضا جدا ويمكن أن يترك الراوتر ويعود للبيت ويسجل هناك</td>
</tr>
<tr>
<td>Castle1</td>
<td>يملك الأمير ويمد بن خلال كروس عاليًا في لندن يبلغ تكلفة أكثر من 50 مليون دولار</td>
</tr>
<tr>
<td>Buy 1</td>
<td>ذهب خالد إلى محل للملابس واستنتج بصلة جديدة بـ 250 دولار ولكن أنه لم يأكل بشيء غريب</td>
</tr>
<tr>
<td>Money1</td>
<td>دُمِت أختي إيه السوق لنفس ملابس العيد ولكنها كانت تريز بيروف أكثر لأن ملابس عادية</td>
</tr>
<tr>
<td>Wear 1</td>
<td>في كل يوم يكتسب حديد ملابس عمل ويذهب إلى المشي ليقضي الوقت مع أصدقائه وشرب لفترة</td>
</tr>
<tr>
<td>Travel1</td>
<td>بعد أن استلمت من رحلتي الهولندية العلوم السياسية جالت جامعتي ممتازة للسفن إلى أمريكا واستكمال دراستي في ماجستير ودكتوراه ولكن فضحت ليقين في يدي</td>
</tr>
<tr>
<td>Market 1</td>
<td>يا سعيد هل تحب أن تأتي معى إلى المشفى لنشرى ملاتس العيد ؟ أريك لن تساهم في لجيك بعض الأشياء</td>
</tr>
<tr>
<td>Money2</td>
<td>عندى بيروف كثير ولكن لا يعرف ماذا تشتري لصديقتي في عيد الحب سوف أسألها ماذا تريز</td>
</tr>
<tr>
<td>Filler 1</td>
<td>قال لي الأستاذ أن الواجب جيد ولكنه يريد أن نكتب على الكمبيوتر</td>
</tr>
<tr>
<td>Poor 1</td>
<td>لن يستطيع محمد أن يقرر الزواج الآن لأن كل الناس تعرف أنه ملتقي وليس عند شغله</td>
</tr>
<tr>
<td>Buy 2</td>
<td>جون يريد أن يستثمر لاب توب وكمبيوتر بمنصة الكمبيوتر ولكن أبوه ليس معه نسخ كلغة</td>
</tr>
<tr>
<td>Travel 2</td>
<td>ولدة صديقي تطورت دائما إلى دول مختلفة في أوروبا وتحدث عن أبحاثها في مؤتمرات كبيرة</td>
</tr>
<tr>
<td>Poor 2</td>
<td>سارة تريز أن تعيش حياة سعيدة ولكن زوجها جون رجل ملتقي لا يأخذ طلع كثيرة من عمله</td>
</tr>
<tr>
<td>Wear 2</td>
<td>يجب على المرأة المسلمة أن تلتقي ملابس مناسبة عندما تخرج من البيت</td>
</tr>
<tr>
<td>Job 1</td>
<td>سافر سمير إلى سعودية للبحث عن فتاة فتاة ذات شكل جردي يمكن أن يكون في أهله ولده وهو في منزله</td>
</tr>
<tr>
<td>Money 2</td>
<td>يريد مي شراء سيارة جديدة ولكنه ليس عند للبيروف لنفس الكلفية لذلك قرر أن يستأجر سيارة</td>
</tr>
<tr>
<td>Sad 1</td>
<td>كان محمود مشغولًا لأيام كثيرة بعد أن صممًا أهله ماتت في المستشفى وهو خارج بلد</td>
</tr>
<tr>
<td>Filler 2</td>
<td>في يوم الإجازة كان أعمل حفلة في البيت ولعبت مع أصدقائي كل وقت</td>
</tr>
</tbody>
</table>
Appendix B: Snapshot of the cued production test

Participant ..........................  
You are given the first letters of some words you have seen in your online reading. Please these cues to recall the words. Write them as complete as you can

1-ذهب خادم محل الملابس واس... تكلفة جديدة بـ 200 دولار ولكنه أخوه قال إنها غالية جدا

2-رئيس أو أمريك دائما يعيش في ك... كبير وعالي ولا يعرف كثيرا عن مشاكل الناس

3-سيرة نوح الظ... إلى نوريدا في الصيف حيث تسمع بالجو المعتدل والمناظر الجميلة على البحر

4-احتاج أن أشترى سيارة غ... لأن سياري أصبحت قديمة جدا ولاتعمل جيدا في الشتاء

5-إذا أردت أن تذهب إلى حمام السباحة معها فيجب أن نف... ملابس السباحة ونظارة

Appendix C: Form recognition test

Participant ID .......................... Form recognition (post 1)  
From each row, circle the word you recognize you as quickly as you can

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>بصة</td>
<td>غشيم</td>
<td>كفتس</td>
<td>قصيل</td>
</tr>
<tr>
<td></td>
<td>قصبة</td>
<td>شغيم</td>
<td>استبضع</td>
<td>فصنة</td>
</tr>
<tr>
<td></td>
<td></td>
<td>المشغ</td>
<td>كستف</td>
<td>ضعبس</td>
</tr>
</tbody>
</table>
Appendix D: Meaning recall test

II-Meaning recall (post 1)

Participant .....................

Here is a list of words you probably have seen in the online reading. Try to recall a meaning or explanation or any comment you may have on them.

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Meaning or explanation or something related</th>
</tr>
</thead>
<tbody>
<tr>
<td>غشيم</td>
<td></td>
</tr>
<tr>
<td>بزوق</td>
<td></td>
</tr>
<tr>
<td>جعرانة</td>
<td></td>
</tr>
<tr>
<td>سخيم</td>
<td></td>
</tr>
<tr>
<td>كردوس</td>
<td></td>
</tr>
<tr>
<td>استبضع</td>
<td></td>
</tr>
</tbody>
</table>

Appendix E: Meaning recognition test

III-Meaning recognition (post 1)  participant .....................

Choose the correct meaning

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
<th>English</th>
<th>English</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>بزوق</td>
<td>method</td>
<td>tears</td>
<td>money</td>
<td>time</td>
</tr>
<tr>
<td>مقصح</td>
<td>basket</td>
<td>market</td>
<td>apartment</td>
<td>hill</td>
</tr>
<tr>
<td>غشيم</td>
<td>huge</td>
<td>rich</td>
<td>old</td>
<td>smile</td>
</tr>
<tr>
<td>سخيم</td>
<td>joyful</td>
<td>ugly</td>
<td>sad</td>
<td>smile</td>
</tr>
<tr>
<td>فصلة</td>
<td>city</td>
<td>house</td>
<td>train</td>
<td>tent</td>
</tr>
<tr>
<td>نظيف</td>
<td>travel</td>
<td>dry</td>
<td>filter</td>
<td>sing</td>
</tr>
<tr>
<td>كردوس</td>
<td>sky</td>
<td>hotel</td>
<td>castle</td>
<td>jar</td>
</tr>
</tbody>
</table>