Castledown

OPEN ACCESS

Language Education & Assessment
ISSN 2209-3591
https://www.castledown.com/journals/lea/

Language Education & Assessment, 3 (1), 1–12 (2020)
https://doi.org/10.29140/lea.v3n1.171

Assessing Threshold Level of L2 Vocabulary Depth in Reading Comprehension and Incidental Vocabulary Learning

S.J. EHSANZADEH [a]

[a] *Florida International University, USA*
sehsa002@fiu.edu

Abstract

This study measures the threshold level of depth of word knowledge in L2 incidental learning and reading comprehension. I administered the Word Associates Test (WAT; Read, 2004) to measure depth of word knowledge. Based on participants' scores on the WAT, they were divided into three proficiency groups: high proficiency (HP), intermediate proficiency (IP), and low proficiency (LP). As part of data collection, participants were also asked to read a short story and then complete items of the Vocabulary Knowledge Scale (VKS; Paribakht & Wesche, 1997) for the 30 chosen target words. To rate incidental vocabulary acquisition, I administered the VKS test with a repeated measure design in a period of one month. Results indicate that (a) the HPs gained more vocabulary by comparison than the IPs and the LPs, (b) the probabilistic threshold level for L2 incidental learning is from 85 to 135 out of 160 on WAT, and (c) no significant role of depth in reading comprehension has been confirmed.

Keywords: Depth of word knowledge, L2 incidental learning, reading comprehension, and threshold level

Introduction

Wilkins (1972) states that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). With respect to the essential role of lexical repertoire in functioning appropriately in a second language (L2), Schmitt (2010) also thinks that word knowledge contributes significantly to every aspect of L2 proficiency. Traditionally, knowing a word means knowing its form and meaning. In the same way, Schmitt (2008) asserts that "many teachers and learners consider a word 'learned' if the spoken/written form and meaning are known" (p. 333).

However, Nation (2013) believes that aspects of word knowledge are more than form and meaning, and word knowledge could be questioned if only form and meaning are known. Webb (2013) argues that, in addition to having a fair knowledge of vocabulary size or breadth of word knowledge, an

Data Availability Statement: *All relevant data are within this paper.*

English language learner (ELL) should know enough about every word to be able to function appropriately in an L2. These other aspects are technically called depth of word knowledge (Read, 2007) and are as needed as vocabulary size for an ELL (Nation, 2013). In other words, breadth of knowledge is the quantity of number of words, and depth refers to the quality or how well a word is known (Nation, 2013). For instance, an ELL may know many words in English, but he/she may not be able to use these words in accurate contexts or to associate them correctly with other words. Thus, researchers (Ehsanzadeh, 2012; Nassaji, 2004; Qian, 2002; Zhang & Koda, 2017) claim that a distinction between depth and breadth is beneficial, especially for teaching vocabulary and reading to ELLs.

The line of research on reading and vocabulary in L2 has not given the same amount of attention to depth and breadth dimensions despite the fact that both are important in learning an L2, particularly in learning vocabulary and improving reading comprehension at advanced levels. In order to develop a richer network of word knowledge that ELLs could utilize to further improve their word knowledge through reading, Webb (2013) argues that teachers should be made aware of the role of depth and how it can be used in developing vocabulary through reading. A lack of rich depth of knowledge may cause unsuccessful use of words in contexts, a problem commonly seen in the language of many ELLs. Thus, students should know what it means by knowing a word if they are expected to effectively use them in different contexts. Also, Nation and Webb (2011) argue that depth helps ESL teachers diagnose ELLs' problems in reading and incidental vocabulary learning.

Also, Milton (2009) thinks that depth of word knowledge is less researched and understood as well. In the meantime, threshold level studies (see below for definition) mainly focused on vocabulary size and its effect on language learning (Nation, 2001, 2013). I believe that there is no well-researched study on the threshold level of depth of word knowledge in reading and incidental vocabulary learning, and a study like this that investigates the threshold level of depth in reading and incidental learning is needed.

Threshold level studies

The question regarding whether there is a minimum amount of vocabulary knowledge that marks a make-or-break boundary for whether one achieves successful comprehension has always been an interesting issue for reading and vocabulary researchers. The boundary is referred to as a *threshold level*. Nation (2001) believes that a threshold level can be defined in either of these two ways: The first view is the strong view, and Nation argues that threshold level is an either-or (all or nothing) level; in other words, the learner will understand the text if he passes the threshold level; otherwise, he will have a vague comprehension of the text. The second view, however, takes the threshold level as a "probabilistic boundary" (p. 238) and states that if a learner passes the threshold level, then he stands a good chance for a successful comprehension (Nation, 2013). Laufer and Sim (1985a) took the second view of threshold level and attempted to examine it with interviews and comprehension questions. Results of their research show that the threshold level is roughly 65 to 70% of vocabulary size. Laufer (1989b) later examined the needed percentage of vocabulary for successful comprehension. Results of her study show that a score of 95% and above on vocabulary size will likely lead to the full comprehension of a text. Hu and Nation (2000) believe that the coverage of 80% of the text is the threshold level for the all-or-nothing view (the ELL will be on the verge of losing the meaning of the text) and 98% for the probabilistic threshold level. They further claim that 95% is the threshold level for a minimum of adequate comprehension in which the ELL will comprehend the basics of the text. In a recent study, Schmitt, Jiang, and Grabe (2011) believe that there should be a positive relationship between reading scores and text coverage; however, they doubt the existence of a confirmed threshold level. Although precursory research is not clear about the existence of a numerical threshold level and

the minimum number of needed vocabulary size, the existence of a probabilistic threshold level of depth of word knowledge has not been well researched.

Incidental L2 vocabulary acquisition through reading

Incidental vocabulary acquisition is the learning that occurs without conscious focus on the object of the vocabulary learning. Ender (2014) argues that incidental learning is a by-product of other cognitive exercises involving comprehension. On a general level, L2 incidental learning is the unintentional picking up of information when learners are not assessed by an upcoming retention test (Hulstijn, 2013). R. Ellis and Shintani (2014) also believe that incidental learning can be investigated by exposing learners to input data, which they are asked to process for meaning, and then investigating (without warning) whether they have acquired any L2 linguistic properties as a result of the exposure.

Depth of vocabulary knowledge in reading comprehension and incidental word learning

The line of research on the relationship between depth of word knowledge and incidental vocabulary acquisition through reading is scant. In two empirical studies, Qian (1999, 2002) investigated the intercorrelations among depth of word knowledge, breadth of knowledge, and reading comprehension. He found that scores on the vocabulary size, depth of word knowledge, and reading comprehension were highly correlated. The results of his study revealed that depth of word knowledge made a unique contribution to reading comprehension. In another study, Nassaji (2004) provides evidence concerning the central role of depth of knowledge in lexical inferencing through reading comprehension and supports the findings of previous research that "L2 learners need good vocabulary knowledge to be able to successfully derive word meanings from context" (p. 126). Ehsanzadeh (2012) also compared the role of depth and breadth in lexical inferencing success. He thinks that inferential success is the starting point of incidental learning. Results of his study show that both dimensions have a share in lexical inferencing success; however, he claims that depth's share is more than what we thought to be.

Having considered the significant role of depth of word knowledge in L2 incidental learning through reading, one may consider whether there is a threshold level of depth of word knowledge that must be crossed to guarantee adequate text comprehension and incidental vocabulary learning. The present research attempts to focus on depth of word knowledge and its threshold level in incidental vocabulary learning and reading comprehension. I addressed the following research question:

> *What is the threshold level of depth of word knowledge assessed through the Word Associates Test (Read, 2004) in reading comprehension and incidental vocabulary learning?*

Method

Participants

Participants were 120 undergraduate university students who were recruited from a larger population at an English language college, Iran. Participants were four-year university students in English literature. Based on their GPA of the last three years of studying English literature and their scores on general English proficiency courses that they had passed at the university, they were considered advanced language learners. Of the participants, 60 were female and 60 were male. All the participants were native Persian speakers.

Assessing L2 vocabulary depth

The Word Associates Test (WAT), developed by Read (1993, 2004), and a think-aloud retrospection

protocol were the measures used to assess the participants' depth of word knowledge in the present study. According to Read (2007), the WAT measures in-depth word knowledge by means of word associates. The test contains 40 items, and each comprises one adjective as stimulus word followed by eight options in two boxes, four of which associate paradigmatically or syntagmatically with the stimulus word. The WAT is not a multiple-choice test type wherein each item should have only one correct answer; the test is different from normal multiple-choice tests in that each item of the test always has four correct answers. However, these four answers are not evenly spread between the boxes. Asking learners to choose the associates of a word from a given set of choices, rather than the associates being elicited through a free association paradigm, thus allows researchers to have good control over the responses to be given by the learners. In scoring, each correct option was awarded one point. The maximum possible score, therefore, is 160 for the 40 items. However, Zhang and Koda (2017) argue that how a WAT item is to be best scored is still an issue under debate. The separation of depth proficiency level with WAT in this study is revealed below in the analysis section. The reported test reliability is 0.98 (Read, 2007). WAT has been used in a number of depth studies as an ideal test format (Ehsanzadeh, 2012; Qian 2004).

To further investigate the participants' depth of word knowledge and its threshold level for L2 incidental learning through reading, I conducted a think-aloud protocol. A think-aloud protocol is a version of a verbal report in which participants state their thoughts which also reveals their strategic processes during text comprehension (Soria, 2001). Throughout the think-aloud retrospection, prompts such as "Why did you choose it?" or "How did you guess?" were given. The participants' retrospection and the probes were in Persian, as it was expected that verbalization in L1 would generate more information. The think-aloud retrospection protocols were recorded for later analysis.

Gauging incidental vocabulary learning

I administered the Vocabulary Knowledge Scale (VKS; Paribakht & Wesche, 1997; Wesche & Paribakht, 1996) in three testing sessions during a period of one month to measure incidental vocabulary acquisition. The VKS test has an easy format to use. It poses five questions, combining self-report and performance items to demonstrate knowledge of specific words in the written form. The test items reveal word knowledge, "ranging from total unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence" (Paribakht & Wesche, 1997, p. 179). The VKS was used in a number of studies to measure the retention of lexical items (Benglelail & Paribakht, 2004; Fraser, 1999; Kweon & Kim, 2008; Paribakht, 2005).

The VKS is highly practical and can be used with any set of words. However, an obvious limitation to the VKS is that it only tests written ability but not oral and aural abilities. Alongside, Read (2000) criticizes the VKS by noting that "it is doubtful whether learners' developing knowledge of second language words can be meaningfully represented by a single linear scale" (p. 136). I decided to use the VKS test for the present study because it has the advantage of taking track of the gradual development of vocabulary knowledge throughout a period of time. In the present study, all the test levels were used and scored according to the VKS scoring procedure. Test reliability that authors reported is 0.99.

Reading comprehension test

The reading comprehension test chosen for the present study was the Graduate Record Examinations (GRE) reading comprehension test. According to ETS (2011), "the purpose of the reading comprehension questions is to measure the ability to read with understanding, insight, and discrimination" (p. 12). Because the GRE reading comprehension test is an international standardized test, it was expected that it would determine the participants' level of text comprehension and their

general reading comprehension ability. The test items were taken from the actual GRE General Test administered in 2009 (ETS, 2011, Form GR 19-17, pp. 136-137, 148-150). The reading comprehension test contains two relatively short passages and two long ones, with 22 multiple-choice questions in total. In scoring, each correct answer was awarded one point. As there were 22 questions, the maximum possible score was then 22.

The text and target words

To reach the desired text density (see section 'Threshold level studies' above), I examined a number of graded readers and short stories. From the books I examined as the candidates for the present study, I chose a short story, "The Mirror" (Kay & Geleshenen, 2001), as the reading text for this study.

I simplified the text by replacing some low frequency words with high frequency words to ensure that the short story was suitable for the present participants. I also gave some explanatory sentences to make the complex structures more readable throughout the text for the participants. After simplifying the text, I calculated the text readability by using the Flesh scale. Results showed a readability index of 99. The rationale for choosing such a high readability index was to make sure that the participants knew all other words in the text; in this way, I would be sure that only target words would affect the results of the study.

I chose 30 words as the target words for the present study. Research indicates that at least 95% of text tokens should be known to participants for successful inferencing from context to take place, and a minimum of that figure will lead to the transfer of reading strategies (Hsueh-Chao & Nation, 2000; Nation, 2013). Nation (2013) also argues that 95% coverage of running words is still a heavy load of unknown vocabulary, and densities, like 1 in 50 (98% coverage), would be better. However, the optimal rate of the unknown/known ratio seems to be between 96 to 99% coverage of text tokens (Waring & Takaki, 2003). Therefore, I decided to choose 30 target words to reach the 98% unknown/known words ratio in the reading text. I also attempted to choose the words with different numbers of word syllables in order to counterbalance the effect of word length on the retention trend. Another criterion focused on parts of speech, as target words were attempted to be chosen from all parts of speech. The context of the story provided significant clues throughout the text as for the meaning of the target words.

To ensure that these target words were unknown to participants, I replaced the target words by pseudo words. A pseudo word is "an invented word constructed according to the orthographic and morphological rules of the target language" (Pullido, 2007, p. 73). The main advantage of pseudo words is that they allow us to control the learners' background knowledge and to ensure that the target words would not be known before reading.

The pseudo words were constructed in a way to look like plausible English words. I then checked them for plausibility, and implausible words were discarded. In the next step, I underlined the pseudo words and fitted them into the short story. Then the new version of the reading was further evaluated by two colleagues for naturalness of the contexts comprising the target words and smoothness of the text as a whole. The pseudo words are given in Appendix A.

Treatment procedure

I conducted three VKS posttests (immediate VKS1, VKS2, and VKS3) to gauge the actual incidental learning through reading over a period of one month. Data collection was in the form of paper-and-pen testing. The first and main testing session was about three hours. A colleague and I prepared the reading materials for the participants and asked them to read the short story in no more than 40 minutes. During the second hour of this session, participants also took VKS1 and the WAT. To gain more insight

into the participants' depth of word knowledge, I asked the participants to provide think-aloud retrospection protocols on the WAT items. In order not to intervene in the participants' test taking process, a colleague and I conducted the retrospection protocols shortly after the participants completed the WAT. Before actuating the retrospection process, the participants were given individual instructions concerning the think-aloud protocol to help familiarize them with the process. All the protocols were transcribed and then analyzed. This session was ended with the reading comprehension test (GRE).

Two weeks after the first session, the participants took the VKS test a second time (VKS2). In the third testing session, one month after the first session, the third vocabulary posttest (VKS3) was administered. Through the use of pseudo words in this study, I hope that any change in participants' vocabulary knowledge is due to the reading that took place and changes are not simply due to the one month of data collection. In the meantime, the participants were not informed of the three vocabulary posttests.

Data Analysis and Results

In the first step of the data analysis, I conducted a repeated measure analysis of variance (ANOVA) to investigate the relationship between L2 vocabulary depth and incidental vocabulary acquisition. ANOVA does not provide statistical information on the threshold level; however, it can confirm whether there is a significant difference between leaners with greater and lower depth of vocabulary knowledge which, in turn, will help us in determining the threshold level. ANOVA has been used in a number of vocabulary size and threshold level studies (Nassaji, 2004; Qian, 2002; Tsuraya et al., 2018).

To probe whether L2 learners with greater depth of vocabulary knowledge differ from those participants with lower depth of knowledge in their rate of L2 incidental learning, I divided the participants into three proficiency groups, namely high proficiency (HP) group, intermediate proficiency (IP), and low proficiency (LP). This division was based on the participants' percentile rank for their scores on the WAT. The participants occupying the first one third of the ranks (the first through the 40th ranks on the results of the WAT) were grouped as HP; those occupying the second one third of the ranks (from the rank 41 to the rank 80) were designated as IP; and the participants whose ranks were from 81 through 120 were then labeled as the LP group. Each proficiency group contains 40 students. The HPs are learners with greater depth of word knowledge, and the 40 participants in the LP group were thus supposed to be learners with lower depth of vocabulary knowledge.

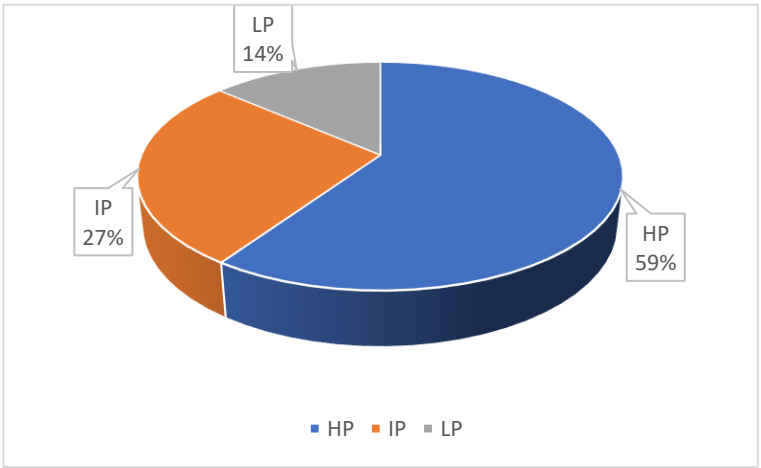

Figure 1 *Three proficiency groups' mean scores on WAT (figures are in percentages)*

In terms of the relationship between L2 vocabulary depth and the results of incidental L2 vocabulary acquisition, ANOVA with repeated measures (within-subject factors) is particularly susceptible to the violation of the assumption of sphericity. Therefore, I conducted Mauchly's test of sphericity. Results of Mauchly's test of indicate that the assumption of sphericity had not been violated, χ^2 (2) = 3.343, *p* = .188. Table 1 shows the results of Mauchly's test.

Table 1 *Mauchly's test of sphericity for the three proficiency groups*

Within Subject Effect				
	Mauchly's W	Approx. Chi-square	df	sig.
Vocabulary Depth	.343	3.343	0.68	.188

To investigate the existence of a probabilistic threshold level, I also conducted a post hoc multiple comparison test to locate differences among the three proficiency groups in terms of incidental L2 vocabulary learning. The post hoc test is supposed to be done all together among the three proficiency groups. Results of the pot hoc revealed that the mean difference among the three proficiency groups was significant, (F (1, 120) = 84.93, p≤ 0.05, effect size (d) = 0.68)). Results obtained from ANOVA are displayed in Figure 2.

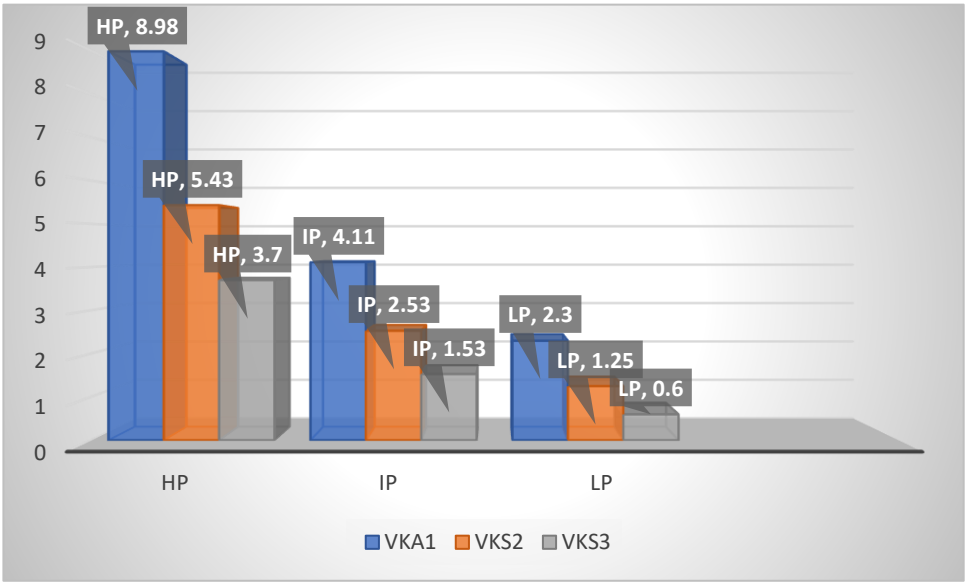

Figure 2 *Mean scores (SD = 4.0) on L2 incidental vocabulary learning for the three proficiency groups (N=40 for each proficiency group)*

Figure 2 shows that HP's mean on L2 incidental vocabulary learning was significantly higher in VKS posttests compared to those of IP and LP groups' means, F(1, 120) = 123.58, p ⩽ 0.05, effect size (d) = 0.8. HP's mean for VKS3 is 3.70, which is way higher than IP's (1.53) and LP's (0.60). It is worth it noting that the mean scores of VKS1 for HP, IP and LP groups are statistically different, demonstrating that WAT appropriately distinguished the participants in terms of their depth proficiency.

The difference that exists among these three groups in terms of depth of word knowledge and the gained vocabulary supports the existence of a probabilistic threshold level of depth of word knowledge among the participants of this study. However, ANOVA only confirms whether there is a significant difference among groups but not the threshold level of group abilities, the participants' scores on WAT show that those ELLs with a score of ≥85 (85 is the HP minimum score on WAT) stand a higher chance for L2 incidental word learning through reading.

Unlike incidental learning, participants' scores on the reading comprehension test across the three proficiency groups didn't show a significant difference. This was also confirmed by running an ANOVA, revealing no significant mean differences due to depth of word knowledge. The ANOVA showed that these differences were not significant, $F(1, 120) = 54.47$, $p > 0.05$, effect size (d) = 0.24. The results of the ANOVA are presented in Table 2. This result and the fact that no significant mean difference is seen among the three groups with respect to their reading scores indicates that picking up a border line as the probabilistic threshold level for reading comprehension cannot be supported statistically. This finding which is contradictory to many precursory studies in the field (Ehsanzadeh, 2012; Harkio & Pietila, 2016; Qian, 2004) might be due to potential influence of pre-existing differences among study groups in terms of L2 proficiency and reading comprehension ability, as they can serve as a confounding variable influencing participants' incidental vocabulary acquisition and reading comprehension achievements.

Table 2 *Mean scores on reading comprehension test throughout the three proficiency groups*

	HP	IP	LP	F	d
Reading Comprehension	25.00 (4.70)	23.42 (4.01)	22.08 (4.00)	54.47	0.24

Standard deviations are in parentheses. N=40 for ach proficiency group. p> 0.05

Think-aloud retrospection protocol of WAT

To further investigate the participants' depth of word knowledge and the strategies that they employ in dealing with the items of WAT, I conducted a think-aloud retrospection protocol. Verbal reports have been used in research studies as a method of getting at the mental processes that language learners use (Anderson, Bachman, Perkins, & Cohen, 1991). Think-aloud protocols involve participants thinking aloud as they are performing a set of specified tasks. Participants were asked to talk about whatever they are looking at, thinking, doing, and feeling as they go about the task. This enabled me to see the first-hand process of WAT completion (rather than only its final product). Protocols were audio-recorded, so that I could go back and refer to what participants did and how they reacted. All protocols were transcribed. Two colleagues also checked the recording and transcription of the protocols.

The analysis of the think-aloud protocols was based on the written transcripts. For the present study, I adapted He's (1998) transcription conventions. I analyzed the introspective think-aloud data on the basis of the three proficiency groups and with a consideration of two criteria: the strategies that participants employed in matching the WAT stimuli with four of the eight provided options, and the participants' depth of word knowledge score and the role that it might have on participants' answer. Also, I made a distinction between valid and invalid attempts. An attempt was considered valid if the participant talked aloud about his thoughts and gave a reason why he chose a particular option, and also the chosen options were discussed well. There is a total of 19,200 possible attempts (120 participants × 4 answers for each question with a total of 40 questions). Of all the attempts, I rejected 383 attempts as invalid. I then encoded the 18,817 protocols on the basis of the proficiency level and the employed strategies. Two colleagues also encoded 75% of the protocols and they reached a 97.26% inter-rater level of agreement.

The results obtained from protocols show that LPs often used translation strategy, as they immediately attempted to translate the stimulus word and the eight options in Farsi, the participants' first language. Then they tried to match the meanings together, looking for the closest synonym or the best collocation. On the other hand, the translation strategy was rarely seen among learners in the HP and IP groups. Choosing an association based on grammatical analysis is the strategy which was common among HPs.

An example from a participant in HP group on item 31 of the WAT is presented below:

Item 31: **Original**

Careful Closed First Proud	Condition Mind Plan Sister

Participant 6 (P6): …ahhh… "Original" is an adjective. Its noun is origin. I know that "al" makes it an adjective. So, you know, I choose "first" in the right-hand column.
Researcher (R): Why did you choose this option?
P6: Well, I said, I know that origin is the noun of this word. Origin means earlier steps, it is first things, so that I think "first" should be the best answer for this question.
R: Ok, what about the second box?
P6: I know that "plan" is the right answer here.
R: Why?
P6: "Plan" is a noun that can come with an adjective like "original". I also have seen these two together in my readings, I know that they come together.

In dealing with the collocational section of the WAT, LP learners again tended to translate the words into Farsi. Then they checked it out with the collocational possibility in their mother tongue. However, I noticed that IPs and HPs mostly relied on collocational reasoning in English. Below are two more examples from a participant in the IP group on item 37 and another participant in the LP group on item 39 of the WAT:

Item 37: Dramatic

Exciting Official Surprising Worried	Adventure Change Patient Salary

P28: (While looking at the words in the collocational section, matching them with the stimulus word, and reading them slightly) I choose "dramatic change" …ahhh… you know, I feel that this is the best option. It sounds like real English.
R: What do you mean by real English?
P28: I heard it, like in movies. It sounds reasonable to me because change can be dramatic. Change is a noun and dramatic is an adjective, so an adjective and a noun make a correct grammatical structure.
R: So, why not "dramatic salary"?
P28: Oh, no, "dramatic salary" seems wrong. I have not seen it before. I'm sure, yeah, "dramatic change" sounds OK.

Item 39: Beautiful

Enjoyable Expensive Free Loud	Education Face Music Weather

P18: Beautiful … ummm …
R: Do you know this word?
P18: Yeah, I know this word. In the second column, I want to choose weather. I think this is the best match.
R: Why did you choose weather?
P18: Oh … It's clear. We use it in this way to describe things like a beautiful flower or a beautiful picture. "Beautiful weather" means good weather.
R: Can you tell me more?

P18: Yes, beautiful should be used with a noun, in my language we say in this way.

Discussion and Conclusion

Results support the significant role of depth of word knowledge in L2 incidental learning. Results also confirm the existence of a threshold level of depth of word knowledge for L2 incidental vocabulary learning. According to the obtained data from this study, thus, those ELLs with a score of ≥85 (85 is the HP minimum score on WAT) stand a higher chance for L2 incidental word learning through reading.

Although there have been numerous studies which consistently support the role of depth of vocabulary knowledge in reading comprehension (Nassaji, 2004; Qian, 2002; Schmitt, 2014; Webb, 2013), the result of this study shows no significant mean difference among three proficiency groups with respect to the role of depth in reading comprehension. This contradictory finding with precursory studies could be related to limitations of this study. Confounding variables for incidental vocabulary acquisition such as participants' motivation and L2 proficiency is one limitation of the study. Another limitation is confounding variables in reading comprehension, namely as decoding skill, background knowledge, and reading strategies. Meanwhile, being a quasi-experimental study is another limitation of the study that might affect the results of the study.

Findings of the think-aloud retrospections also revealed that the LPs and the IPs have commonly resorted to the translation strategy and checked out the collocational possibility in Farsi, while, the HPs tended to utilize their richer network of word knowledge to associate options to the stimulus words of the WAT. This observation demonstrates that the threshold level of depth of word knowledge could be considered as the minimum level of lexical proficiency learners need to approach collocational possibility more in English rather than in their mother tongue. Therefore, findings of this research cry out for a change in EFL contexts, as the focus of vocabulary teaching is mostly on enlarging the number of words, breadth of word knowledge, rather than on enriching depth of knowledge (Nation, 2013); as a result, it led to the situation that learners mostly rely on translation strategy in dealing with the collocational possibility, the same as LPs in this study. Concerning the results of this study and the role of depth of word knowledge, it is recommended to incorporate tasks whose foci are on depth of word knowledge (Webb, 2013). In this regard, Webb thinks that teachers can raise their students' awareness of depth knowledge by using activities like vocabulary notebooks (Schmitt & Schmitt, 1995) that explicitly put ELLs' focus on different aspects of word knowledge.

References

Anderson, N., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources. *Language Testing, 8, 1,* 41–66.

Educational Testing Service. (2011). *GRE: Practicing to take the general test.* Princeton, NJ: Educational Testing Service.

Ehsanzadeh, S. J. (2012). Depth versus breadth of lexical repertoire: Assessing their roles in EFL students' incidental vocabulary acquisition. *TESL Canada Journal, 29*(2), 24–41.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research.* London: Routledge.

Ender, A. (2014). Implicit and Explicit Cognitive processes in incidental vocabulary acquisition. *Applied Linguistics, 10,* 1–26.

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

Harkio, N., & Pietila, P. (2016). The role of vocabulary depth and breadth in reading comprehension:

A quantitative study of Finnish EFL learners. *Journal of Language Teaching and Research, 7*(6), 1079–1088.

He, A.W. (1998). Answering questions in LPIs: A case study. In R. Young & A.W. He (Eds.), *Talking and testing* (pp. 101–116). Amsterdam: John Benjamins.

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

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2632–2637). Hoboken, NJ: Wiley-Blackwell.

Kay, J., & Geleshenen, R. (2001). *Discovering fiction (Student Book 1): A reader of American short stories.* Cambridge: Cambridge University Press.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From humans thinking to thinking machines* (pp. 176–201). Clevedon: Multilingual Matters.

Milton, J. (2009). *Measuring second language vocabulary acquisition.* Bristol: Multilingual Matters.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Canadian Modern Language Review, 61,* 107–134.

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

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 147–200). Norwood, NJ: Ablex.

Pullido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics, 28*(1), 66–86.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56,* 282–307.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading comprehension. *Language Learning, 52,* 513–536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355–371.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). *Word Associates Test* (Version 4.0). Retrieved March 21, 2011, from http://www.lextutor.ca/

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *IJES, 7* (2), 2007, 105–125.

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

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

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4)*,* 913–951.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*, 26–43.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT Journal, 49,* 133–143.

Soria, J. (2001). A study of Ilocano learners' lexical inferencing procedures through think-aloud. *Second Language Studies, 19,* 77–110.

Tsuraya, A., & Atmowardoyo, H. (2018). A cross-sectional study on vocabulary size among different levels of university students. *ELT Worldwide,* 5(1), 64–72.

Webb, S. (2013). Depth of vocabulary knowledge. *The encyclopedia of applied linguistics*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53,* 13–40.

Wilkins, D. A. (1972). *Linguistics in Language Teaching*. London: Arnold.

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education, 1*, 11–24.

Appendix 1

List of pseudo target words

Pseudo word	**English equivalence**
Armantisse (V)	Suspect
Batcock (V)	Lean
Berrow (N)	Notice
Cigbet (A)	Baggy
Combustulate (V)	Surround
Eluctant (A)	Orphaned
Gleap (N)	Boat
Haubard (N)	Behind
Humberoid (A)	Abandoned
Jousle (V)	Fight
Loruste (A)	Charming
Madline (V)	Suffice
Menstruable (A)	Disappointed
Nase (V)	Use
Oestrogeny (N)	Tele-screen
Pertrugant (N)	Astronaut
Pulark (V)	Walk
Quorant (A)	Deflated
Retrogradient (N)	Wrapping
Sabrident (A)	Different
Slaunder (V)	Spend
Smeg (A)	Nice
Snagert (N)	Doctor
Soramp (V)	Stop
Stace (V)	Say
Trudgeon (N)	Family
Vauche (V)	Aware
Waley (N)	Privacy
Woolnough (N)	Movement
Zerdox (V)	Write