



---

# **Investigating the Impact of Task-Induced Involvement Load on Incidental Second Language Vocabulary Acquisition through Reading**

---

By

**Dr. Amal Omar Alobaid**

*Lecturer, Administrative and Social Sciences  
Department, King Saud University, Riyadh, Saudi  
Arabia*

---

## Investigating the Impact of Task-Induced Involvement Load on Incidental Second Language Vocabulary Acquisition through Reading

Amal Omar Alobaid

Lecturer, Administrative and Social Sciences Department, King Saud University, Riyadh, Saudi Arabia

E-mail: amalomar211@hotmail.com

### ABSTRACT:

This study aimed to investigate the effects of task-induced involvement load on the L2 incidental vocabulary acquisition through reading. The study made use of the experimental method to meet such an end. A number of instruments for fulfilling the research main purpose were utilized, namely Oxford placement test, and the vocabulary size test, and Wesche & Paribakht's (1996) Vocabulary Knowledge Scale. Descriptive statistics were employed to explore the statistical differences between the groups. Once this difference was shown to exist, inferential statistics were used to explain whether differences were significant or not. One-way ANOVA followed by Bonferroni post-hoc test was carried out to compare groups. The participants were 45 pre-intermediate ESL students in the preparatory year in King Saud University. The results of the study revealed that task-induced involvement has considerable gains in developing vocabulary acquisition and retention across the three administered tasks. The study concluded with a number of pedagogical implications and suggestions for further research.

**Keywords:** Task-Induced Involvement Load, Incidental Vocabulary, Acquisition.

استقصاء أثر المشاركة المستحثة بالمهام على اكتساب مفردات اللغة الثانية العرضية من خلال القراءة

أمل عمر العبيد

محاضر بقسم العلوم الانسانية والادارية، جامعة الملك سعود ، الرياض ، المملكة العربية السعودية

البريد إلكتروني: amalomar211@hotmail.com

ملخص :

هدفت الدراسة الحالية إلى استقصاء أثر المشاركة المستحثة بالمهمة على اكتساب المفردات العرضية في تعلم اللغة الثانية من خلال القراءة، واعتمدت الدراسة على المنهج التجريبي من خلال توظيف عدد من الأدوات التي تمثلت في اختبار تحديد المستوى الخاص بجامعة أكسفورد (Oxford)، واختبار حجم المفردات، ومقياس معرفة المفردات والذي أعده ويسك وبإريباخت (Wesche & Paribakht 1996)، وتمثلت عينة البحث في عدد ٤٥ طالبًا من المقيدين بالمرحلة قبل المتوسطة بجامعة الملك سعود، وقد قام البحث بتوظيف الإحصاء الوصفي والإحصاء والاستدلالي لاستكشاف الاختلافات بين مجموعات البحث، كما اعتمد البحث على تحليل التباين أحادي الاتجاه ANOVA متبوعًا باختبار Bonferroni للمقارنة البعدية بين المجموعات، وقد كشفت نتائج الدراسة أن المشاركة الناتجة عن المهمة لها أثر في تطوير اكتساب المفردات والاحتفاظ بها عبر المهام الثلاث المكلف بها الطلاب، واختتمت الدراسة بعدد من المتضمنات التربوية والمقترحات لمزيد من البحوث المستقبلية.

الكلمات المفتاحية: عبء المشاركة المستحثة بالمهمة، المفردات العرضية، الاكتساب.

---

## Research Background

Vocabulary acquisition is a vital component of learning a language. It is critical to have a considerable range of vocabulary at one's repertoire for efficiently communicate in a foreign language. Without enough vocabulary knowledge, learners may not be able to express themselves or they would not have the ability to fully engage with the language or understand others. Moreover, vocabulary is the building block of language skills indicating that assigning time and effort for acquiring vocabulary can significantly improve the individual's ability to communicate in a foreign language and boost the overall language learning.

First of all, it is essential to make a distinction between the meaning of incidental and intentional vocabulary acquisition. Eysenck (1982) notes that such a distinction lies in the issue of pre-learning: intentional learning will warn subjects about the existence of a follow-up retention test, whilst incidental learning will not. That is to say, intentional learning involves a learner's desire to commit particular information to memory, whilst incidental learning means that any vocabulary learning will be a by-product of an activity that may not target lexical learning directly. It is important not to confuse the difference between intentional and incidental learning with the notions of implicit and explicit learning. Both intentional and incidental learning may occur explicitly, whereas implicit learning can only be incidental. Implicit learning refers to the acquisition of knowledge without conscious operations. Explicit learning, on the other hand, is a conscious operation where a learner will make and test hypotheses to find an underlying structure. In this regard, Ellis (1994) argues that implicit learning takes place whilst learning perceptual aspects of lexical items: that is, when acquiring phonetic and phonological characteristics of a word. Conversely, understanding the meaning of words is learnt explicitly, as this process requires conscious processing of semantic and conceptual characteristics, as well as attention to the form-meaning relations.

Nation (2001) defines incidental vocabulary acquisition as the process whereby vocabulary is acquired whilst focusing on other matters. A learner will focus on the message conveyed rather than just focusing on learning lexical items as a separate aspect of language learning. Similarly, Schmidt (1994) assumes that incidental learning occurs when students intend to learn a particular thing and so happen to learn lexical items along the way. For instance, a learner who wishes to communicate a particular message may in the process acquire a particular item of vocabulary. In addition, Joe (1998) argues that incidental vocabulary acquisition occurs when learners focus on context, e.g., when working on comprehending a

---

text. Nation (2013a) considers incidental learning from context as one of the most crucial sources of vocabulary learning.

Further to this, incidental learning allows researchers to study the effects of information processing. Laufer and Hulstijn (2001) note the types of designs through which this can be achieved. First, subjects may be exposed to material without instruction to learn. For instance, participants may be asked to read a text for comprehension, after which they would be tested on word retrieval. An alternative test design could involve instructing the subjects to learn, but not the information targeted in a follow-up test. An example would be to ask subjects to read and understand a text, and then to prompt them for the unfamiliar words presented in the text.

As previously mentioned, apart from having a methodological purpose, incidental vocabulary learning may refer to learning without the intent to do so. The most frequently used example is learning vocabulary as a result of reading. The effectiveness of incidental vocabulary learning is supported by Paribakht and Wesche (1999), who regard learners' attempts to understand new lexical items in a particular context as the source of incidental vocabulary acquisition. On a similar note, Laufer and Hill (2000) claim that activities such as reading, listening, and speaking trigger the learning of vocabulary incidentally.

There are some authors who argue that there is no possible way of determining whether lexical items have been learnt either incidentally or intentionally (see Gass, 1999). This might lead to reject considering intentional and incidental learning of vocabulary as direct opposites, which is something that Laufer (2001) proposes. Barcroft (2009) suggested considering these two approaches as tendencies in the learning process, where learner goes through a sequence ranging from more intentional to more incidental. However, other authors such as Nation (2013a) contend that incidental and intentional learning complement each other.

Despite the different orientations in defining incidental vocabulary acquisition, none of the definitions is considered authoritative. With regard to the current study, the researcher will adapt Schmidt's (1994) definition that incidental vocabulary acquisition is learning without the intent to learn. Learning psychologists concede that more attention dedicated to the formal and semantic aspects of lexical items together with richer associations with students' current

---

knowledge means that there are higher chances that the new information will be retained.

Further to this, both learners and teachers are aware that motivation is the key to achieving a high degree of proficiency in L2. The more motivated students are known to perform better, and educationalists make a conscious attempt to increase such motivation by providing tasks and materials that are interesting to the student. The Involvement Load Hypothesis discussed in this paper exemplifies the motivational dimension to vocabulary learning.

Laufer and Hulstijn (2001) developed the Involvement Load Hypothesis with the belief that learning and retention of L2 words depends on the involvement load (mental processing) induced by a task. This particular hypothesis is based on Craik and Lockhart's (1972) processing depth theory. This theory addresses the impact of the depth of processing new words based on the idea that any new input requires a series of stages of processing. The authors' main finding in the field of cognitive psychology was that different kinds of processing have different effects on the recall and retention of newly encountered information. Applying this to the field of learning vocabulary, it can be said that a greater depth implies a greater degree of semantic or cognitive analysis of the new word.

Depth of processing theory has been criticised because of its vagueness in regard to both what constitutes the richness of processing as well as the ways in which different levels are compared and measured. However, Lockhart and Craik (1990) pointed out that their approach should serve as a framework whereby new research would surpass their theoretical account. In the Involvement Load Hypothesis, Laufer and Hulstijn (2001) introduced a construct of involvement with marginally more observable and measurable cognitive and motivational factors. This probably overcomes the ambiguity and non-operational notion of the depth of processing, as proposed by Craik and Lockhart (1972).

The aim of Laufer and Hulstijn (2001) is to put these general cognitive notions to use in regard to L2 vocabulary learning tasks. The Involvement Load Hypothesis suggests that need, search, and evaluation are components of a task-induced involvement load, which determine the effectiveness of any incidental vocabulary task. The need component is the motivational, non-cognitive dimension of involvement. It indicates whether knowledge of a new lexical item is required for completing a task. Need is deemed to be moderate when it is imposed on the learner by an external agent, e.g., by the task itself or by the teacher. On the other hand,

---

need may be strong when learners are self-motivated—for instance, as in a learner looking up a word in a dictionary when reading.

In addition, search and evaluation are the cognitive dimensions of involvement load. Search refers to the learner's attempt to ascertain the meaning of an unknown L2 word, or to find a way of expressing a concept with an L2 word in order to achieve a task. It can be either present, where the learner seeks the meaning of the unknown L2 word, for instance, by consulting a dictionary or teacher, or absent, where no such effort is needed (e.g. when marginal glosses are provided). Evaluation entails a process of comparison between a particular lexical item and other words to determine how suitable that word is to a particular context. A moderate evaluation would entail recognising the differences between words or word senses in a particular context. Conversely, strong evaluation entails deciding which words can combine with a new word.

Not all of three factors —need, search, and evaluation—must be present when a learner processes a word in a task. It is the combination of factors with their various degrees of prominence that is the essence of the involvement load. For example, a learner is asked to write sentences with new words that are unknown to the learner. As the task is imposed by an external agent, it induces a moderate need (1); the search is present (1) as it requires looking for the new words by the student, and it involves a strong evaluation (2), as the student must evaluate the new words to find the appropriate context.

To test the effectiveness of the Involvement Load Hypothesis, Laufer and Hulstijn (2001) examined the studies in the field of L2 vocabulary learning with motivational and cognitive dimensions. The authors then checked the results of these studies against the Involvement Load Hypothesis to corroborate its validity. In the case of Hulstijn (1992), the results indicate that learners retained words better when they were required to infer word meanings from context by choosing from multiple choice options, rather than when they were given a synonym in a reading task. In this particular study, Laufer and Hulstijn (2001) argue that the differences among word retention lie in the lack of evaluation in the synonym-condition versus the presence of evaluation in the multiple-choice condition. They found that both conditions involved moderate need and no search.

Newton (1995) found that words that were negotiated for meaning (that is, requesting clarifications) were retained at a higher degree compared to non-negotiated words. Similarly, the results of Ellis, Tanaka & Yamazaki (1994)

---

indicate that interactionally modified input is associated with students retaining more items compared to items presented in pre-modified input (i.e., input prepared beforehand rather than negotiated). Similar results are confirmed in the study by Ellis and He (1999). Given these findings, it is likely that word negotiation implies need as well as a search for the meaning of those lexical items. If there is no negotiation, the learner does not need to know them, and consequently does not search for the meaning.

Joe (1995, 1998) noted that students were more successful at retaining words used in original contexts (created by the learner) rather than those used in non-original contexts (provided by the task). This means that words used in an original context generated by the learner are better retained; this process involves a higher level of evaluation. Similarly, Hulstijn and Trompeter's (1998) results indicated that looking up the words in an L1-L2 dictionary to use them in a composition proved more successful than simply looking those words up for comprehension purposes; writing composition requires more evaluation than in reading task. Cho and Krashen (1994) found that the use of dictionaries as a self-imposed task entailed learners acquiring more lexical items than through reading. This contributes to the claim of the Involvement Load Hypothesis regarding the relation between levels of need in a task and the vocabulary learning.

Although these studies did not set out to examine the Involvement Load Hypothesis, their findings suggest that the involvement load is linked to the effectiveness of vocabulary tasks, and that the hypothesis can be deemed a well-composed construct that merits further attention in the field of L2 vocabulary research.

After interpreting previous studies in light of the Involvement Load Hypothesis, Hulstijn and Laufer (2001) conducted their own study to investigate the effects of the task-induced involvement load on learning and retaining vocabulary. This study involved two groups of English as a Foreign Language (EFL) learners—79 in the Netherlands and 128 in Israel—who all exhibited a high level of proficiency in English. Both groups performed three tasks that differed in involvement load in order to test the initial learning and subsequent retention of ten target words. Subjects were tested after one week in Israel and after two weeks in the Netherlands.

The authors sought to confirm the hypothesis that tasks with higher involvement loads would lead to higher levels of vocabulary retention. The first

---

task was a reading comprehension task with marginal glosses where students were given a text with ten multiple-choice comprehension questions. The task induced moderate need but no search or evaluation. The second task consisted of reading comprehension with filling in the blanks. This task induced moderate need, no search and moderate evaluation. Finally, the third task asked the students to write a composition using the target words. This task induced moderate need, no search, and strong evaluation.

The results revealed slight differences between the two studies. Whilst in the Hebrew-English experiment, the results fully supported the Involvement Load Hypothesis, the results of the Dutch-English experiment only partially supported it. The reason is that in the latter experiment, the third task resulted in higher retention compared to the first two tasks, but there were no differences between the first two tasks.

Hulstijn and Laufer (2001) pointed out that there are some shortcomings to their study, and therefore, to the validity of their results. First of all, the researchers argued that in order to prove their hypothesis, the experiment would have required a wider spectrum of tasks. Further research should thus look into various tasks with different involvement loads in order to compare their effects on incidental vocabulary learning.

Keating (2008) also conducted research related to Involvement Load Hypothesis. The author tested 79 beginning learners of Spanish on one of three vocabulary learning tasks that differed in terms of degree of involvement. The task involving the least effort was the reading comprehension task, followed by a reading comprehension task supplied with target words. The highest degree of involvement was observed from a sentence writing task. The researcher assessed retention immediately after the task, and then again after two weeks. The results supported the Involvement Load Hypothesis, as retention was highest in the sentence writing task where involvement load was higher than in the other two tasks.

The contribution by Keating is relevant in that he was one of the first researchers to test beginning learners. This is important because more advanced readers have a greater capacity to infer meanings of unknown words. In this sense, I also aim to contribute to the hypothesis by testing pre-intermediate EFL learners. Keating (2008) further took into account the amount of time taken to complete

---

tasks, which has been shown to have an important effect on performance. This will be also taken into account in the current research.

Kim (2008) conducted two experiments with ESL learners with different levels of proficiency. One experiment consisted of three tasks with different levels of task-induced involvement. The second experiment consisted of two tasks with the same level of involvement. In experiment one, participants were randomly assigned to the same tasks used by Hulstijn and Laufer (2001). Experiment two engaged participants in two tasks: either writing a composition or writing original sentences with the use of word glosses.

Regarding experiment one, the results of initial word learning matched those of Hulstijn and Laufer (2001), in which the Composition group gained the highest score of all, while the Gap-fill group and the Reading group did not perform significantly different from each other. In terms of word retention, the results fully supported the Involvement Load Hypothesis in that more involvement load entailed higher retention levels. With regard to experiment two, the author reported that proficiency did not appear to be significant in terms of results.

Another contribution to the validity of the Involvement Load Hypothesis comes from Eckerth and Tavakoli (2012). The authors investigated the differential effects of word exposure frequency and elaboration of word processing in L2 learners. The 30 participants were advanced L2 learners of English at a UK university, and as in Kim (2008), were controlled in terms of time. The results of the study indicate that exposure frequency and elaboration of word processing yield similar effects in the initial word learning. On the other hand, the task-induced load resulted in significant effects compared to the word exposure frequency in terms of word retention. However, the authors pointed out that their findings were limited in that both the environment (classrooms) and the aspects of the research design such as immediate post-tests may have focused the attention of learners towards vocabulary acquisition. Hence, learning would not be entirely incidental. It is also not assumed that this limitation is not applicable in this study as the participants may perceive the tasks as a vocabulary test and so the results might provide limited support for the involvement load hypothesis. Yet, it is important to note that performing tests in the classroom does not necessarily mean ‘under classroom conditions’. Thus, the results continue to have some kind of validity.

Other studies such as Teng (2015) aimed to assess the validity of the Involvement Load Hypothesis when using translation tasks. Teng (2015) produced

---

three translation tasks for 60 native-Chinese low-proficiency learners of English. The three tasks varied in the amount of involvement mode: the first was simply translation; the second was translation plus fill-in the blank exercises; and the third was translation plus sentence writing. The author then tested the students two months after the initial experiment, measuring word form recognition, meaning recall, and word production. Once again, the results of the test supported the Involvement Load Hypothesis. This particular study was innovative in that only a few researchers had addressed the efficiency of translation tasks on vocabulary learning prior to the study.

As Kim (2008) notes, the Involvement Load Hypothesis has certain limitations. First of all, his study along with that of Hulstijn and Laufer (2001) are limited to one particular aspect of vocabulary learning, i.e., learning word meaning. To represent the effects of task-induced involvement in all aspects of vocabulary learning, further research would be required to examine other aspects of vocabulary learning, such as form, meaning, form-meaning, and mapping.

Kim (2008) also indicates a suggestion contrary to the assumption made by Laufer and Hulstijn (2001) that each degree of any component has the same contribution to the overall involvement index of the task. Kim (2008) argues that the extent in which degrees of each component (moderate and strong) might not contribute to an overall involvement load equally. The author suggests that strong evaluation might induce much higher involvement load than the moderate evaluation. It would be necessary then to test tasks with the different degrees of the same component in order to either support or undermine Kim's suggestion.

Other limitations to the theory are brought forward by Eckerth and Tavakoli (2012). The authors argue that the Involvement Load Hypothesis cannot account for some variables involved in incidental vocabulary learning through reading. For instance, there is no certainty that students read the glosses given with their tasks; and if they did read them, one cannot be sure how long they dedicated to reading such information. Another issue arises from the level of difficulty to find the correct meaning for fill-in-the-gap tasks. One may wonder if there are any contextual cues or reliance on the ability of students to guess. It is also possible that students derived the wrong meaning and retained such meaning over time. In this case, the focus is not on learning-retention, but on inferencing capacity. The current study attempted to handle the above issue by ensuring that the context in the (~~gab~~ fill Gap-fill) task is within the participants' comprehension, and also by consulting teachers of the participants regarding the participants' ability to get the correct answer.

---

However, Eckerth and Tavakoli (2012) believe that in order to overcome such limitations, further research should focus on finding how learners perceive and accomplish tasks. Other methods of testing should also be considered, e.g., think-aloud protocols or implicit ways to test representation and access to lexical knowledge (see Wesche and Paribakht, 1999).

## **Research Purpose and Questions**

It may be ascertained from reviewing the literature that the validity of the predictions following from the Involvement Load Hypothesis requires further investigation. The present study will attempt to contribute to the validity of the hypothesis by answering the following research questions:

- I. What are the effects of task-induced involvement load on the initial word learning of pre-intermediate ESL learners?
- II. What are the effects of task-induced involvement load on the medium-term word retention of pre-intermediate ESL learners?

The investigation takes place in an EFL setting, using three different kinds of tasks (A, B, and C). The involvement load induced by the tasks is operationalized through the three components of need, search, and evaluation. The study will cover the meaning aspect of vocabulary learning. Tasks will involve different degrees of the same component (evaluation) to investigate Kim's (2008) above suggestion. Time on tasks will be controlled in order to add more to the argument discussed earlier regarding the effect of time on the results of task-induced involvement studies. However, due to time restrictions, it will not be able to test learners' attitudes, as recommended by Eckerth and Tavakoli (2012).

## **Research Methodology**

### **Research Design**

According to Patton (1990: 39) it should be recognized that 'different methods are appropriate for different situations'. Thus, the main purpose of the study should be considered in determining the design of the research. The aim of this study is to investigate the effects of task-induced involvement load on L2 vocabulary learning and retention. The independent variable in this investigation was the task conditions operationalized through manipulating the three components need, search, and evaluation. The two dependent variables were the initial

---

vocabulary learning and the medium-term word retention, and they were both represented as scores. To explore the relation between these variables we conducted an experiment in the form of pre-test, treatment, immediate post-test and delayed post-test. The study controlled some factors such as age of the participants, their L1 and their English proficiency level.

## **Participants**

Given that choosing the right subjects is a crucial issue in the validity of the research, (Bell 2014) asserts on making efforts in selecting 'as representative a sample as possible' (p.150). In the current research, the participants were chosen from the accessible contacts of the researcher who met the requirements of the study. The participants were 45 pre-intermediate ESL students in the preparatory year in King Saud University. The preparatory year is divided into three levels – A, B, and C. This division parallels the one of the Common European Framework of Reference for Languages (CEFR). Given that participants in level B of the preparatory year are a mixture of the B1 and B2 levels of the CEFR, for more preciseness the Oxford Placement Test was used to measure students' proficiency level in English. The students' scored in the range of 26–34 (classified as being at the pre-intermediate level) were chosen to participate in the study. During the term the study was conducted, a three-month course, participants were exposed to English-teaching sessions of three hours five days a week. The participants were all females with Arabic L1 background and their ages ranged between 18 and 20. The reasons for choosing these particular participants were many. First of all, few studies have investigated the effects of the involvement load hypothesis on low-proficiency learners, thus we focused in the current study on these pre-intermediate learners. In addition, their level of English proficiency was still sufficient to complete most of the tasks and for the required vocabulary learning to occur, and they were easily available as I worked in the same place (Dornyei, 2007).

## **Research Instruments**

### **Selecting the Passage and the Target Words**

The first step in designing the tasks for this study was choosing a passage and target words. In terms of choosing the passage, because assessing reading comprehension was not one of the aims of this research, it was essential to make sure that the passage was within the comprehension of the participants. It has been stated that in order to comprehend a text without difficulty, a reader should have

---

95% (98% for higher levels) coverage of the vocabulary in that text (Nation, 2001). Thus, in order to measure the participants' vocabulary size, Version 1 of The Vocabulary Size Test by Nation & Beglar (2007) was administered (see 3.4.3). The average of the participants' vocabulary level was within the range of 2000 words. Thus a passage with a 95% coverage of the 2000-level vocabulary was nominated. Vocabulary Profile (VP) BNC-20 software was used to measure the coverage of the vocabulary in the text. This software is provided through the Complete Lexical Tutor website. The passage is a modified version of 'Honey for you Honey for me' from the York Assessment of Reading for Comprehension (YARC) secondary test by Stothard et al. (2010). This passage was chosen mainly because of its seemingly interesting theme and plot for the participants, which would help increase the motivation for learning. Lightbown & Spada (2006) relate aptitude for learning to high levels of motivation and interest. Thus, the kind of input was carefully chosen, as 'the choice of content can be a major factor stimulating interest' (Nation 2013a:102).

The target words were taken from the same candidate passage. The motivation and interest have been considered here as well, but this time at the word level. Words that play a central role in the text would have a greater chance of being learned than words that do not figure centrally in the plot (Nation 2013a). This was confirmed previously by Elley (1989), who found a relationship between the importance of the words to the text (importance was rated by teachers) and the gain in vocabulary knowledge. Nation (2013a) also indicated that interest could be built in the tasks. Thus, 'important' target words were chosen for both the reading passage and the combined activity. As the participants' vocabulary size was in the range of 2000 words, the chosen target words were of a higher level. They comprised two verbs, four nouns and four adjectives: fatal (adj), pale (adj), ungrateful (adj), greedy (adj), rove (v) , innovate (v) , companion (n) , portion (n) , expert (n), method (n).

To ensure a high probability that these words were not known to the participants, a pre-test was conducted. As done by Hulstijn & Laufer (2001), a group of pre-intermediate students, who were in similar course to the participants but not taking part in the experiment, were asked to indicate their knowledge of the target words. The results of this pre-test showed that those students indicated their unfamiliarity with the target words. Since the participants in our study were of a similar level of proficiency, it was taken for granted that most of the words would be unfamiliar to them. In addition, the teacher of the course was consulted with

---

regard to the probability of the target words being unfamiliar to the participants. As a final step, during the immediate post-test, participants were asked about their previous knowledge of the target words. Although not all students responded, 17 of them indicated knowing the word 'method', while 13 indicated knowing the word 'expert' and only 9 pointed out their pre-knowledge of the word 'companion'.

## **The Tasks**

Three tasks (A, B & C), integrated in different experimental conditions that differed in involvement load, were designed for the current investigation. The task-induced involvement load was operationalized by manipulating the three components of need, search and evaluation as suggested by Hulstijn & Laufer's (2001) Involvement Load Hypothesis.

**Task A:** Glossed reading comprehension. This task was adapted from Eckerth & Tavakoli (2012). The task involved reading the chosen passage and then answering two sets of comprehension questions about the passage. As indicated by Nation (2013a), paying attention to an aspect of language by noticing it is the first cognitive process enhancing learning. Nation confirmed that salience of the word in the text has an impact on noticing (2013a). For this reason, the target words were underlined and written in bold. Unlike Eckerth & Tavakoli (2012), who glossed the target words in L2, in this task each word was glossed with its brief definition in both L1 (Arabic) and L2 (English) as well as its parts of speech. The inclusion of the L1 translation was done because of what Jiang (2000) proposed in his lexical development model, which is based on information represented in the lexical entry. According to this model, at the beginning stages of learning an L2 word, only the form is acquired in L2 while the meaning is acquired in L1. The students were given the passages first and were asked to read it for meaning and to make use of the marginal gloss, as they were to be tested in comprehension later. After that, they were handed the questions and asked to answer them. According to Eckerth & Tavakoli (2013).

**Task B:** Reading comprehension plus gap-fill. This task involved reading the same passage as in Task A and answering just the first set of comprehension questions as in Task A. However, in this task the ten target words were removed from the text and replaced with empty spaces. The deleted target words were reprinted randomly on a separate page. Each of these words was explained using the same L1 and L2

---

definitions used in the gloss of Task A. The participants' task was to read the text, fill in the gaps and then to answer the following comprehension questions.

**Task C:** Writing original sentences. This task is a combination of Eckerth & Tavakoli's task 3 (2012) and Keating's task 3 (2008). Like Task A, this task asked participants to read the same glossed passage but with no comprehension questions, as done in Eckerth & Tavakoli (2012). On the next page they were provided with the same list of the target words along with their definitions, as in Task B, and then they were asked to write a sentence using each of these words, as done in Keating (2008). The instructions, as well as the teachers who were administering the task, insisted on the participants using the words in sentences that differed from the ones in the passage. In the same manner as the two previous tasks, the need was moderate and no search was required. However, evaluation is strong in this task, as it requires the participants to evaluate the target words against previously known words so they can create original sentences. In total, the involvement index of this task is 3.

### **Oxford Placement Test**

Allan's (2004) Oxford Placement Test 2 was employed to determine the learners' proficiency level. According to Allan (2004) the test is easy to administer, economic and with an objective scoring system providing meaningful and consistent scores. The test performs reliably, with Allan (2004) reporting its high item reliability across test populations.

### **The Vocabulary Size Test**

The Vocabulary Size Test (VST) by Nation & Beglar (2007) was administered for two main reasons: first, to make sure that there are no differences between the participants with regard to their level of vocabulary knowledge, and second, to measure the participants' vocabulary size with the aim of choosing target words of a higher level than theirs, as well as to choose a suitable text for the tasks.

The Vocabulary Size Test by Nation & Beglar (2007) was designed to measure the English vocabulary size for learners of English as a foreign or second language. A large number of researchers have validated and used the Vocabulary Size Test for assessing L2 vocabulary capacity (Beglar, 2010). The test's reliability and consistency, as reported by Beglar (2010), is 0.96.

---

The original test is divided into 14 levels, with each level having 10 items organized in a four-item multiple-choice format, with the difficulty of the items becoming greater level by level. Each word represents the value of 100 words; thus every level represents 1,000 words, and consequently the total vocabulary number of the whole test is 14,000 items. The test used the British National Corpus as a basis for the frequency count of word families. The test-taker's score is multiplied by 100 to get the total vocabulary size. For instance, if the total test score of a learner was 35 out of the 140 items, then the vocabulary size of that person is 3,500 words. Yet, due to the participants' low level of English proficiency only 5 levels out of the 14 were administered. So, only the words from first 1,000 to the fifth 1,000 were used, and since each item was worth one point, thus the maximum score is 50, which indicates that the participant's vocabulary size is 5,000. However, only the level of 2,000 words was reached by the participants.

Nation stated that since this test focused only on written receptive vocabulary knowledge, designing the test should not require any kind of knowledge other than vocabulary and reading skills. Elgort (2012) indicated that sitting the vocabulary test with writing choices in test-takers' L1 resulted in scores around 10% higher than when they were written in L2. Nation (2012) attributed this to the learners' immediate comprehension of the meaning without being hindered by any difficulties of grammar or other unknown words. Thus, some versions of the test were issued in bilingual format, but unfortunately there was no English–Arabic version to use in the current study. However, in Nation and Beglar's (2007) monolingual Vocabulary Size Test which was used in this study, the language for writing the choices was simple and much easier than the target word. For example, definitions used for the first and second 1000 word level used only words from the first 1000 of West's (1953) General Service List (Nation, 2012).

This test performs reliably for three reasons: first, it works in accordance with requirements and expectations, distinguishing learners of different proficiency levels according to their scores of the target words, as well as identifying the different levels of test-takers' vocabulary capacity; second, the test's scoring and result interpretation is simple; third, the test includes clear and unambiguous items that have been very carefully planned, formed, and finally trialled tested???

### **Assessment**

The current study used Wesche & Paribakht's (1996) Vocabulary Knowledge Scale (VKS) in the immediate and delayed post-tests to measure the

---

dependent variables, which are the participants' initial word learning and medium-term word retention. The VKS was adapted in this study for three reasons. First, as defined by Wesche & Paribakht's (1996) the VKS is a test of vocabulary depth rather than breadth, in which it tracks the progress of words in the students' memory. Second, its purpose is to evaluate the acquisition of new lexical items from reading L2 texts. Hence, it has been used for measuring incidental vocabulary acquisition. Third, the VKS measures the passive and active knowledge of the vocabulary at the same time.

Wesche & Paribakht (1996: 28) defined the VKS as a tool using a five-point scale featuring self-report and performance items. The scale represents the depth to which a lexical item has been acquired. This ranges from complete unfamiliarity to the ability to use the word accurately in a sentence. That is, students need to provide information regarding their knowledge of the tested words using the category scale in Figure 1.

**Figure 1** VKS elicitation scale of self-report categories (Wesch & Paribakht, 1996)

---

**Self-report categories**

- 1) I don't remember having seen this word before.**
- 2) I have seen this word before, but I don't know what it means.**
- 3) I have seen this word before, and I think it means\_\_\_\_\_ . (synonym or translation)**
- 4) I know this word. It means\_\_\_\_\_ . (synonym or translation)**
- 5) I can use this word in a sentence:\_\_\_\_\_ (If you do this section, please also do section IV.)**

---

Read and Chapelle (2001) characterize the VKS as a test that tracks the progress of words in the students' memory. The VKS allows researchers and practitioners to make inferences about learners' knowledge of words. In terms of test purpose, inferences are made at item level as it measures the growth in word knowledge. In other words, knowledge of word meaning is assessed without taking context into account, which is the case for the current study.

Although it is a self-report test, a strong correlation between the test-takers' ratings and the test scores has been found, with Wesch & Paribakht reporting this correlation as ranging between 0.92 and 0.97. Additionally, Paribakht & Wesch stated that the responses elicited by the test-takers are accepted as reliable.

Bruton (2009) highlights some of the shortcomings of the test design around VKS. One of the issues is that the VKS can only be used with decontextualized words. In cases where the learners know the word, they may produce a grammatical sentence, but they are not aware of which meaning may be sought in cases of homographs. However, in the case of the current investigation, the required meanings of the target words were already provided to the participants, thus the issue of decontextualization is not of a problem. Another issue that Bruton (2009) identifies is that the VKS makes it impossible for students to prompt for L2 word forms as the form is already given. This might be true for studies investigating the vocabulary knowledge needed for speaking or writing. In this study, using the VKS is appropriate as we are only seeking the written receptive word knowledge and the progress of this knowledge in the students' memory.

Figure 2: VKS scoring categories: meaning of scores (Wesch & Paribakht, 1996)

Self-report categories	Possible scores	Meaning of scores
I.	→ 1	The word is not familiar at all.
II.	→ 2	The word is familiar but its meaning is not known.
III.	→ 3	A correct synonym or translation is given.
IV.	→ 4	The word is used with semantic appropriateness in a sentence.
V.	→ 5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

The VKS follows a scoring procedure as shown in Figure 2. Scores 1 and 2 are given to responses to categories I and II respectively. A score of 2 also indicates responses to any higher categories that fail to provide an appropriate synonym or translation. Although the difference between categories III and IV is slight, representing how sure test takers are about their knowledge of the word, a score of 3 (indicating participant's passive knowledge) would be given to a correct response to both of them. Coming to category V (indicating participant's active knowledge), while a score of 4 is given for semantically correct use of the target word, a score of 5 is given when a word is used with both semantic appropriateness and grammatical correctness.

In the current study's immediate and delayed post-tests, the ten target words plus three no-target words were presented to the participants in a chart using the VKS scale. The instructions asked the participants to indicate their level of knowledge of the meaning of words by ticking any of the categories. However, if they ticked categories III, IV, or V they needed to prove their knowledge by either providing the translation or synonym (category III or IV) or by using the target

---

word in a sentence (category V). Each word received a score between 1 and 5, thus the total score ranged between 10 and 50.

### **Hypothesis**

Since the literature review showed that the involvement load hypothesis is valid for all L2 proficiency levels, it was hypothesized that the immediate and delayed post-test scores would be highest in Task C (writing original sentences), lower in Task B (reading comprehension plus gap-fill) and lowest in Task A (reading comprehension).

### **Procedures**

The participants were divided randomly into three groups each comprising 15 participants. Each group received the same reading passage and the target words but the conditions of the reading were different in each task (A, B & C). The three groups had the same amount of time for accomplishing their tasks, 35 minutes. After completing the tasks, the participants took an immediate post-test to measure their initial vocabulary learning. Then, a week later, they took a delayed post-test to assess their medium-term vocabulary retention. Both post-tests took about ten minutes and they were of the same format, but the item order was different. All tasks and tests took place at school during class time.

The tasks and tests were piloted on students who were not participating in the study, but who were of a similar level of proficiency to the study's actual participants. The piloting of the study was administered for two main reasons. First, to know the time needed for conducting the tasks and the tests. Second, to notice any unexpected problems or mistakes. Although Dornyei (2005) referred to reliability of procedures and instruments as the consistency of the results within the same participants in different circumstances, due to the required incidental nature of learning it wasn't applicable to pilot the study on the actual population of the study. However, the advantages of piloting the study were clear, as some significant changes were applied to the content of the tasks and tests. For example, in Task C. it was recognized that instructions asking students to create sentences differed from the ones in the text were not clear, so they were modified and capitalized to place more focus on them.

Two days before starting the data collection, the study was explained to the participants by the researcher, and they read the information sheet and signed the

---

consent form. Although the participants have been told that they would be doing some kind of vocabulary and reading activity, they were not expecting to be given a vocabulary post-test. After signing the consent form, the researcher administered the Oxford Placement Test and the Vocabulary Size Test (Nation & Beglar 2007). The treatments and post-tests were administered with the help of my colleagues in the preparatory year at King Saud University.

### **Statistical Analysis**

This study used a quantitative approach. Although some advantages, such as focusing on unique areas of individuals, can be obtained through a qualitative approach, Dornyei (2007) indicates that the precise measurement of the quantitative method, along with its generalizability of producing reliable and replicable statistics in other contexts, makes it as advantageous as the qualitative methodology. To analyse the results of the participants' post-tests, the SPSS program was used. Descriptive statistics were employed to explore whether there was a difference between the groups. Once this difference was shown to exist, inferential statistics were used to explain whether differences were significant or not. One-way ANOVA followed by Bonferroni post-hoc test were carried out to compare between groups. Statistical differences were considered significant at  $P \leq 0.05$  levels.

### **Results of the Research**

Detailed reports of the results pertaining to the initial learning (immediate post-test) and medium-term word retention (delayed post-test) experiments are provided. Results of both post-tests are explained in the form of histograms. Each of the three participant groups' results are separated in each histogram chart (using different colours) in order to allow a more simple description and visual search of the results. Also, more precise information regarding the number of vocabulary items acquiring each of the Vocabulary Knowledge Scale (VKS) scores "1, 2, 3, 4 or 5" in either parts of the study is provided in the form of pie charts. Lastly, the results are compared similarly in statistical terms (using a One-Way ANOVA and Pearson's correlation, respectively).

This approach will help to realize the differences between the three participant groups' results in both post-test investigations. It will also help to identify the nature of the relationship between initial word learning and delayed word retention using the same experimental tasks. Therefore, this approach will enable showing to what extent *task induced* involvement load specifically may prove a useful theoretical tool in vocabulary acquisition.

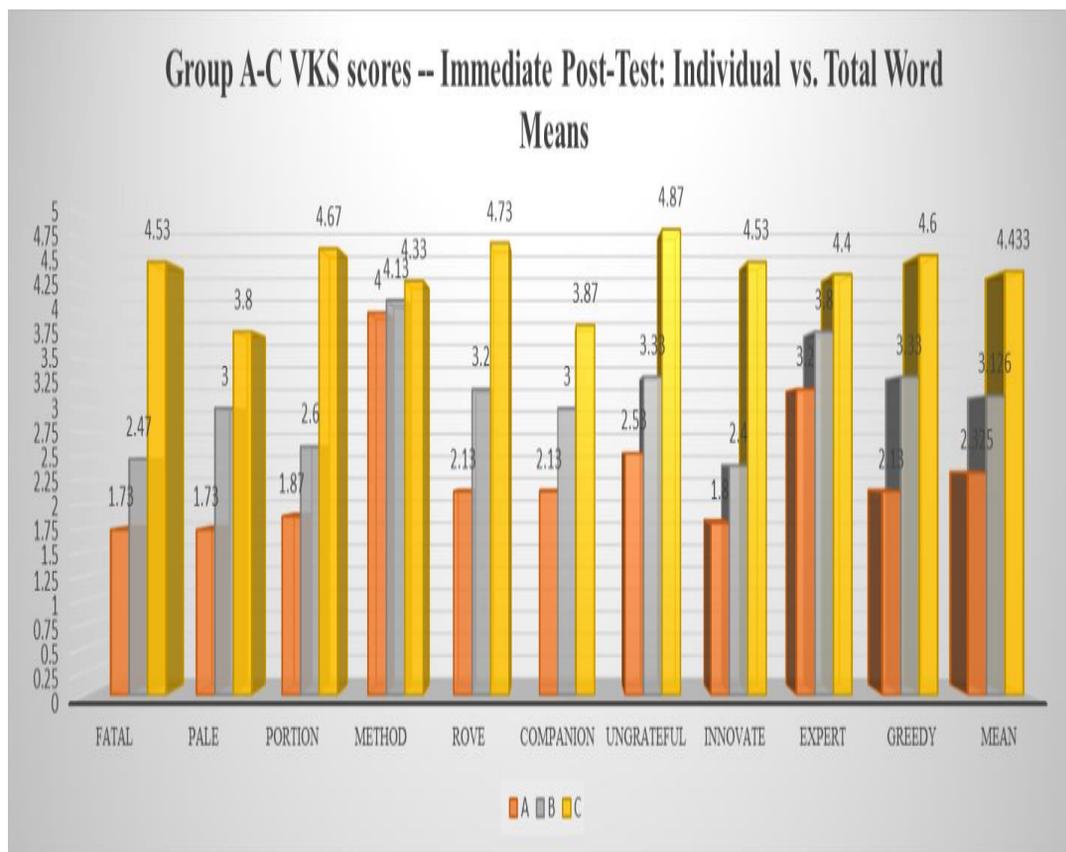


Figure 3: Group results in the immediate post-test and total word means for each group

Figure 3 describes the main descriptive statistics of the results for experimental groups A, B and C in the immediate word learning experiment. In general, It is shown that the mean score of almost all target words (when compared with the average scores 2.33, 3.13 and 4.43 for groups A, B and C, respectively) is the highest in group C (writing original sentences), lower in group B ('fill-in-the-gaps') and the lowest in group A (glossed reading comprehension). The mean scores of word forms 'Fatal', 'Portion', 'Rove', 'Ungrateful' and 'Innovate' all receive moderate improvements in VKS in group B compared to group A (compare red and grey bars for each word form, respectively) whilst improvements when comparing gain in group B and C are much greater (between nearly 1.5-2 on average, compare grey and yellow bars, respectively). For the rest of word forms, little or almost no improvement is shown based on the chosen learning tasks for 'method' and 'expert' (respectively), while 'pale', 'greedy' and 'companion' receive similar moderate gains based on the three tasks.

In summary, for the majority of words investigated, one of two patterns based on task assignment arise; for 5/10 words, task C receives much more notable gains than for either task 1 or 2, whereas for 3/10 words the gains made in tasks B and C compared to task A are fairly similar (for 2/10 words, minor gains were made based on task assignment).

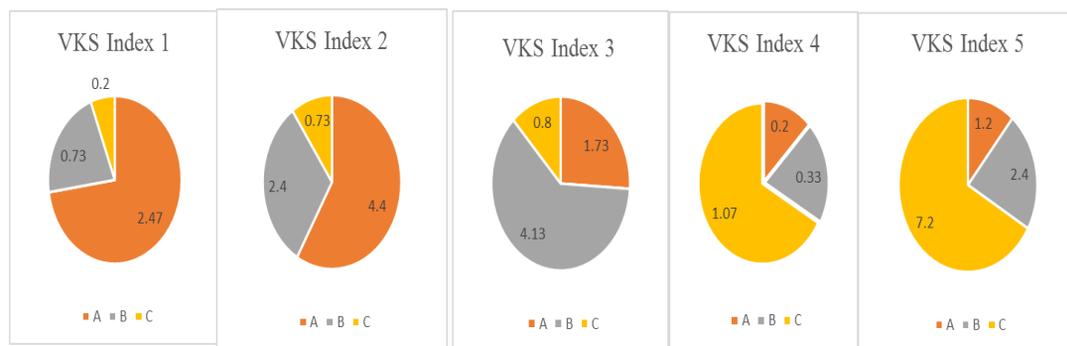


Figure 4: The results for VKS indexes 1-5 across groups A (orange), B (grey) and C (yellow) in the immediate post-test.

Figure 4 indicates the differences among the means of the number of vocabulary items acquiring each VKS score (1-5) per group (see points on scoring made in 3.6). The minimum score in the VKS for each item is 1, and the maximum is 5. It can be deduced from the five pie charts corresponding to the five indexes that the distribution of the scores of the VKS also indicates the superiority of group C over the other two groups, and consequently that the assigned task and the index given to the participants are correlated. In terms of active knowledge (score 4 or 5), group B had about two times higher the number of vocabulary items scoring 5 than group A, while group C is about three times higher than group B with regard to score 4 and 5. In terms of passive knowledge (score 3), the superiority went to group B, followed by A and lastly C; this was understandable, as most of the participants of this latter group indicated rather active knowledge. So, by counting the mean proportional VKS score for familiar word forms (i.e. for indexes 3-5) it is

---

found to be the highest in group C (3.02), lower in group B (2.29) and lowest in group A (1.04).

### Statistical Comparison of the Immediate Post-test Results

To examine the effects of task-induced involvement load (independent variable) on participants' initial vocabulary learning (dependant variable); the VKS results of the immediate post-test were formally tested by one-way ANOVA. The results display that the following words have very highly significant differences as well as a considerably high effect size ( $\eta^2$ ): 'fatal',  $F= 34.587$ ,  $p < 0.001$ ,  $\eta^2 = .622$ , 'pale',  $F= 11.53$ ,  $p < 0.001$ ,  $\eta^2 = .354$  'portion',  $F= 32.659$ ,  $p < 0.001$ ,  $\eta^2 = .609$ , 'rove',  $F = 21.073$ ,  $p < 0.001$ ,  $\eta^2 = .501$  'companion',  $F=6441$ ,  $p=0.004$ ,  $\eta^2 = .235$ , 'ungrateful',  $F = 33.05$ ,  $p < 0.001$ ,  $\eta^2 = .611$  'innovate',  $F= 40.971$ ,  $p < 0.001$ ,  $\eta^2 = .661$  'greedy',  $F = 26.142$ ,  $p < 0.001$ ,  $\eta^2 = .555$ . The result for the word 'expert',  $F= 1.676$ , is marginally significant,  $p = 0.05$ ,  $\eta^2 = .133$ . Finally, 'method' displays no statistically significant difference  $F= .361$ ,  $p = n.s.$ ,  $\eta^2 = .017$ . Most importantly, the overall results across all ten word forms shows that the effect of task assignment for the immediate post-test (between groups A, B and C) is very highly significant,  $F= 76.841$ ,  $p < 0.001$ ,  $\eta^2 = .785$ . Since the main effect here is very high and the p statistic very low (reflecting its high significance), this result lends credence to the hypothesis on L2 vocabulary initial learning.

Post hoc analyses using the Bonferroni post hoc criterion for significance display existence of significant differences for the three assigned tasks among number of the 10 researched word forms in the immediate post-test. In order to make visual inspection of these differences more straightforward., Table 1 was designed to show a vertical comparison of the means. Within each column, the superscript 'a' indicates a mean significantly lower than the one with 'b', and the one with 'b' is significantly lower than the one with 'c'. Any two means sharing the same superscript indicates that no significant differences were found from the post-

---

hoc analysis. Starting with the total column, it reveals that group C significantly outperformed group B, and that group B in turn significantly outperformed group A. However, this effect is not applicable in some cases when considering the words individually. While results for ‘fatal’, ‘portion’ ‘rove’, ‘greedy’, and ‘ungrateful’ were similar to the overall one, differences in ‘pale’ were found to be significant only between Group A and B but not between B and C. As for ‘companion’ and ‘expert’, group A did not differ significantly from group B, but it did differ from C.

Table 1 Comparison of means using the results of Bonferroni post-hoc test ( in the Initial learning experiment) (Mean  $\pm$  Std. Deviation)

Group	fatal	Pale	portion	Method	rove	companion	ungrateful	innovate	expert	greedy	total
A	1.73 $\pm$ 0.799 <sup>a</sup>	1.37 $\pm$ 0.704 <sup>a</sup>	1.87 $\pm$ 0.915 <sup>a</sup>	4.00 $\pm$ 1.134	2.13 $\pm$ 0.990 <sup>a</sup>	2.13 $\pm$ 1.125 <sup>a</sup>	2.53 $\pm$ 1.060 <sup>a</sup>	1.80 $\pm$ 0.775 <sup>a</sup>	3.20 $\pm$ 1.568 <sup>a</sup>	2.13 $\pm$ 0.915 <sup>a</sup>	23.27 $\pm$ 4.847 <sup>a</sup>
B	2.47 $\pm$ 0.990 <sup>b</sup>	3.00 $\pm$ 1.195 <sup>b</sup>	2.60 $\pm$ 1.242 <sup>b</sup>	4.13 $\pm$ 0.990	3.20 $\pm$ 1.521 <sup>b</sup>	3.00 $\pm$ 1.309 <sup>ab</sup>	3.33 $\pm$ 0.816 <sup>b</sup>	2.40 $\pm$ 0.910 <sup>b</sup>	3.80 $\pm$ 1.207 <sup>ab</sup>	3.33 $\pm$ 1.113 <sup>b</sup>	31.27 $\pm$ 4.682 <sup>b</sup>
C	4.53 $\pm$ 1.060 <sup>c</sup>	3.80 $\pm$ 1.521 <sup>b</sup>	4.67 $\pm$ 0.724 <sup>c</sup>	4.33 $\pm$ 1.113	4.73 $\pm$ 0.594 <sup>c</sup>	3.87 $\pm$ 1.506 <sup>c</sup>	4.87 $\pm$ 0.352 <sup>c</sup>	4.53 $\pm$ 0.915 <sup>b</sup>	4.40 $\pm$ 1.056 <sup>b</sup>	4.60 $\pm$ 0.737 <sup>c</sup>	44.33 $\pm$ 4.562 <sup>c</sup>
sig	0.000	0.000	0.000	.699	0.000	0.004	0.000	0.000	0.050	0.000	0.000

## Delayed Word Retention

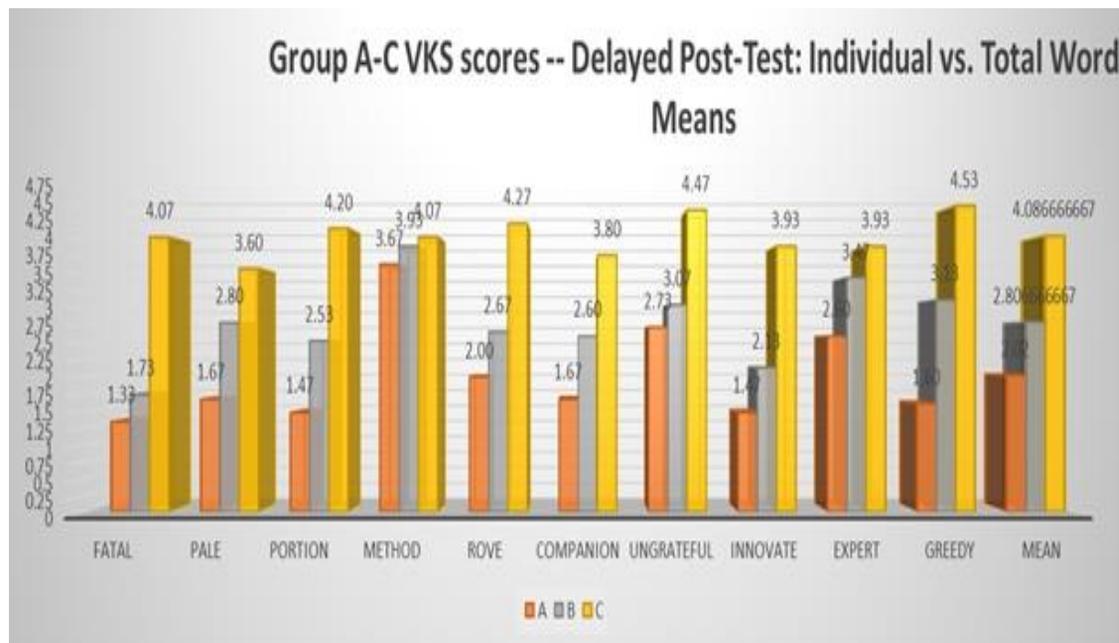


Figure 5: Group results in the delayed post-test for ‘fatal ... greedy’ and total word means for each group

Figure 5 describes the main descriptive statistics of the results for experimental groups A, B and C in the delayed word retention post-test. In summary, the results of the delayed word retention post-test are qualitatively similar to those of the immediate word learning post-test (when comparing the mean score of individual words with the average scores 2.02, 2.81 and 4.09 for groups A, B and C respectively). Scores of ‘Fatal’, ‘Portion’, ‘Rove’, ‘Ungrateful’ and ‘Innovate’ all increased in VKS in task 2 compared to task 1 (compare red and grey bars for each word form, respectively) whilst improvements gained when comparing tasks 2 and 3 are much greater (between nearly 1.5-2 on average, compare grey and yellow bars, respectively). For the remaining word forms, little or almost no improvement is evidenced based on task for ‘expert’ and ‘method’ (respectively), whilst ‘pale’, ‘greedy’ and ‘companion’ receive similar and more moderate gains based on the type of task assigned to learners.

In summary, for the majority of words studied in the delayed word retention part of the study, one of two patterns based on task assignment emerge; for 5/10 words, task 3 receives more notable gains than for either task 1 or 2, whereas for 3/10 words the gains made in tasks 2 and 3 compared to task 1 are fairly similar (for 2/10 words, more minor gains were achieved based on tasks assigned to learners).

Finally it is important to note that the one-week delay between the two post-tests has only induced about 10% drop in word retention for each task type, 2.33, 3.13 and 4.43 (immediate post-test) contra 2.02, 2.81 and 4.09 (delayed post-test) for each of the three tasks assigned. These differences translate to average mean distinctions of nearly 13.3% (task A), 10.2% (task B) and 5.5% (task C) in each of the two test scenarios reflecting immediate learning and medium-term retention.

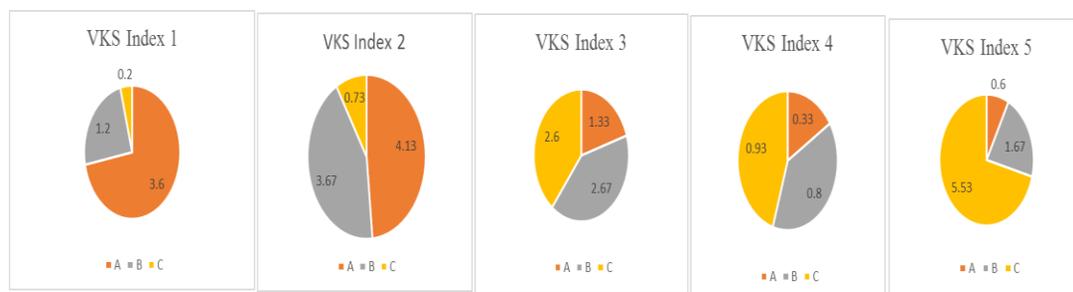


Figure 6: The results for VKS indexes 1-5 across groups A (orange), B (grey) and C (yellow) in the delayed post-test

The organisation of figure 6 is identical to that of figure 3. To summarise the comparison of the VKS indexes acquired by the participants in the delayed post-test, the trend is similar to that in the immediate post-test. When comparing the results as presented across indexes, the following average scores for familiar words (i.e. for indexes 3-5) are (0.75) for group A, (1.75) B and (3.02) C. It is thus clear that the self-report scores for familiar words for medium-term retention are similar as for initial learning, although the observed averages for groups A and B are somewhat lower while the average score for group C remains much the same.

---

## Statistical Comparison of the Delayed Post-test Results

To examine the effects of task-induced involvement load (independent variable) on participants' medium-term word retention (dependant variable); , the VKS results of the delayed post-test were formally tested by one-way ANOVA. Results of the individual words are displayed as follows: 'fatal',  $F= 64.761$  ,  $p < 0.001$ ,  $\eta^2= .755$ , 'pale',  $F= 8.575$ ,  $p < 0.002$ ,  $\eta^2 = .290$ , 'portion',  $F= 20.661$ ,  $p < 0.001$ ,  $\eta^2 = .496$ , 'rove',  $F= 15.754$ ,  $p < 0.001$ ,  $\eta^2 = .429$ , 'companion',  $F, 19.3$   $p < 0.001$ ,  $\eta^2 = .429$ , 'ungrateful',  $F=13.458$ ,  $p < 0.001$ ,  $\eta^2 = .391$ , 'innovate',  $F= 28.179$ ,  $p < 0.001$ ,  $\eta^2 = .573$  and 'greedy',  $F=41.012$ ,  $p < 0.001$ ,  $\eta^2 = .661$ . The result for the word 'expert'  $F= 4.595$  is also statistically significant at  $p < 0.02$  and  $\eta^2 = .191$ . Finally, 'method' displays no statistically significant difference in this context,  $F = .510$   $p = n.s.$ ,  $\eta^2 = .024$ . However, as for initial vocabulary learning, the differences between the total average of groups A, B and C are very highly significant,  $F= 46.534$ ,  $p < 0.001$ , and with 68% of the total variance is accounted for by the treatment effect.

The results of the Bonferroni post hoc test is displayed in Table 1. Similar to the one of the immediate post-test, the comparison post-hoc of the delayed post-test across the three groups reveals that the total average of the performance of Group C was significantly the best of all groups. However, the means of the retention of some words has more varied differences than in the initial learning post-hoc analysis. What is recognizable here is that while significant differences related to the word forms ('fatal' and 'rove') disappeared between groups A and B, differences with regard to 'greedy' remained significant between all the groups as they were in the initial learning of words.

### Implications of the Results

There are five key implications arising from the results presented outlined as follows:

a) The indexing of the task involvement load can be seen to be operationalised

---

through ‘need’, ‘search’ and evaluation’ as reflected in the assigned three tasks. Also, since the nature of the tasks themselves cannot be shown to require either more or less time or more word retrievals (i.e. compared to each other), only task-induced involvement load can explain these findings of vocabulary acquisition. Therefore, this study casts some doubt on Folse’s (2006) claims concerning depth of processing not being a factor in terms of vocabulary learning and that the number of word retrievals as considered alone can largely explain acquisition of new lexemes.

- b) The results clearly show that the nature of the task (involvement load) correlates with learners’ results, so that using new words in original sentences (task C) leads to better initial learning as well as better medium-term retention than doing different types of reading comprehension activities (tasks A and B). Also, the average of the initial learning of words per task one week after the immediate post-test drops the least in group C (5.5% vs. 10.2% and 13.3% in tasks 2 and 1, respectively). Overall, these findings suggest that depth of processing *is* a key issue in terms of learning new words for pre-intermediate stage learners, even when other theoretical and methodological factors are being held constant ( Laufer and Hulstijn, 2001, p. 15). More importantly, this finding supports Teng’s (2015, p. 94) claim, according to which

“...the present study suggests that task (sic) supplemented with word-focused activities that require high degrees of evaluation is beneficial for learners to elaborately process words, which in turn, improve their learning of unknown words. Another implication is that this study highlights the fact that, overall, a task with high involvement load is beneficial for learners” Teng (2015, p. 94)

Thus, since the key factors of ‘need’, ‘search’ (see e.g. Kim, 2008, Teng, 2015, Keating, 2008 and Hulstijn and Laufer, 2001) were of the same order as a

---

whole in tasks A, B, & C – but with evaluation being moderate in task B and strong in task C – it can be stated that these three key factors relating to word learning are theoretically intact. In other words, ‘need’, ‘search’ and ‘evaluation’ can be used to operationalize task-induced involvement load as a theoretical concept in vocabulary learning and retention.

- c) In terms of initial word learning, Kim's study (2008) did not result in significant differences between "gap-fill" and "reading comprehension" tasks (representing Tasks B and A respectively in the current research). Kim suggested that this might be due to a disparity in the degree of contribution (moderate and strong) of each component to the overall involvement load, proposing that "strong evaluation induces much greater involvement in processing a word than the moderate evaluation" when the other two components are held constant (2008, p. 125). In general, this claim seems to be at odds with the results of the current study, as significant differences were found in the overall results from the three tasks in both post-tests. However, if the results of individual words are examined, there are some cases in which the "gap-fill" group did not perform significantly better than the "reading comprehension" group (see “fatal” and “rove” in Table 1) in the delayed post-test, which lends some support to Kim's suggestion. What is notable about these two word forms is that they are from frequency levels higher than the other target words (5000 and 4000 level respectively). Thus, it can be suggested that "strong evaluation might be the most influential factor" (Kim, 2008, p. 125) for learner's medium-term retention of words where there is a frequency two or three levels higher than the learners' vocabulary knowledge level.
- d) Since a relatively large proportion of the participants (13 and 18 of 45) knew the meanings of the word forms ‘expert’ and ‘method’ (respectively) prior to the study being conducted, it is not surprising that these words receive the least amount of support in the statistical analysis presented. For example, the

---

minimum and maximum average scores of words recalled for these two words in either the immediate or delayed post-tests range from 2.6 to 4.4, whilst for the remaining 8 words range from 1 to 4.73. However, regarding 'expert' it was noticed that although a number of participants have indicated their previous knowledge of the word, some of them gave the translation of 'experience' instead. Thus, learners' previous knowledge of vocabulary items may distort the interpretation of results of this kind, while learners' previous knowledge of similar vocabulary items might affect the vocabulary learning and retention. Therefore, it is important to balance such aspects carefully in designing learning materials (Nation, 2001, 2013) and in conducting further research on similar topics.

- e) James (1890, p. 662) originally stated: “all improvement of the memory lies in the line of elaborating the associates”. Laufer and Hulstijn (2001) affirm that essentiality of this claim lies in that more and more attention should be paid to the formal and semantic properties of words and to the kinds of associations learners make in acquisition with existing words in their memory, (Kim, 2008) also asserts similar claims. For example, learners may be able to establish similarities and contrasts between existing and new forms of knowledge. In addition, Laufer and Hulstijn concede the significance of the motivation factor in promoting success in second language learning (Dörnyei and Ushioda, 2009). While the significance of the first two claims can be exemplified through two theoretical properties of the word forms investigated in this research (which are linked to concept of depth of processing), the significance of motivation is related to methodological aspects of the study. Firstly, since all the words were from three different parts of speech rather than from only a single part of speech (compared to Folsse's 2006 exclusive focus on verbs), it can be seen that the processing of the words as a whole required greater semantic and syntactic depth than processing of only one particular part of speech; after all, both the semantic and syntactic properties and syntactic ordering of different parts of

---

speech are constrained in English (Chomsky and Halle, 1968). Thus, the findings of this study can be seen to reflect task-induced involvement load more strongly in terms of lexical diversity and processing depth. Secondly, the same claim for complexity of processing and evaluation can be applied in which the selected words comprise variety of syllable structure; monosyllabic, disyllabic and trisyllabic lexemes with several types of sounds. Thus, whatever properties of the stimuli are considered, only depth of processing (reflecting high involvement load) can sufficiently explain the evidenced improvement in acquisition and lexical recall. Lastly, since all the 45 participants selected for this study were pre-intermediate learners studying an intensive 3-month course in English with several lessons received daily, this suggests that a high degree of motivation among learners is essential in participant recruitment (Dörnyei and Ushioda, 2009).

### **The Relationship of the Findings of this Research to Ones in Previous Literature**

Some previous research (Folse, 2006 and Keating, 2008) has suggested that the time spent studying new words (verbs) or the use of nonsense words may offer more suitable explanations for the key factor in acquisition, and to what extent, e.g. lexical frequency can be controlled.

However, despite these two issues, the findings of this study suggest that two other factors are much more important in terms of explaining vocabulary learning. The first factor lies on operationalising processing depth and to what extent it can be adequately expressed in relation to involvement load. That is, involvement load can be seen as theoretically more complex than merely comprising ‘need’, ‘search’ and especially ‘evaluation’ (Kim, 2008, p. 128). The Second factor refers to having a suitably strong focus on incidental learning (Schmidt, 2002 and Nation, 2013).

Although a suitable integration of the two approaches of incidental and intentional learning may be seen as an effective way to vocabulary acquisition in

---

vocabulary learning, the claims by Nation (2013) on incidental learning through suitably designed input remain important in this context:

“Each meeting provides a small opportunity to learn about the form, meaning and use of the word within the contexts in which it occurs. After the first meeting with the word, the learners may meet the word again in the book and when they do they have a chance to recall what they learnt from the previous meeting. This recall is probably done subconsciously” Nation, 2013, p. 2.

Firstly, this contention may be seen to be in contrast with the claims made by Schmidt (1994) on incidental vocabulary learning being defined as the learner having no strong intent to learn; and consequently conflicts with my claims. However, Nation (2013) does recognise that this process is almost certainly subconscious. Furthermore, Jiang (2000) has stated that learners first establish new L2 word entries in the L2 lexicon with initial access to formal information. This process is then followed by the semantic and syntactic specifications of the new word's L1 translation equivalent. It can thus be induced, that the meaning of any L2 word is acquired first in L1, even if the definition was provided in L2. Since the learners in this study are at a pre-intermediate stage in acquiring English, it is shown that adequately operationalizing key parameters for high involvement load in incidental learning is theoretically important. A similar claim might not be true for more advanced or highly proficient learners. However, such learners may have passed beyond the key stage referred to by Jiang (2000, pp. 51-53); according to this proposition, when learners become more experienced in using their L2 (this may transpire much later on in the language acquisition process), semantic, syntactic, morphological and formal information about L2 word are established in the entry, rather than merely referring to their L1 equivalents.

Since task 3 in the experiment designed for this study constitutes the kind of approach adhered to by Joe (1995, 1998), the claims made on task-induced involvement receive further support in this instance. It may (or may not) that

---

students in Joe's studies used more time and thus a larger number of word retrievals to complete their tasks. However time spent on a task and the number of word retrievals, though important criteria in acquisition, form only one part of acquiring lexical items. Task-induced involvement load, designing suitable tasks for learning as well as level of motivation (along with attitudes, see e.g. Gardner, 1985) may further affect the acquisition of new lexical items.

## **Assessing the Theoretical Validity of Operationalising Involvement Load**

### **Theory**

In their research on the applicability of involvement load as a theoretical construct in lexical learning, Laufer and Hulstijn (2001, p. 9) refer to several different issues that remain unresolved with respect to *what* aspects of cognitive-linguistic processing affect lexical access and retrieval. More importantly, a range of central theoretical issues relating to *how* these aspects of information processing in language acquisition are to be modelled remain elusive. For example, different aspects of explicit and implicit learning of lexical items do not appear to have been fully resolved yet, since it is not clear at what point learners begin referring to formal aspects of representation in their L2 learning. Similarly, Laufer and Hulstijn (2001) recognise the significance of different types of motivation in L2 acquisition and stress the qualitative differences in processing between different words as well as the distinctive ways of ‘searching’. It is the point of variation of levels of motivation that been dealt with in the design of the current study's tasks. Different kinds of motivation were built into the tasks, e.g. interest in the text, importance of the target words within the text, as well as the motivation resulted from the need to comply with the task requirements. Although Laufer and Hulstijn (2001) recognise the potential significance of having different levels in modelling depth of processing, little insight is provided into what this may involve in theoretical terms with respect to the degree/strength of the involvement load that participants are subjected to in different tests and experiments. Whether processing depth remains “deep” or “shallow” remains a debatable point of discussion, since we do not yet

---

have adequate technological ways of fully measure linguistic processing in this way. Thus, instead of modelling the key components of ‘need’, ‘search’ and ‘evaluation’ as a set of +/- variable values (cf. Laufer and Hulstijn, 2001, p. 18), it should prove more worthwhile to establish consensus on an accepted set of purely formal linguistic and/or cognitive properties that impinge upon ‘depth of processing’. For example, Dörnyei’s (1994) attempt to model motivation (reflecting ‘need’ in Laufer and Hulstijn, 2001) as a continuum ranging from the level of the learner to that of the learning context and the language learnt itself. Aspects of ‘evaluation’, in turn, might include different formal aspects of lexical representation, such as semantic, syntactic, phonological and morphological properties (see Jiang, 2000, p. 53), whereas ‘search’ could include external resources vs. internal ones (such as dictionaries contra teachers or fellow learners). Thus, it seems theoretically appealing to suggest that involvement load involves much more than a simple presence or absence of the three variables of ‘need’, ‘search’ and ‘evaluation’.

### **Evaluation**

It is crucial to investigate whether the way in which the experiment was designed would have served to distort or skew the results in some respect. In terms of focusing on pre-intermediate learners, this can be justified on two grounds as follows: firstly, such learners have not been the focus of previous research. Secondly, investigating the role and relevance of task-induced involvement load in the context of learners at an intermediate stage in their learning may provide a sound basis for further research on similar topics. After all, pre-intermediate learners are at a stage in learning where they will understand and be able to produce a large number of well-formed expressions and sentences in their L2, potentially even without frequent reference to their L1 and its translation equivalents in linguistic processing, for example. In this sense, focusing on learners at this stage in their L2 development may allow a more solid and theoretically appealing basis for further research on second language learning. That is, focusing only on beginners or

---

advanced/highly proficient learners would probably not allow to draw as theoretically engaging conclusions.

It was deemed necessary to ensure that the participants were not highly familiar with the new words. Although the pre-test performed with the other students at a similar level and the teacher suggest that the participants were not familiar with the majority of words, this cannot be proved beyond reasonable doubt, for two reasons: a) the other participants are different individuals and b) it is not possible to control whether participants receive exposure to any given word outside the classroom.

### **Pedagogical Implications**

The current findings have significant educational implications represented mainly in the possible instructional applicability of the Involvement Load Hypothesis. Teachers and instructors can manipulate and enhance their students' vocabulary learning through formulating tasks based on the claims of the hypothesis. Beside the central theoretical position of depth of processing in L2 lexical learning, Nation (2013) has also pointed towards the importance of having well designed materials (e.g. with respect to the number of newly encountered lexical items) and the extent to which extensive reading (whether linked skills, narrow or repeated reading) can help to maximise student learning. Thus, another potential issue remains central in pedagogical terms in this context. This issue is represented in the potential role of extensive reading and negotiation of meaning (see Nation, 2015 and e.g. Newton, 1995) in directing L2 students onward in their learning process in this respect. Considering this, it will be crucial to aptly balance explicit teaching and learning with implicit learning, in order to increase learner motivation and aptitude for L2 acquisition.

---

## **Suggestions for Further Research**

Several strands of further research have become apparent from the discussion of the results. First, from the above discussions on learner levels (whether beginner, intermediate or advanced) an interesting area for further research has become evident. Although Jiang (2000) suggests that the representation of L2 lexical information, in the learner's mind, is distinct from that in the L1, Jiang also recognises that extensive and contextualised exposure to the target language is essential in order to incorporate different types of information in L2 lexical entries. Thus, investigating such integration in intermediate learners of English (lower or upper-intermediate) and especially in the way their L2 is mentally represented will constitute a useful avenue for further research in L2 vocabulary acquisition. Thus, as also suggested by Kim (2008, p. 126) "It will be important for future studies to investigate a wider range of proficiency levels in a variety of educational contexts". That is the concept of task-induced involvement load might not have the same theoretical implications in all research and/or educational contexts.

Another important avenue for further research in L2 vocabulary acquisition is related to the optimal amount of time in the vocabulary tasks that may help learners to better retain lexical items in long-term memory. It is known from previous research on L1 acquisition that even the first language may be subject to linguistic attrition (Schmid, 2004). This was noticed in this study, represented in the erosion of some words a week after they were gained from tasks with approximately low involvement load. Therefore, the significance of the amount of repeated exposure to L2 lexical items and their formal and contextual properties will become more important if fluency and efficient usage of vocabulary in the L2 is to be maintained by L2 learners, especially for words that were not processed in high levels of depth at the initial learning. This proposed aspect of retention is very relevant to further research in the field of

---

task-induced involvement load theory, as it could help to show the theoretical significance of how word recall and retention as reflected through ‘need’, ‘search’ and ‘evaluation’ develop over time.

Finally, whether there are different levels in processing linguistic information remains a third area for further research. Hawkins (2003) shows that meaning remains the central property that listeners pay initial and most attention to in perceptual terms. Therefore, ascertaining in more detail to what extent reference to more formal properties of words remains essential for lexical acquisition by L2 learners. As for the previous potential areas of further research, this third area is closely related to deepening the understanding of operationalizing involvement load theory adequately.

---

## References

- Allan, D. (2004). *Oxford Placement Test 2: Test Pack*. Oxford University Press.
- Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly*, 43(1), 79-103.
- Beglar, D. (2010). A rasch-based validation of the vocabulary size test. *Language Testing*, 27(1), 101-118.
- Bell, Judith. (2010). *Doing Your Research Project*. Open University Press. Retrieved 18 October 2015, from <<http://www.mylibrary.com?ID=302868>>
- Bruton, A. (2009). The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly*, 6(4), 288-297.
- Cho, K. S., & Krashen, S. D. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, 662-667.
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York, Harper & Row.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. NJ: Lawrence Erlbaum Associates.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, 16(2), 227-252.
- Elgort, I. (2013;2012;). Effects of L1 definitions and cognate status of test items on the vocabulary size test. *Language Testing*, 30(2), 253-272.

- 
- Elley, W. (1991). Acquiring Literacy in a Second Language: The Effect of Book-Based Programs. *Language Learning*, 41(3), 375-411.
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 174-187.
- Ellis, N. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *Consciousness in second language learning*, 11, 37-57.
- Ellis, R. and He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21: 285-301.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language learning*, 44(3), 449-491.
- Eysenck, M. W. (1982). 'Incidental learning and orienting tasks' in C. R. Puff (ed.): *Handbook of Research Methods in Human Memory and Cognition*. New York: Academic Press. 197- 228.
- Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *Tesol Quarterly*, 40(2), 273-293.
- Gardner, R. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Arnold.
- Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21, 319–33.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31(3), 373-405.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. *Vocabulary and applied linguistics*, 113-125.
- Hulstijn, J. H., & Trompetter, P. (1998). Incidental learning of second language

- 
- vocabulary in computer-assisted reading and writing tasks. In D. Albrechtsen, B. Henriksen, I. M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy* (pp. 191–200). Odense, Denmark: Odense University Press.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied linguistics*, 21(1), 47-77.
- Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11(2), 149-158.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied linguistics*, 19(3), 357-377.
- Keating, G. D. (2008). Task effectiveness and word learning in a second language: The Involvement Load Hypothesis on trial. *Language Teaching Research*, 12(3), 365-386.
- Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language learning*, 58(2), 285-325.
- Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. In Foster Cohen, S.H., Gracia-Maya, M. and. Cenoz, J. (eds.), *Eurosla Yearbook*, 5, 223-250, Amsterdam: Benjamins.
- Laufer, B., & Hill, M. (2000). What Lexical Information Do L2 Learners Select in a CALL Dictionary and How Does It Affect Word Retention? *Language Learning & Technology*, 3 (2) (January 2000): 58-76.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics*, 22(1), 1-26.

- 
- Lightbown, P. M. & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.
- Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 44(1), 87.
- Martínez-Fernández, A. (2008). Revisiting the Involvement Load Hypothesis: Awareness, Type of Task and Type of Item. *Selected Proceedings of the 2007 Second Language Research Forum*, edited by Bowles, M., Foote, R., Perpiñán, S, and Bhatt, R., Somerville, MA: Cascadilla Proceedings Project, 210-228.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82.
- Nation, I. S. P. (2013a). Designing reading tasks to maximize vocabulary learning. *Applied Research on English Language*, 3(1), 1-8.
- Nation, I.S.P. (2013b). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 2-13.

