THE EFFECT OF TASK-INDUCED INVOLVEMENT LOAD ON INCIDENTAL VOCABULARY ACQUISITION AMONG ELT STUDENTS

TUNCAY KARALIK

(Yüksek Lisans Tezi)

Ocak 2016

THE EFFECT OF TASK-INDUCED INVOLVEMENT LOAD ON INCIDENTAL VOCABULARY ACQUISITION AMONG ELT STUDENTS

TUNCAY KARALIK

MA Thesis

Program in Foreign Language Education

MA in English Language Teaching Program

Supervisor: Assist. Prof. Dr. Ali MERÇ

Eskişehir

Anadolu University Graduate School of Educational Sciences

January, 2016

"This MA thesis has been supported by Scientific Research Unit of Anadolu University. Project No: 1508E604"

JÜRİ VE ENSTİTÜ ONAYI

Tuncay KARALIK'ın "THE EFFECT OF TASK-INDUCED INVOLVEMENT LOAD ON INCIDENTAL VOCABULARY ACQUISITION AMONG ELT STUDENTS" başlıklı tezi 22.01.2016 tarihinde, aşağıda belirtilen jüri üyeleri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği programı yüksek lisans tezi olarak değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	**İmza**
Üye (Tez Danışmanı)	: Yard.Doç.Dr. Ali MERÇ	.....
Üye	: Prof.Dr. Zülal BALPINAR	.....
Üye	: Prof.Dr. Gül DURMUŞOĞLU KÖSE	.....
Üye	: Doç.Dr. Belgin AYDIN	.....
Üye	: Doç.Dr. İlknur SAVAŞKAN	.....

Prof.Dr. Esra CEYHAN
Anadolu Üniversitesi
Eğitim Bilimleri Enstitü Müdürü

ABSTRACT

THE EFFECT OF TASK-INDUCED INVOLVEMENT LOAD ON INCIDENTAL VOCABULARY ACQUISITION AMONG ELT STUDENTS

TUNCAY KARALIK

Anadolu University Graduate School of Educational Sciences

Program in Foreign Language Education – MA in English Language Teaching

January, 2016

Supervisor: Assist. Prof. Dr. Ali MERÇ

Vocabulary knowledge is at the heart of language learning as both comprehending and producing messages depend on knowing words. Second/foreign language learners need to acquire a great deal of vocabulary in a limited time to be successful in nearly all aspects of the target language Therefore, a large body of research has been devoted to investigating the best ways to help learners in their endeavour.

In recent years, incidental acquisition of vocabulary has been found to be beneficial in expanding vocabulary knowledge. In this regard, how incidental vocabulary can be promoted is an intensely investigated subject. Incidental vocabulary acquisition through reading texts has been found to be facilitated by additional vocabulary tasks drawing the learners' attention to specific words. Therefore, designing the best vocabulary tasks has been an important issue for teachers. In order to address this concern, Laufer and Hulstijn (2001) proposed the Task-induced Involvement Load Hypothesis positing that the absence or presence of three components (evaluation, search and need) in a task determines the overall involvement load level induced by the task. Evaluation is related to learners' judgement on whether a specific word fits into a

context. Search explains whether the meanings of words are searched by consulting dictionaries or other sources. Need is the motivational dimension of the construct and explains whether the drive to learn the word is extrinsically or intrinsically exposed.

Using the framework of Laufer and Hulstijn (2001), this study aims to investigate the effect of four vocabulary tasks with varying distributions of evaluation, search and need components on incidental vocabulary gain and retention. The subjects were 139 first year ELT students from eight intact classes taking Academic Reading Course in the fall semester of 2015-2016. The subjects were assigned to four groups randomly: fill-in with glossary, fill-in by searching, retelling with glossary and retelling by searching. Two different texts were designed for four tasks. Each text contained 10 target words accompanied by reading comprehension questions. Prior to the implementation of the tasks, the participants' familiarity with the target words was tested through a pre-test which required providing the L1 equivalent or English explanations of the target words. The groups were to complete the reading comprehension tasks and the vocabulary tasks they were assigned to. The fill-in groups completed the gaps in the text with the target words, with or without glossary. The retelling groups wrote their own versions of the text by incorporating the target words, with or without glossary. Once the tasks were completed, the worksheets were collected and an unannounced post-test similar to the pre-test was administered. Similarly, two weeks later, an unannounced delayed post-test was conducted. For marking, each correct answer was counted as 1 and the incorrect ones as 0. Partially correct answers were awarded 0.5 point.

The comparison of the groups on immediate and delayed post-tests using one-way ANOVAs showed that the tasks with higher level of involvement load yielded higher vocabulary gain and retention. However, the only significant differences were between retelling by searching and fill-in groups on the delayed post-test, which provided partial support for the Task-induced involvement load hypothesis.

Keywords: incidental vocabulary acquisition, involvement load level, task-induced involvement load, vocabulary gain and retention, vocabulary tasks

ÖZET

GÖREV KAYNAKLI KATILIM YÜKÜNÜN İNGİLİZCE ÖĞRETMENLİĞİ BÖLÜMÜ ÖĞRENCİLERİNİN RASTLANTISAL KELİME ÖĞRENİMİ ÜZERİNDEKİ ETKİSİ

TUNCAY KARALIK

Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü

Yabancı Diller Eğitimi Anabilim Dalı-İngilizce Öğretmenliği Programı

Ocak, 2016

Danışman: Yard. Doç. Dr. Ali MERÇ

Dışarıdan gelen mesajları anlamak ve mesaj üretmek kelime bilgisine bağlı olduğu için, kelime bilgisi dil öğrenmenin merkezinde yer almaktadır. Yabancı dil öğrenenler dil öğrenme sürecinin her aşamasında başarılı olabilmek için kısa bir süre içerisinde yüksek sayıda kelime öğrenmek durumundadır. Bu bağlamda dil öğrenenlere yardımcı olabilmek için en iyi kelime öğrenme ve öğretme yollarını araştıran çok sayıda çalışma yapılmıştır.

Son yıllarda rastlantısal kelime öğreniminin kelime dağarcığını artırmada faydalı olduğu bulunmuştur. Bu noktada rastlantısal kelime ediniminin nasıl hızlandırılacağı sıkça çalışılan bir konu olmuştur. Okuma parçaları yoluyla rastlantısal kelime öğreniminin, öğrencinin dikkatini belli kelimelere çeken kelime görevleri vasıtasıyla hızlandırıldığı sıkça ifade edilmiştir. Bu nedenle en iyi kelime öğrenme etkinliklerini tasarlamak, öğretmenler açısından önemli hale gelmektedir. Bu doğrultuda, Laufer ve Hulstijn (2001) herhangi bir kelime görevinde varlığı ve yokluğu o görevin katılım yükü düzeyini belirleyen üç bileşenli (değerlendirme, arama ve ihtiyaç) "Görev Kaynaklı Katılım Yükü Hipotezi"ni ortaya atmışlardır. Buna göre, katılım düzeyinin

kelime öğrenimini etkilediği varsayılmıştır ve daha yüksek seviyelerin daha fazla kelime öğrenmeyi sağlayacağı düşünülmektedir. Değerlendirme, öğrencinin bir kelimenin belli bir bağlama uygun olup olmadığı konusunda karar vermesidir. Arama, öğrencinin bir kelimeyi bir kaynakta arayıp aramamasıyla ilgilidir. İhtiyaç ise, öğrencinin bir kelimeye neden ihtiyacı olduğuyla ilgilidir.

Bu çalışma, Laufer ve Hulstijn (2001)'in önerdiği çerçeve kapsamında değerlendirme, arama ve ihtiyaç bileşenlerinin farklı dağılımlarını içeren dört kelime görevinin, kelime edinimi ve kalıcılık üzerine etkisini ölçmeyi amaçlamıştır. Çalışmaya 2015-2016 akademik yılı güz döneminde Anadolu Üniversitesi Eğitim Fakültesi İngilizce Öğretmenliği programında verilen Akademik Okuma dersini almakta olan sekiz ayrı şubeden 139 öğrenci katılmıştır. Katılımcılar sözlük kullanarak boşluk doldurma, sözlükçe yardımıyla boşluk doldurma, sözlük kullanarak yeniden yazma ve sözlükçe yardımıyla yeniden yazma olarak dört farklı etkinlikten birine rastgele atanmıştır. Her biri 10 hedef kelime ve okuduğunu anlama soruları içeren iki okuma parçası, dört farklı etkinliğe göre tasarlanmıştır. Öğrencilerin hedef kelimeleri bilip bilmedikleri, kelimelerin Türkçe veya İngilizce karşılıklarının sorulduğu bir ön test yardımıyla belirlenmiştir. Gruplar okuduğunu anlama sorularını yanıtlama ve kendilerine verilen kelime etkinliklerini bitirmekle görevlendirilmiştir. Boşluk doldurma grupları, verilen parçalardaki boşlukları hedef kelimeleri kullanarak doldururken, yeniden yazma grupları hedef kelimeleri kullanarak parçaları yeniden yazmıştır. Etkinlikler tamamlandıktan sonra çalışma kâğıtları toplanmış ve haber verilmeksizin ön teste benzer bir son test uygulanmıştır. Aynı şekilde, iki hafta sonra, kalıcılık testi uygulanmıştır. Öğrencilerin testlerde vermiş olduğu her doğru cevap 1 ve kısmen doğru cevaplar 0,5 puan olarak hesaplanırken yanlış cevaplara puan verilmemiştir.

Grupların son test ve kalıcılık testlerinde aldıkları puanların Tek Yönlü ANOVA istatistiksel yöntemi kullanılarak karşılaştırılması sonrasında daha yüksek katılım yükü içeren görevlerin hem kelime edinimine hem de kalıcılığa daha yararlı oldukları ortaya çıkmıştır. Ancak, tek anlamlı fark sözlük kullanarak tekrar yazma ve boşluk doldurma gruplarının kalıcılık puanları arasında bulunmuştur. Bu yönüyle çalışma "Görev Kaynaklı Katılım Yükü Hipotezi" için kısmi destek sağlamıştır.

Anahtar Kelimeler: rastlantısal kelime öğrenimi, katılım yükü, görev kaynaklı katılım yükü, kelime edinimi ve kalıcılığı, kelime etkinlikleri

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to many people but for whom this thesis wouldn't have been written.

Firstly, I would like to express my sincere thanks to my thesis supervisor Assist. Prof. Dr. Ali MERÇ for being patient and attentive through all the processes of material selection, implementation and reporting the findings. Without his constructive feedback and guidance, this thesis wouldn't have been realised.

I also would like to express my gratitude to my thesis committee, Prof. Dr. Zülal BALPINAR, Prof. Dr. Gül Durmuşoğlu KÖSE, Assoc. Prof. Dr. Belgin AYDIN and Assoc. Prof. Dr. İlknur SAVAŞKAN for their invaluable feedback.

I would like to thank to Assist. Prof. Dr. Hasan ÇEKİÇ, Assist. Prof. Dr. Hülya İPEK, Dr. Müge KANATLAR, Zuhal KIYLIK, Samet DENİZ, Aslı AKÇAY, Merve BAKKAL and Sibel ATAY for their understanding and allowing me their course hours to conduct the implementation.

I would like to thank to my office mate, Yusuf ÖZTÜRK for his support during the implementation and proofreading processes and also for his technological guidance.

Finally, I have to present my sincere thanks to TÜBİTAK for their contribution to the improvement of science by funding the future scientists.

Tuncay KARALIK

Eskişehir, 2016

TABLE OF CONTENTS

ABSTRACT iv

ÖZET vi

ACKNOWLEDGEMENTS ix

LIST OF TABLES xiii

LIST OF ABBREVIATIONS xiv

CHAPTER 1. INTRODUCTION 1

1.1. Statement of the Problem 5

1.2. Aims and Research Questions 7

1.3. Significance of the Study 8

CHAPTER 2. REVIEW OF LITERATURE 10

2.1. Introduction 10

2.2. Vocabulary Knowledge and Teaching of Vocabulary 10

2.2.1. Intentional/Incidental Leaning of Vocabulary 11

2.3. Task-induced Involvement Load Hypothesis 13

2.3.1. Internal Structure of Task-induced Involvement Load 15

2.3.2. Important Considerations for Involvement Load Studies 16

2.4. Review of Related Studies on Task-induced Involvement Load Hypothesis 18

2.4.1 Effect of Evaluation Component 19

2.4.2. Effect of Evaluation and Search Components 22

2.4.3. Path for Further Verification of Task-induced Involvement Load Hypothesis 26

CHAPTER 3: METHODOLOGY 27

3.1. Introduction 27

3.2. Participants 27

3.3. The Research Context 28

3.4. Instruments 28
3.4.1. Reading Texts 29
3.4.1.1 Criteria for Selecting the Texts and Target Words 29
3.4.2. Vocabulary Tasks 33
3.4.2.1. Fill-in by Searching 34
3.4.2.2. Fill-in with Glossary 34
3.4.2.3. Retelling by Searching 34
3.4.2.4. Retelling with Glossary 35
3.4.3. Vocabulary Tests 36
3.5. Procedures 36
3.5.1. Consenting 37
3.5.2. Piloting 37
3.5.3. Implementation 37
3.6. Analysis of the Data 39
CHAPTER 4: RESULTS AND DISCUSSION 40
4.1. Introduction 40
4.2. Results 40
4.2.1. Tasks with Different Involvement Load Levels 41
4.2.1.1. The First Implementation 41
4.2.1.2. The Second Implementation 44
4.2.2. Tasks with Equal Involvement Load Levels 47
4.2.2.1. The First Implementation 47
4.2.2.2. The Second Implementation 48
4.3. A Brief Summary of the Results 49
4.4. Discussion of the Findings 50
4.4.1. The Effect of Tasks with Different Task-induced Involvement Loads 51

4.4.2. The Effect of Tasks with Equal Task-induced Involvement Loads 58
CHAPTER 5: CONCLUSION AND IMPLICATIONS 61
5.1. Introduction 61
5.2. An Overview of the Present Study 61
5.2.1. Tasks with Different Task-induced Involvement Load Levels 62
5.2.2. Tasks with Similar Task-induced Involvement Load Levels 63
5.2.3. Summary 63
5.3. Implications 65
5.4. Suggestions for Further Studies 70
REFERENCES 72
APPENDICES 85
Appendix I. The First Text 85
Appendix II. The Second Text 88
Appendix III. Fill-in by Searching for the First Text 92
Appendix IV. Fill-in by Searching for the Second Text 95
Appendix V. Fill-in with Glossary for the First Text 98
Appendix VI. Fill-in with Glossary for the Second Text 101
Appendix VII. Retelling by Searching for the First Text 105
Appendix VIII. Retelling by Searching for the Second Text 108
Appendix IX. Retelling with Glossary for the First Text 111
Appendix X. Retelling with Glossary for the Second Text 115
Appendix XI. Vocabulary Tests for the First Text 119
Appendix XII. Vocabulary Tests for the Second Text 120
Appendix XIII. Consent Form for the Instructors 121

LIST OF TABLES

Table 1. Components of Task-induced Involvement Load..... 16
Table 2. Task-induced Involvement Levels of Tasks in Hulstijn & Laufer (2001)..... 18
Table 3. Distribution of Components in Sarani et al. (2013)..... 23
Table 4. Frequency and Percentage of Encounters of Target Words in COCA 31
Table 5. The tasks and Total Task-induced Involvement Load Levels 33
Table 6. Number of the Participants across Tasks and Implementation..... 35
Table 7. Implementation Schedule 38
Table 8. Immediate Vocabulary Gain Scores for the First Implementation 41
Table 9. One-way ANOVA for Immediate Post-test Scores of the First Implementation 42
Table 10. Vocabulary Retention Scores for the First Implementation 43
Table 11. One-way ANOVA for Delayed Post-test Scores of the First Implementation43
Table 12. Immediate Vocabulary Gain Scores for the Second Implementation 44
Table 13. One-way ANOVA for Immediate Post-test Scores of the Second Implementation 45
Table 14. Vocabulary Retention Scores for the Second Implementation..... 46
Table 15. One-way ANOVA for Delayed Post-test Scores of the Second Implementation 46
Table 16. The Mean Scores of Groups across the Implementations 49

LIST OF ABBREVIATIONS

ELT: English Language Teaching

FBS: Fill-in by Searching

FWG: Fill-in with Glossary

RBS: Retelling by Searching

RWG: Retelling with Glossary

TILH: Task-induced Involvement Load Hypothesis

TILL: Task-induced Involvement Load Level

TW: Target Word

VG: Vocabulary Gain

VR: Vocabulary Retention

CHAPTER 1. INTRODUCTION

A person grows with his vocabulary. Concepts broaden and become more complex as an individual acquires new, more precise terms in which to express them.

Charles Croll (1971: 378)

In its simplest definition, knowing words or vocabulary refers to being able to identify the form and meaning of a word according to Nation (2001) and as Croll (1971) stated in the excerpt, knowledge of vocabulary is at the very core of understanding the concepts around thus helping to express what is present in the environment. In this sense, developing a larger body of vocabulary knowledge can be accepted as the key for expressing one's ideas more easily and give meaning to the messages coming from the environment.

Folse (2006) underlined that having a good deal of vocabulary knowledge is a prerequisite for second language learners to perform well in four main skills: listening, reading, writing, and speaking. In other words, lack of vocabulary knowledge will lead to unexpected consequences such as communication breakdowns, misunderstanding of the message received and more importantly difficulties in expressing oneself. Nation (2001) also suggests that whether receptive or productive, all language skills require learners to know words either to comprehend messages or to express themselves

The fact that vocabulary knowledge affects language learners' performance in main skills becomes more salient when studies aiming to uncover the relationship between lexical knowledge and competence in four main skills are reviewed. Starting with the reading skill, a detailed review by Jeon and Yamashita (2014) demonstrates that success in reading comprehension can be attributed to many factors such as L2 decoding, L2 grammar knowledge, L1 reading comprehension, L2 phonological knowledge, L2 orthographic knowledge, L2 morphological knowledge, word reading, and metacognition. Vocabulary knowledge is also among these factors and it has always been found to be a significant contributor to reading comprehension. Studies on the effect of vocabulary knowledge on reading comprehension have generally focused on the correlational relationship between the size of vocabulary language learners have and

reading comprehension scores (Grasparil & Hernandez, 2015; Kharaghani & Ghonsooly, 2015; Ma & Lin, 2015; Mokhtari & Niederhauser, 2013; Price, Meisinger, Louwerse & D'Mello, 2015; Quinn, Wagner, Petscher & Lopez, 2015; Strasser, & Rio, 2014; Sidek & Rahim, 2015; Silverman, Proctor, Harring, Hartranft, Doyle & Zelinke, 2015; Tighe & Schatschneider, 2014; Zhang, McBride-Chang, Wong, Tardif, Shu & Zhang, 2014). The unanimous finding of the studies on reading comprehension and vocabulary knowledge was that the relationship between these two constructs was strong with vocabulary knowledge contributing significantly to reading comprehension regardless of both the context the studies were conducted and the covariates tested.

Research has also focused on the relationship between vocabulary knowledge and listening comprehension in L2. Similar to the ones regarding reading comprehension, studies on listening also highlighted the significant relationship between vocabulary knowledge along with other variables such as morphosyntactic skills, word reading, phonological knowledge and listening comprehension (Babayigit, 2014; Florit, Roch & Levorato, 2014; Kim, 2015; Ling, 2015; Vandergriff & Baker, 2015; Wang, 2015). The results of the mentioned studies underlined that decoding oral input was highly affected by vocabulary knowledge capacity. Another perspective in the area of vocabulary regarding its effect on listening was participants' perceptions of the difficulties while listening to a listening text. Generally, learners were found to be experiencing vocabulary related difficulties such as giving up listening and getting discouraged in listening because of encountering too many unknown words (Hamouda, 2013; Nowrouzi, Tam, Zareian & Nimehchisalem, 2015; Solak & Altay, 2014).

In addition to the studies on the effect of vocabulary knowledge on comprehending messages, there are also studies exploring the effect of vocabulary knowledge on language learners' success in producing written and oral language. To this end, the vocabulary stock of language learners was found to be a direct determiner of writing difficulties learners encounter as the studies showed that In cases where the vocabulary knowledge stock of the learners was found to be deficient, difficulties in arranging the ideas and word choice emerged (Al Seyabi & Tuzlukova, 2014; Begriche, 2013; Bouhitem, 2013; Huy, 2015; Putra, 2014; Rudy, 2013; Sawaki, Quinlan & Lee, 2013; Sofian & Salam, 2015; Yang, 2015).

The other productive skill, speaking, is also highly affected by vocabulary stock language learners have. Similar to the studies regarding writing, research has also focused on the difficulties learners have while speaking due to lack of vocabulary knowledge (Al Nouh, Abdul Kareem & Taqi, 2015; Khotimah, 2014; Perez Manzanilla & Diaz Cabrera, 2014; Tahir, 2015; Wahyuni, Basri & Mushuri, 2014) such as hesitations, silence for too long trying to decide which words to use; and the positive relationship between vocabulary size and speaking performance (Anova, Antoni & Kasyulita, 2015; Fhonna, 2014; Nurfuah, 2013).

In sum, vocabulary knowledge plays a direct role in comprehension and production of language in a second/foreign language. First, vocabulary knowledge is a strong predictor of competence regardless of the context and the covariates under control. Second, lack of vocabulary knowledge results in difficulties for learners to use the language to convey messages both orally and in written forms.

The problems encountered by learners in four main skills have encouraged researchers to determine how many words learners need to know in order to overcome these difficulties. To this end, researchers have been trying to come up with some threshold levels of vocabulary knowledge to help language learners succeed in basic skills. Among these researchers, Laufer and Nation (1995) posited that 95% coverage of the lexical items in a written discourse allows learners to comprehend the passage. Later, Hu and Nation (2000) found that 98% of the words in a written discourse had to be known for a reader to comprehend the text. Similarly, Adolphs and Schmitt (2000) suggested that around 96% of the words in a spoken discourse should be known to maintain daily conversations and Kaneko 2015 found that 96% lexical coverage was needed for comprehending academic spoken English. What these coverage sizes mean is that, learners have to master a great number of words for understanding the language in these contexts. For example, Schmitt (2000) advocated that approximately 3000 word families were required to read authentic texts and 10000 words to comprehend academic texts. In another study, Schmitt (2008) suggested that 5000-7000 word families had to be known for success in oral discourse. In a more recent study, Webb (2013) discovered that knowing 5000 to 13000 word families will provide learners with a coverage of 98% in academic spoken English. In a similar recent study, Kaneko (2015) tried to calculate how many word families have to be known for a coverage of

96% of academic spoken discourse and the analysis showed that knowing the most frequent 6000 word families resulted in slightly above the needed coverage. These numbers of vocabulary knowledge required for competence in basic skills pose a great learning challenge. The crucial role vocabulary knowledge plays in determining success in promoting competence in the four skills led researchers to acknowledge the importance of teaching vocabulary as Karbal (1975) suggested, teaching of vocabulary is one of the first steps to take in teaching four main skills to language learners.

However, learning vocabulary is not the same for native speakers of a language and those who learn that language as a foreign language. Regarding this issue Folse (2006) states that native speakers of a language acquire the needed vocabulary for expressing themselves and giving meaning to the incoming messages by being exposed to the natural linguistic stimuli in their environment, which is not the case for foreign language learners. Most of the time, the only chance for language learners to be exposed to the input in the target language is their classes, where the exposure to the input is too small when compared to the opportunities native speakers have. The important role played by vocabulary knowledge in determining the success in four skills along with high numbers of words to be mastered and the limited exposure to natural input makes the learning of vocabulary a big challenge for language learners. As Schmitt (2008) puts it "taken together, this amounts to a substantial lexical learning challenge, one which many/most learners fail to meet" (p. 329).

Addressing this challenge, Paribakht and Wesche (1999) maintained that teaching vocabulary in a way that is similar to how native speakers acquire lexis is not feasible for language teaching programmes for two reasons: a) language teaching programmes aim to equip the learners with the needed vocabulary in a limited time, b) the vast amount of input present in natural settings does not exist in classroom settings. What they actually put forward is that since language learners do not have the opportunities native speakers have in terms of exposure to input, teachers have to compensate for this lack of exposure by seeking ways to help learners acquire the needed vocabulary through manipulation of the classroom practices.

To this end, the ways to foster vocabulary knowledge expansion should be investigated to help language learners catch up with the requirements of main skills in a limited time, which is usually confined to classroom hours.

1.1. Statement of the Problem

The role of vocabulary knowledge in four main skills is a highly salient one. Since foreign language learners are deprived of the large amount of input to be exposed to out of classes, opportunities for the learners to gain high numbers of vocabulary have to be provided by teachers. Therefore, ways to foster vocabulary acquisition has been the focus of many researchers. Mainly, studies have focused on what strategies learners use to acquire vocabulary and proposed strategies conducive to vocabulary acquisition (Croll, 1971; Fan, 2003; Fraser, 1999; Gu & Johnson, 1996; Harmon, Buckelew-Martin & Wood, 2010; Lawson & Hogben, 1996; Little & Kobayashi, 2015; Mallet, 1975; Namaghi & Malekpur, 2015; Pflaum, 1973; Readence & Searfoss, 1980; Sagarra & Alba, 2006).

According to Nagy and Herman (1987), however, such direct instruction aiming to expand vocabulary knowledge of the language learners may not be the only and best source for vocabulary expansion as their study suggested that most of the vocabulary stock their participants had were not taught in the class by applying word learning strategies such as rehearsal and mnemonics. Nagy and Herman instead acknowledged the value of reading in and out of class in leading to vocabulary learning. This kind of vocabulary learning by engaging in reading and listening is incidental learning and it involves picking up words without paying special attention to commit the words into long-term memory (Schmidt, 1994; Gass, 1999). Following the influential proposition of Nagy and Herman (1987), most researchers held that incidental acquisition of words through processing the input is a valuable resource for expanding vocabulary knowledge among these are Brown (2000), Day and Bamford (1998) and Krashen (1989). However, because of the unforeseen nature of incidental vocabulary acquisition, expecting the learners to pick up substantial and specific vocabulary is not an appropriate step to take within L2 pedagogy as Paribakht and Wesche (1997) suggested:

> "It appears from research on incidental vocabulary acquisition that if systematic development of L2 vocabulary is desired, it cannot be left to the students themselves. They cannot be expected to "pick up" substantial or specific vocabulary knowledge through reading exposure without guidance. Given the limitations of decontextualized vocabulary instruction, the question for L2 pedagogy is whether L2 vocabulary acquisition can be enhanced through instructional intervention in the context of meaningful language use."(p. 177).

Here, 'instructional intervention' can be operationalised through various ways the most important one being tasks defined as "an activity or an action which is carried out as the result of processing or understanding language" (Richards, Platt, Weber, & Inman, 1986: 289). Later, Paribakht and Wesche (1997; 1999) acknowledged the importance of meaning-based vocabulary tasks by stating that when learners' attention is drawn to the form and meanings of specific words (selected as targets by the teacher) through vocabulary tasks, the chances that these words will be retained will increase.

Many studies investigated whether reinforcing the reading texts and leading the learners to attend to specific vocabulary items by designing some vocabulary tasks had an effect on incidental vocabulary acquisition (Folse, 2006; Hulstijn, Hollander & Greidanus, 1996; Joe, 1998; Min, 2008; Pichette, de Serres & Lafontaine, 2011; Peters, 2012; Pichette, de Serres, & Lafontaine, 2011; Rassaei, 2015; Rott, Williams & Cameron, 2002; Sonbul & Schmitt, 2009; Wesche & Paribakht, 2000). The findings of the mentioned studies proved that getting the students involved in an additional task was more beneficial than reading only condition in terms of incidental vocabulary acquisition. Therefore, the proposition Paribakht and Wesche (1997; 1999) held that incidental vocabulary acquisition through reading is better when learners are engaged in additional tasks was proven.

The fact that getting learners involved in additional vocabulary tasks leads to acquisition of more vocabulary raises an important concern: the difficulty in designing and choosing the best tasks for bringing about the highest vocabulary gain (VG) and vocabulary retention (VR). In this regard, analysing vocabulary tasks in terms of their requirements and their potential contribution to promoting vocabulary acquisition is an

essential process for teachers and researchers. One such attempt was Task-induced Involvement Load Hypothesis (TILH) proposed by Laufer and Hulstijn (2001). Laufer and Hulstijn posited that existence or absence of three components of evaluation, search and need in a task determines the overall Task-induced Involvement Load Level (TILL), which in turn determines how cognitively loaded a task is (see section 2.3. for a detailed account of the TILH). Their assumption is that increasing the level of these components in a specific task makes it more cognitively loaded. Worded differently, the amount of incidental VG and VR is assumed to be contingent upon the overall TILL induced by a task from the learners, with higher loaded tasks assumed to be more beneficial than lower loaded tasks.

By using the framework of TILH, it is possible for teachers, researchers and material designers to design several different vocabulary tasks with different levels of cognitive requirements. Therefore, utilizing the framework of TILH will make it possible to design the most appropriate task for increasing incidental vocabulary acquisition in a specific context.

1.2. Aims and Research Questions

The framework of task-induced involvement load hypothesis (TILH) assumes that tasks with higher presence of search, evaluation and need components will bear higher involvement loads. The higher involvement load a task bears, the higher vocabulary gain (VG) and vocabulary retention (VR) rates this task will lead to. Designing several vocabulary tasks with different task-induced involvement load levels (TILL), the present study aims to explore whether the assumptions of the construct are valid in determining the most successful tasks in promoting vocabulary acquisition as well as to find out the most beneficial task to promote incidental vocabulary acquisition among English Language Teaching (ELT) students. The research questions to be answered at the end of the present investigation are:

1) Do different vocabulary tasks with varying levels of involvement load differ in terms of their contribution to vocabulary gain and retention?

2) Do different vocabulary tasks with similar levels of involvement load operationalised by different components (search and evaluation) lead to similar amounts of vocabulary gain and retention?

1.3. Significance of the Study

The fact that vocabulary knowledge is a prerequisite for being successful in four skills makes teaching of vocabulary an important part of language teaching programmes. However, the needed numbers of vocabulary and the limited time for reaching these high numbers pose a great task for both teachers and learners to accomplish. In this regard, investigating the best ways to boost vocabulary acquisition to equip the learners with the needed vocabulary stock in a short time is of crucial importance for language teaching programmes.

As Paribakht and Wesche (1997; 1999) underlined, the deficiencies of direct vocabulary instruction through decontextualized drills in promoting vocabulary acquisition have led many to go for incidental acquisition of vocabulary through contextualised and natural input. However, as incidental way of vocabulary acquisition through reading is an unpredictable process in which there is no control over what is acquired and how much time it takes to acquire that piece of vocabulary, ways to provide guidance for the learners have been investigated intensely. The nearly unanimous finding of these enquiries is that incidental vocabulary acquisition through reading is best achieved by designing reading based tasks which require learners to pay attention to specific words. The proven value and usefulness of vocabulary tasks in drawing the students' attention to the forms and meanings of specific target words (TW), thus leading to considerable VG and VR raises the concern of determining the best task to implement in a specific context. Evaluating vocabulary tasks based on their cognitive requirements and contribution to VG and VR can be a good starting point for selecting and designing the most appropriate tasks for a specific context.

The present study therefore aims to investigate four different tasks in terms of their cognitive requirements and contribution to vocabulary acquisition within the framework of TILH proposed by Laufer and Hulstijn (2001). The results of the study are expected to add to our knowledge about the efficacy of different tasks in terms of

their contribution to incidental vocabulary acquisition. In this sense, the implications can guide teachers and material designers in preparing appropriate tasks for vocabulary development of language learners.

In addition, the present study will shed some light on the predictive validity of the TILH as the tasks under investigation operationalise varying TILLs and different distributions of evaluation, search and need components. How different tasks with different and equal TILLs affect incidental vocabulary acquisition and how the different components in the framework operates will be explored.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction

This present inquiry is aimed at investigating the effects of different tasks on incidental vocabulary acquisition through reading within the framework of Task-induced involvement load by Laufer and Hulstijn (2001). The present chapter will be devoted to explaining the differences between intentional and incidental learning of vocabulary as the two approaches to vocabulary acquisition. Then, the factors contributing to incidental vocabulary acquisition will be covered. After that, reasons leading to the proposal of task-induced involvement load hypothesis will be explained. Moreover, a detailed account of the internal structure and the assumptions of this construct for incidental vocabulary acquisition will be provided. Some important considerations and caveats for studies on the task-induced involvement load will also be discussed in the light of the hypothesis. In addition, some studies investigating the effects of different vocabulary tasks designed in the light of the framework of task-induced involvement load on incidental vocabulary gain and retention will be reviewed. Finally, the significance of the related literature on the present investigation will be explained.

2.2. Vocabulary Knowledge and Teaching of Vocabulary

Nation (2001) provides a detailed framework of what is included in "knowing words". According to this framework, the most basic steps of knowing a word consist of recognising the meanings and the forms of words in a language. From a psychological perspective, according to Baddeley, in order for a word to be learned or acquired, three steps are important: noticing, retrieval and generation (cited in Van Polen, 2014, 7). Noticing means paying attention to specific words when they are used in context for various reasons such as the salience of the word or the potential of that specific word to fill a gap in the learner's mind. Retrieval requires the noticed word to be taken from the schemata and the meaning of this word to be matched with the form or vice versa. Generation includes using the word for conveying meaning.

In terms of providing opportunities for either one of these three steps to occur, it is often stated that L1 context and L2 context are substantially different from each other. Folse (2006) states that while L1 context provides abundant stimuli for native speakers,

L2 learners of that language cannot find many opportunities for exposure to much input. Therefore, while the three steps in learning a word in an L1 context is somehow taken for granted, the situation is not the same for L2 learners to acquire the needed number of words to be competent in all aspects of the language.

Since learning a large number words in a short time is a must for L2 learners to have a good command of the target language, language classes are generally open to any manipulation to bridge the gap between the needed numbers of words to be known (Paribakht & Wesche, 1997). In order to provide this 'manipulation', two types of vocabulary learning, seemingly dichotomous, have been acknowledged through the literature of vocabulary acquisition. The first one is intentional learning and the other is incidental learning.

2.2.1. Intentional/Incidental Leaning of Vocabulary

Intentional vocabulary learning has been referred to as paying deliberate attention to commit the lexical items to long term memory by applying some retention strategies such as rehearsals and mnemonics for preparing a later recall test (Gass, 1999; Hulstijn, 2003; Schmidt, 1984).

On the other hand, the second approach posits that vocabulary items can also be acquired without deliberate attention to commit the lexical items into long-term memory for later use. Generally, incidental vocabulary learning has been used to refer to situation where learners acquire words by focusing their attention to something else rather than specifically trying to learn the words they encounter (Day & Bamford, 1998; Krashen, 1989; Nagy & Herman, 1987).

Regarding the effectiveness of the intentional vocabulary learning, many researchers have advocated that the number of vocabulary items acquired under this condition is highly limited and most of the vocabulary stock learners have cannot be attributed to deliberate word learning strategies (Hulstijn, 1992; Krashen, 1989; Nagy & Herman, 1987). Given that the intentional learning condition leads to limited amount of vocabulary acquisition, then how incidental vocabulary acquisition can be fostered in a way that allows learners to acquire the needed lexical items in a limited time is an

important concern. Paribakht and Wesche (1997) underline that incidental learning has an unpredictable nature as learners may pay attention to some words while totally ignoring some other. Therefore, they posit that learners cannot be expected to learn all the words they encounter while reading thus some interventions are needed to draw the learners' attention to important words so that these words will not be ignored. Such interventions may include glossing of some words, increasing the encounters of specific words, or encouraging dictionary use for the meanings of specific words. All these interventions are proved to be beneficial for students to pay attention to the meanings of specific words in a reading text thus eliminating the chances that students will ignore the selected words.

Several studies have tested the effect of the times a specific word is encountered in a reading on the acquisition of those words and it was found that the more frequently occurring words attracted the students' attention and were acquired. It is also evident that the reoccurrence of unknown words are effective in creating a need in the students to attend to the meanings of these words better (Eckerth & Tavakoli, 2012; Hulstijn, Hollander & Greidanus, 1996; Rott, 1999; Van Polen, 2014).

Another factor contributing to incidental acquisition of words is providing glossary accompanying the text. Studies investigating the benefit of providing glosses for specific words in a text indicated that the glossed words were retained better than those which were inferred only (Cobb, 1997; Hulstijn, 1992; Hulstijn et al., 1996; Watanabe, 1997).

Encouraging the learners to use dictionaries on any encounters of unknown words was also tested whether it contributed to incidental acquisition of vocabulary. The results demonstrated that consulting dictionaries was also useful in drawing the students' attention to specific words and acquire these words (Knight, 1994; Luppescu & Day, 1993; Newton, 1995).

In addition to these factors, the most important contributor of incidental vocabulary acquisition is the level of processing. According to Eysenck (1982), the retention of the newly received information is contingent upon how that information is processed. In this regard, in order to catch the learners' attention and make them process specific words at different levels, reading-based vocabulary tasks are suggested by

Paribakht and Wesche (1997). In the literature, some studies compared the effect of engaging in a vocabulary task in addition to reading comprehension and reading only on incidental vocabulary acquisition. The results strongly suggested that taking part in additional vocabulary tasks addressing the unknown words in a text led to higher vocabulary retention than reading only conditions (Folse, 2006; Hulstijn, Hollander & Greidanus, 1996; Joe, 1998; Min, 2008; Pichette, de Serres & Lafontaine, 2011). However, determining the levels of processing a vocabulary task bears is a serious concern to compare and design different vocabulary tasks. Addressing this issue, the Task-induced Involvement Load Hypothesis was proposed by Laufer and Hulstijn (2001) to determine a task's processing levels and potential contribution to incidental vocabulary acquisition.

2.3. Task-induced Involvement Load Hypothesis

The huge amount of lexical knowledge needed in order to be competent in the target language and the proven value of vocabulary tasks in increasing vocabulary knowledge compel teachers and researchers to decide which vocabulary tasks are more useful. In an attempt to answer which tasks are better in terms of their contribution to VG and VR, Laufer and Hulstijn (2001) proposed the construct of Task-induced Involvement Load to compare vocabulary tasks based on their cognitive requirements. They give account of three reasons for introducing the TILH.

The first reason is that the former constructs of 'generative model' by Wittrock (1974) and 'depth of processing' model by Craik and Lockhart (1972) were deficient in presenting observable processing levels of a task. These two constructs posited that when learners encounter an unfamiliar piece of information, they process this information by forming links between their existing knowledge and this new piece of information. How many associations are made and how deeply the information is processed later determine the retention time of this new piece of information. According to these two theoretical perspectives, being able to store the new information in the long term memory does not depend on the time it is held in the short term memory but on how deep it is processed in the brain. Laufer and Hulstijn (2001) put their counter argument by comparing the following tasks: practicing the spelling of a newly learned

word and trying to find out the collocations of this word. When compared within the former frameworks, the first task is considered to be shallow while the second task is assumed to require a deeper processing. However, there is not a clear description of the level at which the words are processed and what exactly constitutes a "level".

Another reason, according to Laufer and Hulstijn (2001) is that when learning is considered, not only cognitive but also motivational dimension of tasks should be considered. They state that even though there are several studies investigating how motivation correlates with language acquisition, none of the motivation theories demonstrate how different tasks can evoke different degree of motivation in learner. What they argue is that different vocabulary tasks can lead students to have different levels of motivation towards learning the meanings of the target words (TW), words selected by external agents for the learners to master. Therefore, the motivational dimension should also be taken into consideration in classifying, evaluating, and designing vocabulary tasks, according to Laufer and Hulstijn (2001).

The third reason according to Laufer and Hulstijn (2001) is that learners who are given the opportunity to look up the meanings in a dictionary are better at retaining the unknown words when compared to those who do not look up the words in a dictionary. Studies investigating the effects of making students search the meanings of the unknown words and providing the glossary of these words suggested that when students were engaged in dictionary use or some other type of looking for the meanings of the words, they acquired more words (Knight, 1994; Luppescu & Day, 1993; Newton, 1995). Despite the proven value of looking for the meanings of the words, as Laufer and Hulstijn (2001) state, this element of search is not investigated as an ingredient of a vocabulary task, but rather as a separate entity.

To sum up, these reasons led Laufer and Hulstijn (2001) to propose the TILH which determines the cognitive load of vocabulary tasks based on the presence or absence of three components: evaluation, search and need.

2.3.1. Internal Structure of Task-induced Involvement Load

According to the framework of TILH, evaluation means the judgement of the learners about whether a given word fits into a specific context or their choice of the correct homonym of a word to match a sentence or a piece of discourse. Laufer and Hulstijn described the component as 'selective decision based on a criterion of semantic and formal appropriateness (fit) of the word into its context' (p. 15). The component of evaluation can exist in a task at three levels: moderate and strong or absent. Laufer and Hulstijn (2001) demonstrated how the level of evaluation could adjusted with the following tasks: a fill-in the blanks task requiring the learner to select the appropriate word from a list of words and another task requiring the learner to use a word in meaningful sentences. The first task will pose moderate evaluation as the context where the word will be used is limited to just the list of words and the learner just needs to decide which word will suit best to the gap. However, writing an original sentence using the word will require strong evaluation as the learner needs to make some syntactic and semantic decisions such as what kinds of words can come before and after the TW, what kind of a sentence the word should be used in.

Search component is related to whether the learners need to find out the meaning of the given words himself by using any source like dictionaries or software. If the meaning of a specific word is looked up in a dictionary or in any other sources, then there exists search. However, when the definitions of the words are provided in a dictionary or the meanings can be inferred from the context, then, the 'search' component is not present.

Need is the motivational component of the construct of involvement load and it is related to why a specific word has to be known. The component is operationalised at three levels: moderate, strong or it can be absent. If a word is required by an external factor such as the teacher's demand or the requirement of a task, then need will be moderate. On the other hand, if the learner wants to use the word to serve his own purposes such as talking about something or describing something, need will be strong. Whether the drive to use the word is externally or internally imposed determines the degree of need in an activity.

The TILH by Laufer and Hulstijn (2001) posits that existence or absence of these three components determines the overall Task-induced Involvement Load Level (TILL) of a given task. The existence of the components is graded through minuses (-) and plusses (+). If a component does not exist in a task, it is given – and counted as 0. However, when a task includes a component at moderate level that component is given + and counted as 1. When a component exists in a task at strong level, it is given ++ and counted as 2. Table 1 summarizes the components in the TILH and demonstrates how to measure the existence of each component.

Table 1. Components of Task-induced Involvement Load

	Evaluation	Search	Need
Feature	Cognitive dimension of involvement	Cognitive dimension of involvement	Motivational dimension of involvement
Operationalisation	Comparison of a word to other words to determine whether it fits its context	Attempt to find the meaning of a word	Need for knowing words for the task
Existence	Absent (-) Moderate (+) Strong (++)	Absent (-) Present (+)	Absent (-) Moderate (+) Strong (++)

2.3.2. Important Considerations for Involvement Load Studies

Laufer and Hulstijn (2001) put forward some key assumptions regarding the effect of TILL within the framework of TILH : a) retention of incidentally acquired words depends on the absence or presence of need, search and evaluation in a task; b) words which the learners process with higher involvement load will be retained better than those processed at a lower involvement load level and c) teacher/researcher designed tasks with a higher involvement load will be more effective for retaining words than tasks with a lower involvement load.

There are three important points to take into consideration before assessing or conducting a study aiming at testing the effect of involvement load on VG and VR. Firstly, as suggested by Laufer and Hulstijn (2001), TILH posits that retention of new words are contingent upon the presence or absence of three components with which the encountered words are processed under incidental condition only. Therefore, a distinction between incidental and intentional learning should be made at this point. According to Eysenck (1982), incidental learning refers to a situation where learners are required to process some piece of information without being forewarned about upcoming recall and retention tests. Schmidt (1994) and Gass (1999) described incidental learning as a situation where learners acquire a piece of information when they focus on doing something else such as focusing on completing a task following a reading passage. Combining together these definitions, Laufer and Hulstijn (2001) described incidental acquisition of vocabulary as a situation where the learners pick up words as a result of completing a task without knowing that the retention of the words will be tested later. Conversely, intentional learning means trying to commit the newly encountered words to long term memory using some strategies and being prepared for an upcoming retention test (Eysenck, 1982; Gass, 1999; Schmidt, 1994). When deliberate attention and using rehearsal strategies to commit words to memory are at play, it will not be appropriate to attribute vocabulary acquisition solely to vocabulary tasks. Therefore, studies within the framework of TILH address incidental vocabulary acquisition.

Secondly, pointing to the unpredictable nature of incidental acquisition of vocabulary through reading underlined by Paribakht and Wesche (1997), Laufer and Hulstijn (2001) maintained that every unknown word in a reading text may not be processed with the same involvement load (e.g. some words may be skipped while other words are looked up for their meanings). While the neglected words will not be acquired, those words the learners attend to whose meaning and form will be retained. Therefore, as a response to the suggestion by Paribakht and Wesche (1997) that there is need for guiding learners' attention to specific vocabulary, Laufer and Hulstijn (2001) proposed that teacher/researcher designed tasks can provide this guidance by requiring learners to attend to the selected words in a reading text thus preventing learners from skipping those words.

Thirdly, Laufer and Hulstijn (2001) do not take into consideration the relative difficulty and allotted time of a task in promoting incidental vocabulary acquisition. For example, in terms of efficiency in promoting incidental VG and VR, a 50-minute composition writing task is expected to be equal to a 20-minute matching task as long as the total TILLs of these tasks are similar. By adjusting the existence of components in tasks, it is possible to reach different TILLs and therefore, a matching task can be as loaded as a composition writing task.

2.4. Review of Related Studies on Task-induced Involvement Load Hypothesis

The literature of TILH starts with Hulstijn and Laufer (2001), who tried to provide empirical support for their framework of TILH. For research purposes, three different tasks were put under investigation and their contribution to VG and VR was compared. The tasks were reading comprehension with marginal glosses, reading comprehension plus gap fill and composition writing incorporating the TWs. The first task induced moderate need as the knowledge of the TWs was obliged by the comprehension questions. The second task induced moderate need and moderate evaluation. The third task was loaded with moderate need and strong evaluation. Therefore, total involvement load indexes of the tasks were 1, 2 and 3 respectively. Table 2 demonstrates how the existence of components in these tasks were adjusted.

Table 2. Task-induced Involvement Levels of Tasks in Hulstijn & Laufer (2001)

	Reading comprehension with marginal glosses	Gap fill	Composition writing
Evaluation	-	+	++
Search	-	-	-
Need	+	+	+
Total TILL	1	2	3

The study took place in Israel and Netherlands simultaneously. The post-tests required the participants to provide L1 equivalents or English explanations of the TWs. Each correct answer was counted as 1 and the wrong answers were counted as 0. The

groups then were compared on their immediate and delayed post-test scores. Data analysis showed that composition writing group significantly outscored the other groups on both post-tests in both countries. As for gap-filling and reading comprehension with glossary, there was a significant difference between these groups on both immediate and delayed post-tests in Israel. However, the groups in Netherlands did not differ from each other on the post-tests. The study provided strong empirical support for the hypothesis that tasks with higher TILL yield higher incidental VG and VR.

Some points should be noted regarding this first attempt to verify the TILH. Firstly, all the tasks contained moderate need because the knowledge of the TWs was required by the tasks rather than the learners themselves. This is the case when teachers, researchers and material designers prepare tasks which require knowing the TWs. Therefore, increasing the presence of need is not feasible in classroom environments. The other point is that, search was kept under control by providing a glossary and the tasks differed in terms of evaluation component only. By just adding or subtracting the evaluation factor, the construct of involvement load cannot be fully explored. As Laufer and Hulstijn (2001) expressed their doubts that all the components in the construct of involvement load may not be equally effective, there was also need for further studies which investigate different distribution of search and evaluation components together. Therefore, the literature of TILH framework can be divided into a) the studies investigating the evaluation component only, and b) evaluation and search components together.

2.4.1 Effect of Evaluation Component

Following the introduction of the TILH framework by Laufer and Hulstijn (2001), many studies attempted to verify the validity of the construct by using the design of Hulstijn and Laufer (2001). Among these researchers, Tu (2004) prepared different tasks containing varying TILLs and explored the effect of tasks on incidental vocabulary acquisition. The tasks were reading comprehension with marginal glosses (moderate need as the comprehension questions required the participants to refer to the TWs), reading comprehension plus fill-in (inducing moderate need and moderate evaluation) and writing a composition incorporating the TWs (inducing moderate need and strong

evaluation) with involvement TILLs of 1, 2 and 3 respectively. The groups were compared on their immediate and delayed post-test mean scores. The results showed that composition writing group significantly outperformed the other groups on both post-tests. Similarly, the second highest loaded task, reading comprehension plus gap-fill, yielded significantly better results than reading comprehension only task. The results of the study strongly supported involvement load hypothesis.

Another study by Keating (2008) with a similar design to that of Tu (2004) set out to investigate the effect of involvement load on vocabulary gain and retention via three tasks with different TILL imposed. Search component in the tasks was kept absent with the help of a glossary provided and need component was moderate as the drive to use the words was imposed by the tasks themselves. The first task, reading comprehension with marginal glosses inducing neither search nor evaluation. The second task, reading comprehension plus fill-in induced moderate need, no search and moderate evaluation as the participants were to fill in the gaps by choosing among the given words. The third task was sentence composing task in which the participants were to create original sentences incorporating the target words. This task induced moderate need, no search and strong evaluation. Comparison among the groups showed that participants doing task 2 and task 3 significantly outperformed their peers doing task 1, proving that higher involvement load leads to higher VR.

Implementing similar tasks and using a similar distribution of the components, some other researchers as well provided strong support for the TILH. The results suggested that tasks with higher TILLs were found to be better in promoting incidental VG and VR in reading context (Behbahani, Pourdana, Maleki & Javanbakht, 2011; Feng, 2015; Sarbazi, 2014) and in listening context (Jing & Jianbin, 2009).

The framework was also supported empirically in that relative difficulty and allotted time of the tasks are not important for determining the relative usefulness of tasks in leading to incidental vocabulary acquisition. Kim (2008) compared writing sentences and writing a composition incorporating the TWs. The tasks had equal distribution of evaluation component and their total TILLs were the same. Although the composition writing task required more time and was relatively a more challenging process than writing separate sentences, the results suggested that the participants

completing these tasks did not differ from each other significantly on VG and VR scores.

While there are many studies providing support for the huge effect of the evaluation component, there are some other studies which yielded insignificant results demonstrating that higher presence of evaluation does not always bring about higher VG and VR. For instance, Walsh (2009) prepared two tasks with different levels of involvement and compared them in terms of their contribution to incidental VG and retention. The first task was a blank filling task and it induced moderate need and moderate evaluation. The second task was a sentence composing task which induced moderate need and strong evaluation. Both tasks lacked search component as the TWs were glossed and the participants did not have to search for their meanings. When compared on immediate and delayed post-test scores, the tasks did not differ from each other significantly.

Similarly, Soleimani, Rahmanian and Saiedi (2014) compared the vocabulary gains and retention achieved under three tasks: reading comprehension, blank filling and sentence making. The reading comprehension task induced moderate need as the target words had to be known to answer the questions. Therefore, the involvement load index of the tasks were 1, 2 and 3 respectively. Comparison of the groups on immediate and delayed post-tests showed that both sentence making and blank filling tasks yielded significantly higher vocabulary retention than reading comprehension task. However there was not any significant difference between blank filling and sentence making although the latter led to slightly higher scores. Some other studies as well kept search and need components under control and tested whether increasing the presence of evaluation would make any difference. It was found that simply increasing the presence of evaluation component from moderate to strong did not result in significant differences in VG and VR (Bao, 2015; Beal, 2007; Jahangiri & Abilipour, 2014; Keyvanfar & Badraghi, 2011) and that the initial positive effect of strong evaluation diminished over time (Van Polen, 2014).

In summary, increasing the presence of evaluation component was found to be a significant contributor to incidental vocabulary acquisition as shown by several studies.

Requiring the learners to compare a word to other words and make several associations with their existing vocabulary stock was helpful in promoting incidental VG and VR.

However, there were also cases where the increase in the presence of evaluation component did not yield the expected results. These insignificant results were generally attributed to the proficiency levels of the participants. The participants in the more demanding tasks could not benefit from the higher levels of cognitive load due to their lack of vocabulary depth and writing skills (Van Polen, 2014, Walsh, 2009). As a result of deficiencies in vocabulary stock, the participants were not able to form strong connections between the TWs and the context they created, which in turn hindered retention of the newly encountered words (Van Polen, 2014; Walsh, 2009). The participants' unwillingness to complete the vocabulary test was also another factor leading to the inferior scores for the higher loaded groups (Beal, 2007).

Another factor leading to insignificant results was reported to be the disparity between the expected and the actual TILL of tasks (Li, 2014). Li investigated the online participant behaviours within the framework of TILH and tested whether the assumed TILL would be realised. The results showed that the highest involving task did not lead to significantly higher VG and VR and the expected participant behaviour patterns either. Li (2014) attributed these insignificant results to the fact that the tasks were not efficient in promoting the expected TILL as the participants did not attend to the requirements of the tasks.

2.4.2. Effect of Evaluation and Search Components

In addition to the studies merely devoted to explore whether increasing the presence of evaluation component results in significant differences, there are some other studies exploring the effect of both search and evaluation components at the same time. These studies aimed to further verify the TILH by increasing the presence of search component as well as evaluation component. According to the assumptions of the TILH, an increase in the total TILL, regardless of the component being increased (search, evaluation or need), will bring about significant differences. Therefore, designing tasks with different distribution of the components will help further verification of the construct. Several studies aimed at weighing the effect of both

components in the construct of involvement load and inconclusive results were obtained regarding whether both components were equally effective. A limited number of studies yielded results in line with the assumptions of the involvement load hypothesis.

For weighing the effect of the components in involvement load construct, Sarani, Negari and Ghaviniat (2013) designed six tasks with varying involvement loads. Three of the tasks were receptive and the other three were productive. Such a design would help investigating whether different tasks (receptive, productive) with similar involvement load levels would yield similar results in terms of incidental VG and VR. Receptive tasks were true-false (moderate need), matching (moderate need and moderate evaluation) and multiple choice (moderate need, search and moderate evaluation) inducing involvement load indexes of 1, 2, and 3 respectively. The distribution of the components in the receptive set was designed in such a way that the researcher would be able to see if adding a component is really effective (by adding evaluation in the second task and by adding search in the third task). Similarly tasks in the productive set induced similar levels of involvement load and these tasks were short response (moderate need), fill-in (moderate need and moderate evaluation) and sentence writing (moderate need and strong evaluation). Search component was not present in the productive set. The distribution of each component is presented in Table 3.

Table 3. Distribution of Components in Sarani et al. (2013)

	Receptive tasks			Productive tasks		
	True/False	Matching	Multiple Choice	Short Response	Fill-in	Sentence Writing
Evaluation	-	+	+	-	+	++
Search	-	-	+	-	-	-
Need	+	+	+	+	+	+
Total TILL	1	2	3	1	2	3

There is an important point in the tasks presented in Table 3. When true/false and short response, and matching and fill-in pairs are investigated, it can be seen that equal TILLs of these pairs were achieved with need and evaluation components distributed equally. When sentence writing and multiple choice tasks are compared on

the other hand, it can be seen that equal TILLs were achieved by adjusting the presence of search and evaluation components. Search component was present in the multiple choice task while the already existing evaluation was increased and search was absent in sentence writing task. This design would enable the researcher to test whether the two components were equally effective in contributing to incidental VG and VR.

First, the receptive and productive sets were analysed separately. The analysis on the receptive tasks showed that multiple choice group with the highest involvement load significantly outperformed the other groups. Similarly there was a significant difference between matching and true false tasks on both immediate and delayed post-tests. This result provided strong support for the assumption that increasing the TILL, whether by adjusting search or evaluation, will increase vocabulary retention as well. Another finding supporting this assumption was that there was no difference between sentence making and multiple choice groups which contained different distributions of evaluation and search. This result indicated that however the components are distributed, as long as the total TILL is equal there will be no differences in vocabulary retention.

Using a similar design to that Sarani et al. (2013) used; Ghabanchi, Davoudi and Eskandari (2012), Hazrat (2015) and Pourakbari and Biria (2015) also found similar results proving that regardless of which component to increase, increasing the total TILL in a task results in significant differences in terms of VG and VR.

While the three studies above provided support for the assumption that increasing the overall TILL by changing the distribution of either search or evaluation leads to significant differences in vocabulary retention, there are many studies revealing that the components may not be equally effective. One such study was Yaqubi, Rayati and Gorgi (2010) in which three tasks with varying levels of involvement load were compared. This study was aimed at finding out whether all the components in the construct were equally effective on incidental vocabulary acquisition. For research purposes, a reading comprehension plus multiple choice questions task, a reading comprehension plus fill-in task and a reading comprehension plus composition writing task were compared. The first task induced moderate need, search and moderate evaluation. The second task involved moderate need and moderate evaluation while the third task contained moderate need and strong evaluation. The first and the third tasks

induced similar involvement load index of 3 while the index of the second task was 2. The groups' mean scores on immediate and delayed post-tests demonstrated that contrary to the expectations of the involvement load hypothesis, task 2 significantly outperformed the first task despite being lower in terms of involvement load, which means that adding search did not bring about the expected outcomes. As for the comparison between the two equally loaded tasks, it was found that composition writing group significantly outperformed reading comprehension plus multiple choice questions group. This suggests that, the components of search and evaluation may not be equally weighed as the task that induced higher level of evaluation outperformed the task that induced additional search.

Another study directly aiming at comparing equally loaded tasks was Konno, Takanami, Okuyama and Hirai (2009) in which four different tasks with varying involvement load levels were designed. Two of the tasks, comprehension plus gap filling and sentence writing using the target words induced the same level of involvement load (moderate need, search and strong evaluation) with a total of 4 while the other comprehension plus gap fill had an involvement load index of 2 (moderate need and moderate evaluation) and the other sentence writing with a total index of 3 (moderate need, strong evaluation). Comparison among the groups showed that, contrary to the expectations of involvement load hypothesis, the highest loaded tasks did not result in higher vocabulary retention. Sentence writing group, which induced an involvement load index of 3 significantly outperformed both of the reading comprehension plus fill-in groups, although one of which induced an involvement load index of 4. As for the other sentence writing task, which had an involvement load level of 4, it did not produce significantly better results than the other groups. These findings demonstrated that increasing the involvement load by inserting search factor did not yield better results whereas an increase on the factor of evaluation brought about significant changes.

There were also other studies discovering that inserting search component was not as effective as increasing the presence of evaluation (Haratmeh, 2012; Jahangard, 2014; Li, 2014; Marmol & Sanchez-Lafuente, 2013; Wang, Xu & Zuo, 2014; Xu, 2010).

The results of the aforementioned studies were not as assumed by the TILH, and these insignificant findings were attributed to factors related to participants' task performance. The common explanation for the insignificant contribution of search factor is that while some participants may have looked into the words in a more detailed way, others may have just looked up the meaning as in the case of using monolingual and bilingual dictionaries (Jahangard, 2014). Superficial use of dictionaries may also be the reason behind the insignificant effect of search. Individuals may have just preferred to look up the meanings of the words without paying attention to the context provided by dictionaries (Konno et al., 2009). Possible lack of dictionary use knowledge was another factor diminishing the positive effects of search (Marmol & Sanchez-Lafuente, 2013). Li (2014) based on the look-up behaviours of the participants highlighted that since the participants who were theoretically required to search the meanings of the TWs did not actually get involved in this process, the effect of search component was not evident.

2.4.3. Path for Further Verification of Task-induced Involvement Load Hypothesis

The construct of TILH was proven to be a successful predictor of more effective tasks in boosting incidental vocabulary acquisition. Studies on the effect of increasing the evaluation component and adding search in a task demonstrated that equipping any task with a higher TILL led to significant difference in terms of VG and VR. However, some contradictory findings were also obtained which suggested that increasing the overall TILL did not always bring about significantly higher vocabulary acquisition rates. The common point in all these studies was that some learner-related factors such as lack of vocabulary depth and writing skills, and deficiencies in exploiting an external source for searching the meaning of word prevented the components of TILH from taking effect. Therefore, in order to investigate the TILH appropriately, relatively higher proficiency participants should be chosen.

CHAPTER 3: METHODOLOGY

3.1. Introduction

This study aimed to explore the effect of task-induced involvement load on incidental vocabulary acquisition through different vocabulary tasks prepared in the light of involvement load framework. Therefore, different task-induced involvement load levels were achieved with different tasks assigned to different groups. The present chapter will provide detailed information about the steps of the experiment. The chapter starts with an account of the subjects chosen for the study and the context where the experiment was conducted. There will also be information about the instruments including the reading texts, vocabulary tasks and vocabulary tests. Finally, how the implementation was undertaken and the relevant data were collected and analysed will be explained in details. Additionally, several criteria for selecting the instruments will be covered.

3.2. Participants

The subjects of the present study were first-year students enrolled in the English Language Teaching program at Anadolu University. The study employed two reading texts implemented at different times and therefore, the number of the participants attending each implementation was different. A total of 131 participants were included in the data analysis of the first implementation while 139 participants were present in the second implementation. The age range of the participants was 17-21. The participants were from eight intact classes enrolled in the Academic Reading course.

The proficiency levels of the subjects participating in the study were upper intermediate and above as either they had just passed the preparation school examination of the institution, or they were already exempt from preparatory education or they had completed the preparatory education. When the participants were compared in terms their proficiency scores on the preparation school examination conducted at the beginning of the academic calendar, the classes were found to be similar to each other. Therefore, the comparison allowed the researcher to assume that the groups were homogeneous and proficiency differences would not interfere with the overall results of the study.

3.3. The Research Context

The study was conducted in the fall term of 2015-2016 academic year. In the term the study was undertaken, the participants in the sample were taking basic language skills courses such as Contextual Grammar, Spoken Communication, Written Communication, Listening Comprehension and Academic Reading.

The implementation of the vocabulary tasks under investigation took place in the 'Academic Reading' course as the course addressed academic reading texts. The present study, which also made use of academic reading texts, would lead to minimum interference with the course in terms of the content. Moreover, it would be convenient to keep track of the attendance of the participants as they were registered for the same course.

The classes taking this course met once a week and the course took three class hours. The first two weeks of the course were devoted to presenting previewing, skimming and close reading strategies. In the weeks following the presentation, as homework, the students were expected to analyse one article assigned for the following week using the strategies they had been presented. In the lesson, the passage of the week was discussed as whole class first and then analysed with the guidance of the instructors. As for the examination, the students were expected to demonstrate how to apply reading strategies (previewing, skimming and close reading) to a reading text. There was also another part asking the students to provide definitions of some words and use these words in sentences.

3.4. Instruments

For research purposes, there were several instruments utilised in the present study. The participants were required to read two different texts accompanied by reading comprehension sections and complete vocabulary tasks and they were measured on their VG and VR scores through a set of vocabulary tests.

3.4.1. Reading Texts

Similar to the previous studies on the construct of TILH, this study also made use of reading texts accompanied by reading comprehension sections along with different vocabulary tasks with varying TILLs.

Two different reading texts were used in the present investigation. Implementing two different texts would enable the researcher to explore whether the results would be consistent over different texts and different TWs. The texts were taken from an online source called IELTS Mentor, access to which is granted for anyone interested in preparing for International English Language Testing System (IELTS) proficiency exam. The texts selected were 'The Triune Brain' for the first implementation and 'Practical Intelligence Lends a Hand' for the second implementation.

3.4.1.1 Criteria for Selecting the Texts and Target Words

While choosing the reading texts, four points were considered. The first point was that of reading difficulty, and both texts were to be similar in terms of reading ease. The texts were measured in terms of readability through the Flesch-Kincaid Grade Level readability formula proposed by Kincaid (1975), and the results showed that both texts were suitable for college or above level readers. The two texts being comparable in terms of reading difficulty would eliminate text effects on reading comprehension and thus, on success in vocabulary tasks.

The second concern was whether the reading texts contained enough number of lexical items as most of the studies on the construct of TILH chose 8-15 words as target vocabulary items. These words should be belonging to the same category, and potentially unknown to the participants. As Laufer (1990) suggested that certain grammatical categories lead to more difficulties in acquisition than others. Nouns, verbs, adjectives and adverbs are not acquired in similar patterns, with the adverbs being the most difficult one followed by verbs and adjectives. However, nouns are the easiest category to acquire, according to Laufer (1990). In order to eliminate the effect of different categories, only nouns were chosen for this study. In determining the potential TWs, JACET 8000 Level Marker, an online tool for determining the level of a word, by Shimizu (2008) was used. The words that were in the 5000 band or above

level were identified first, so that the words would be not much frequent. Then, as any encounters with the TWs after the implementation would interfere with the effect of the tasks, Corpus of Contemporary American English (COCA) created by Davies (2008) was utilized to determine the words which were frequently found in Academic and written texts but scarce in spoken discourse. Identifying the words scarcely used in spoken discourse would minimize the possibility that the participants could encounter these words while watching TV series or movies. As for the possibility of the participants encountering the words through written input, the Academic Reading Course books were investigated whether the selected words appeared. The TWs in the 'The Triune Brain' were *aggression*, *equilibrium*, *indifference*, *kinship*, *morality*, *offspring*, *propensity*, *respiration*, *solidarity* and *wit*. As for the TWs in 'Practical Intelligence Lends a Hand' there were *acumen*, *apprentice*, *audit*, *deficiency*, *dilemma*, *distress*, *entrepreneur*, *frustration*, *gratification,* and *recruitment* on the list. Table 4 presents the frequencies of the TWs across different genres in COCA.

Table 4. Frequency and Percentage of Encounters of Target Words in COCA

	Spoken	Fiction	Magazine	Newspaper	Academic	Total
acumen	6%	12%	28%	29%	25%	526
aggression	15%	4%	14%	10%	57%	6505
apprentice	14%	33%	20%	21%	12%	1782
audit	7%	3%	14%	30%	46%	4282
deficiency	5%	4%	30%	12%	49%	3702
dilemma	14%	5%	19%	14%	48%	7413
distress	8%	16%	16%	9%	51%	5748
entrepreneur	11%	2%	40%	27%	20%	7045
equilibrium	3%	10%	15%	7%	65%	2034
frustration	18%	19%	18%	22%	23%	8612
gratification	10%	10%	28%	17%	35%	1114
indifference	6%	27%	19%	14%	34%	2430
kinship	2%	8%	9%	8%	73%	2312
morality	13%	6%	17%	11%	53%	5342
offspring	4%	13%	35%	13%	35%	3604
propensity	6%	5%	17%	9%	63%	1522
recruitment	6%	2%	9%	13%	70%	3993
respiration	7%	17%	19%	3%	54%	630
solidarity	11%	5%	18%	15%	51%	3957
wit	8%	28%	27%	23%	14%	3269

The possible unfamiliarity and usefulness of the selected target vocabulary items were then questioned by consulting the instructors of the course and some experts in the department of ELT. As a result, the words were found to be appropriate to use.

The third point was whether the reading passages were followed by reading comprehension questions, which was convenient as it eliminated the need for creating comprehension questions. Using texts with comprehension questions covering the essential information would enable the researcher to ensure that the participants' attention would be drawn to complete the questions rather than commit the TWs into

long term memory, which is an essential component of incidental vocabulary acquisition research (Hulstijn, 2003). Therefore, the texts were taken from IELTS Academic Reading sample tests. Both of the reading texts were accompanied by a reading comprehension section consisting of two parts. These comprehension questions were checked on their appropriateness and strength to measure comprehension by asking the course instructors and some experts in the ELT department.

The last consideration was the internal organisation of the texts. When compared in terms of presenting the information, the two reading texts were similar since the pattern of organisation used was compare and contrast. Both of the texts aimed to give an account of three different entities and compared them to each other. This pattern would make it feasible to retell the text, which was one of the tasks used in the study.

The first passage, 'The Triune Brain' is an article on the differentiated structure of the brain in animals (reptiles and mammals) and humans consisting of 821 words and 7 paragraphs. A comparison between humans, reptiles and mammals is made in terms of how the brain of each is divided into cortexes. The functions of reptilian cortex, limbic cortex and neocortex are illustrated through examples. In addition to the different functions of the brain in animals and humans, there is also information on what problems occur in case of any damage to the different parts of the brain and some incidents are provided to illustrate these problems. The passage is followed by a reading comprehension section consisting of two parts. The first part requires the readers to match a total of 9 features with the three different cortexes. The second part consists of 4 blank filling questions requiring the readers to provide answers up to two words maximum. (See Appendix I for "The Triune Brain" and comprehension questions).

The second text 'Practical Intelligence Lends a Hand' is an article on three different types of intelligence. The passage contains 912 words and 13 paragraphs. The text is aimed at informing the reader about how academic intelligence, emotional intelligence and practical intelligence differ from each other; and how performance on these three types of intelligence is measured. Some examples of how these three different intelligence types function are provided for the reader. Similar to the 'The Triune Brain', this text is also accompanied by a reading comprehension section consisting of two parts one of which is matching a total of 7 features with three types of

intelligence and the other part consisting of 5 multiple choice questions. (See Appendix II for “Practical Intelligence Lends a Hand” and comprehension questions).

3.4.2. Vocabulary Tasks

For the aim of the present study, the participants were to read the texts, complete the comprehension tasks and a vocabulary task they were randomly assigned to. The vocabulary tasks designed to operationalise varying levels of involvement load were fill-in the blanks by searching (FBS), fill in the blanks with glossary (FWG), retelling by searching (RBS) and retelling with glossary (RWG). The overview of the tasks and their involvement load indexes are shown in Table 5.

Table 5. The tasks and Total Task-induced Involvement Load Levels

	By searching		With glossary	
	Fill-in	Retelling	Fill-in	Retelling
Evaluation	+	++	+	++
Search	+	+	-	-
Need	+	+	+	+
Total TILL	3	4	2	3

As seen in Table 5, the participants were divided into two broad groups as by searching and with glossary, which would enable the researcher to see whether adding or removing search component had any effect on incidental vocabulary acquisition. As for evaluation, there were two tasks inducing different levels of it so that whether increasing or decreasing the level of evaluation had any effect on vocabulary retention would be verified. As demonstrated in Table 5, the highest involving task was retelling by searching followed by fill-in by searching and retelling with glossary. The least involving tasks was assumed to be fill-in with glossary. This design would help the researcher test whether tasks with similar task-induced involvement load levels (TILLs) FBS and RWG, operationalized with a different distribution of the search and evaluation components would yield similar results as well as whether different tasks with varying TILLs would result in any significant difference in vocabulary retention.

3.4.2.1. Fill-in by Searching

The participants assigned to fill in the blanks by searching group were required to complete the blanks of a gapped version of the reading passage by choosing the appropriate words from a word list, including the TWs and an additional word functioning as a distractor. The participants were encouraged to consult a dictionary or their smart phones for looking up the meanings of the words in the word list. This task, according to the framework of TILH, induced moderate need as the participants were required to know the meanings of the words by the task; search since the participants were allowed to look up the meanings of the words; and moderate evaluation as the participants compared a limited number of words to each other. The total TILL of this task therefore, was 3, and the participants were given 40 minutes to complete this task (See Appendix III and IV for the task layouts of two texts).

3.4.2.2. Fill-in with Glossary

Fill in the blanks with glossary group was also required to complete the reading passage by filling in the gaps by using the TWs. This time, unlike the fill-in by searching group, both Turkish equivalents and English explanations of the TWs were provided by the researcher by means of a glossary. Therefore, the participants did not have to consult any other source for the meanings of the words and the search component did not exist. In terms of involvement load, the task induced moderate need, moderate evaluation and no search. The total TILL of this task was calculated as 2. The participants were given 30 minutes to complete this task (See Appendix V and VI for the task layouts of two texts).

3.4.2.3. Retelling by Searching

Retelling by searching group was required to complete the reading comprehension questions and retell the text by incorporating the TWs encountered in the reading text. The TWs were written in bold to draw the participants' attention to the words and the context they were used in. Additionally, they were allowed to look up the meanings of the TWs in a dictionary or any other source. The instruction for this task informed the students about what information they had to include in their own versions of the texts.

This task contained moderate need as using the TWs was required by the task rather than the participants themselves. There was search component as the TWs were to be looked up in a dictionary or any other source depending on the participants' preferences. Finally, the component of evaluation was strong as the participants needed to create a context based on their existing linguistic knowledge. The overall TILL of the task was calculated as. The allotted time for this task was 60 minutes (See Appendix VII and VIII for the task layouts of the two texts).

3.4.2.4. Retelling with Glossary

Quite similar to the retelling by searching group, retelling with glossary group was also required to complete the reading comprehension questions and retell the reading text by incorporating the TWs according to the instructions. However, the participants in this group did not have to consult any sources for the meanings of the words as they were provided with marginal glosses of the words. Similar to the retelling by searching task, the TWs were written in bold to draw the participants' attention to the target words. This task induced moderate need; no search as the TWs were already presented through a glossary; and strong evaluation. The overall involvement load index of this task was 3. The time required to complete this task was 50 minutes (See Appendix IX and X for the task layouts of the two texts).

The participants were assigned to one of the vocabulary tasks randomly. Table 6 demonstrates the total number of participants across two implementations.

Table 6. Number of the Participants across Tasks and Implementation

	The First Implementation	The Second Implementation
Fill-in with Glossary	30	33
Fill-in by Searching	39	39
Retelling with Glossary	30	33
Retelling by Searching	32	34
Total	131	139

3.4.3. Vocabulary Tests

In order to test whether the participants knew the meanings of the TWs prior to the implementation and to measure the VG and VR, three versions of the same vocabulary test were used. Similar to the vocabulary tests used in Hulstijn and Laufer (2001), the vocabulary tests used in the present study required the participants to provide the Turkish equivalents or English explanations of the TWs presented as a simple word list. The participants were instructed to put 'X' next to the words they did not know. Similar to Hulstijn and Laufer (2001), correct answers were calculated as 1 and the incorrect answers were counted as 0. If the answer was partly correct, half a point was given to the participant. The layouts of the vocabulary tests were not similar. The appearance order of the TWs on the list was changed to eliminate the possibility that the participants would associate the order of the words with the meanings and use this to their advantage (See Appendix XI and XII for vocabulary tests for the texts).

The vocabulary pre-test was aimed at ensuring that the participants were unfamiliar with the TWs. This test would help researcher find out the participants who knew any of the TWs and score this participant accordingly on the immediate and delayed post-tests.

The immediate post-test was administered to measure the participants' immediate incidental VG. The TWs appeared in a different order from that of the pre-test to prevent the participants using this to their advantage. The correct answers were calculated 1, the incorrect answers as 0 and the partially correct ones were given a value of 0,5.

The delayed post-test was designed in order to find out the effect of different tasks with varying TILLs on VR over time. The scoring was similar to that of the pre- and post-tests.

3.5. Procedures

The present investigation consisted of three stages: consenting, piloting and the actual implementation. The present section aims to provide what these stages included.

3.5.1. Consenting

Before starting the implementations, the informed consent of the institution was obtained through the ethics committee of the university. Also the instructors of the Academic Reading Course were informed about the study and their consent was obtained as well (See Appendix XIII for the consent form for the instructors). As for the participants, the procedural purposes of the study were not told in advance to prevent any of the participants from devoting special attention to commit the target words into their long term memory.

3.5.2. Piloting

After obtaining the consents, the instruments were piloted with another group of first year ELT students who were not included in the actual study. The purpose of the pilot was to find out whether the instructions of the reading comprehension sections and the vocabulary tasks were clear. Another aim was to determine the time required by each vocabulary task. Finally, the TWs were checked whether they were unknown to the participants. Based on the findings of the pilot study, the instructions were found to be clear to understand. As for the time to be allotted for the tasks, FWG, FBS, RWG and RBS tasks required 30, 40, 50 and 60 minutes to complete, respectively. Finally, the TWs were found to be unknown to most of the participants with only one participant knowing the meanings of two words among the 20 TWs.

3.5.3. Implementation

The actual study adopted a between-subjects quasi-experimental design (Creswell, 1994) without a control group: four different tasks with varying involvement load levels were assigned to four groups from eight intact classes. The implementation process lasted for four weeks in total. The classes for the Academic Reading Course were held four days a week and on each day there were two classes meeting for the course. Each two class taking the course on the same day were assigned to one task randomly. Table 7 presents the implementation schedule.

Table 7. Implementation Schedule

	Day 1	Day 2	Day 3	Day 4
	Class C-F	Class B-D	Class A-E	Class G-H
Week 1 Text 1 Vocabulary Tasks and Immediate Post-test	FWG	FBS	RWG	FBS
Week 2 Text 2 Vocabulary Tasks and Immediate Post-test	FWG	FBS	RWG	FBS
Week 3 Text 1 Delayed Post-test	FWG	FBS	RWG	FBS
Week 4 Text 2 Delayed Post-test	FWG	FBS	RWG	FBS

Prior to the implementation of the reading texts, the vocabulary pre-test of the related reading text was administered to check whether the participants knew the meanings of the TWs. The participants were required to provide Turkish equivalents or English explanations of the TWs appearing in an alphabetical order.

After the completion of the pre-tests, the participants were handed out different versions of the texts and vocabulary tasks according to their respective groups. As explained above, different groups completed a different vocabulary task in the allotted time. Once the reading comprehension sections and vocabulary tasks were completed, the worksheets were collected and an unannounced immediate post-test was administered to measure the immediate VG.

As most studies like Hulstijn and Laufer (2001), Keating (2008), Kim (2008) regarding the effect of different vocabulary tasks adopted, a two-week period was set for administering the unannounced delayed post-test to measure the effect of TILL on vocabulary retention. The implementation procedure was the same for both of the reading texts.

3.6. Analysis of the Data

The participants were tested on their knowledge of the TWs through three vocabulary tests. In the vocabulary pre-test the subjects who knew any of the TWs were scored 0 for those words on the post-tests to make sure that the overall score would reflect the actual improvement in vocabulary knowledge.

The subjects were compared in terms of their immediate VG through their scores on the immediate post-test. Each correct Turkish equivalent or English explanation was scored 1 and incorrect ones were scored 0. Partially correct answers (e.g., providing a meaning belonging to a different word category) was given half a point. The maximum immediate VG score was, therefore, 10.

Similarly, the participants were compared on their vocabulary retention via the scores obtained on the delayed post-test. Similar scoring procedure was applied to the delayed post-test.

The groups processing the target words at different TILLs were compared on their immediate and delayed post-test scores to answer the first research question regarding the effect of different TILLs on incidental VG and VR. For the second research question, on the other hand, the groups completing tasks with similar TILLs were compared on both post-tests. Since there were four groups, One-way ANOVA was selected as the statistical analysis method.

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Introduction

The present chapter aims to present the findings of the implementation process and provide a detailed discussion of the results in the light of the previous research on the effect of task-induced involvement load level (TILL) on incidental vocabulary gain (VG) and vocabulary retention (VR). Furthermore, the research questions set in the beginning of the study will be answered, and the answers to the questions will be discussed based on the assumptions of the construct of task-induced involvement load hypothesis (TILH) and the findings of the previous studies.

4.2. Results

The purpose of the present inquiry was to explore the effects of TILL on immediate vocabulary gain and vocabulary retention. Therefore, the answers to the following two research questions were sought:

1) Do different vocabulary tasks with varying levels of involvement load differ in terms of their contribution to vocabulary gain and retention?

2) Do different vocabulary tasks with similar levels of involvement load operationalised by different components (search and evaluation) lead to similar amounts of vocabulary gain and retention?

The present investigation aimed at finding the effect of TILL on incidental VG and VR utilising four different tasks with varying TILLs. Two separate reading texts were prepared to operationalise different TILLs. To this end, the results obtained from the immediate and delayed vocabulary tests will be analysed for both of the two reading texts in the order they were implemented. Firstly the findings of the two implementations regarding the first research question will be presented. After then, the answer of the findings to the second research question will be covered.

4.2.1. Tasks with Different Involvement Load Levels

The present section will be devoted to the present the findings regarding the effect of increasing the level of involvement in a task. To this end, the groups completing different tasks with different task-induced involvement load levels (TILLs) will be compared on the immediate and delayed post-test scores for both implementations.

4.2.1.1. The First Implementation

In order to find out whether the different groups processing the target words at different involvement load levels differed in terms of their immediate VG, they were compared on their immediate post-test scores. The mean scores of the groups on the immediate vocabulary post-test for the first implementation are presented in Table 8.

Table 8. Immediate Vocabulary Gain Scores for the First Implementation

	N	M	SD	Min.	Max.
FWG	30	8,4	,9685	7,0	10,0
FBS	39	8,513	1,5748	5,0	10,0
RWG	30	8,6	1,4937	6,0	10,0
RBS	32	9,172	,9299	7,0	10,0

Firstly, it can be seen in Table 8 that all the groups, regardless of their different TILLs, gained the meanings of nearly all of the ten target words. The tasks actually led to substantial amounts of incidental vocabulary gain. When the table is investigated closely, it is evident that the lowest mean score was obtained in the fill-in with glossary (FWG) group with a TILL of 2 (M=8.4, SD=0.968), followed by fill-in by searching (FBS) inducing a TILL of 3 (M=8.513, SD=1.574) and retelling with glossary (RWG) containing a TILL of 3 (M=8.6, SD=1.493). The highest mean score was that of retelling by searching (RBS) (M=9.172, SD=0.929) and total TILL for this task was 4. It was found that slight differences existed among the groups on immediate recall of the words, with the higher loaded groups obtaining higher scores. In order to detect if the difference between the mean scores of the groups on the immediate post-test was

significant, one-way ANOVA was conducted. Table 9 presents the results of one-way ANOVA.

Table 9. One-way ANOVA for Immediate Post-test Scores of the First Implementation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11,357	3	3,786	2,258	,085
Within Groups	212,948	127	1,677		
Total	224,305	130			

Based on the findings of the statistical analysis conducted on the immediate post-test scores of the participants, the groups were found to be not significantly differing from each other ($F(3,127)=2.258$, $p>.05$). Although RBS group got slightly higher scores than RWG group which in turn obtained better results than FBS and FWG respectively, these differences were found to be insignificant.

The results of the immediate post-test showed that an increase in the overall TILL was accompanied by an increase in the VG as well. The groups completing a higher loaded task obtained higher scores. As mentioned in the vocabulary tasks section, the total TILL of the tasks were increased by distributing the components of search and evaluation into the tasks differently. Both inserting search into a task and increasing the level of evaluation component from moderate to strong presence brought about changes in terms of immediate VG scores. However the difference between the groups did not reach significance and there was nearly no difference between the groups in terms of immediate recall.

The groups were compared on their vocabulary retention scores to test the long term effect of TILL as well. In order to investigate the differences between the groups in terms of their vocabulary retention, a similar statistical procedure was applied to the delayed post-test data. Table 10 shows the mean scores four groups obtained on the delayed vocabulary test.

Table 10. Vocabulary Retention Scores for the First Implementation

	N	M	SD	Min.	Max.
FWG	30	5,233	1,6281	2,5	9,0
FBS	39	5,385	1,4117	2,5	8,0
RWG	30	5,633	2,0466	2,5	9,0
RBS	32	6,406	1,3407	4,0	9,0

When Table 10 is investigated, it can be observed that the initial vocabulary gains of the participants decreased to a large extent. However, the higher loaded tasks kept their advantage despite the two-week interval over the lower loaded tasks. There was a similar pattern in the line-up of the scores with the lowest mean score belonging to FWG (*M*=5.233, *SD*=1.628) group with an index of 2. The highest scoring group was RBS (*M*=6.406, *SD*=1.34) with and index of 4 followed by RWG (*M*=5.633, *SD*=2.046) with the involvement level of 3 and FBS (*M*=5.385, *SD*=1.411) with an involvement index of 3 respectively. Once again, the groups were lined up parallel to their TILLs as groups completing higher loaded tasks obtained higher scores. In order to find out if the groups differed from each other significantly, one –way ANOVA was run on the data. Table 11 presents the findings of the analysis.

Table 11. One-way ANOVA for Delayed Post-test Scores of the First Implementation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26,259	3	8,753	3,371	,021
Within Groups	329,783	127	2,597		
Total	356,042	130			

The results of the one-way ANOVA indicated that there was a significant difference among the groups in terms of their vocabulary retention scores ($F(3,127)=3.371$, $p<.05$). In order to detect which groups differed from one another significantly, a Tamhane's T2 post-hoc test was conducted. As the results of the analysis demonstrated, RBS group (*M*=6.406, *SD*=1.34) was found to be significantly outperforming FBS (*M*=5.385, *SD*=1.411) and FWG (*M*=5.233, *SD*=1.628) groups with

a moderate task effect size. However there was not a significant difference between RBS and RWG groups in terms of vocabulary retention.

The findings of the statistical analysis run on the delayed post-test revealed that the groups benefited from higher TILLs as retention scores were higher in the higher loaded tasks. Especially the FBS group with the highest TILL obtained significantly better scores than FBS and FWG groups. Based on the TILH framework, it can be stated that inserting search and increasing the evaluation component to strong presence at the same time leads to significantly higher vocabulary retention scores.

Another finding was that, although inserting search into tasks with similar levels of evaluation component lead to differences in terms of vocabulary retention, this effect was not significant. This was the case in FBS and RBS tasks which contained search but did not lead to significantly higher retention scores than FWG and RWG tasks (lacking search) respectively.

4.2.1.2. The Second Implementation

Similar statistical procedure utilized in the first implementation was applied to the data gathered on the second reading text. Firstly, the groups were compared in terms of their immediate VG on the immediate post-test and their vocabulary retention was compared through their delayed post-test scores. Table 12 presents the mean scores of the groups on the immediate post-test.

Table 12. Immediate Vocabulary Gain Scores for the Second Implementation

	N	M	SD	Min.	Max.
FWG	33	8,758	1,3058	6,0	10,0
FBS	39	8,846	1,1071	6,0	10,0
RWG	33	9,076	1,1465	6,0	10,0
RBS	34	9,294	,8801	7,0	10,0

As it can be seen in Table 12, the groups gained a large amount of the target words, with nearly all the words recalled in the short term regardless of the involvement

load levels. In addition, a similar line-up as in the first reading text was obtained. The groups were in the order parallel to their respective TILLs and it was found that the highest mean score was achieved by the RBS group (*M*=9.294, *SD*=0.88) completing the highest loaded task followed by RWG (*M*=9.076, *SD*=1.146) and FBS (*M*=8.846, *SD*=1.107) groups respectively. The lowest immediate mean score was obtained by the participants in FWG group (*M*=8.758, *SD*=1.305) completing the least involving task. These results signal that increasing the total TILL made a difference. In order to determine whether this difference between the groups was significant, one-way ANOVA was conducted on the mean scores. Table 13 presents the findings of the analysis.

Table 13. One-way ANOVA for Immediate Post-test Scores of the Second Implementation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5,977	3	1,992	1,594	,194
Within Groups	168,757	135	1,250		
Total	174,734	138			

As displayed in Table 13, although the groups differed from each other with the RBS group outperforming the RWG group which in turn obtained a higher mean score than FBS and FWG respectively, these differences were not significant ($F(3,135)=1.594$, $p>.05$).

Regarding the immediate VG scores, similar results as in the first reading text were obtained. As the total TILL increased in a task either with an insertion of search or an increase in the presence of evaluation component, immediate recall scores also increased. However, the difference in the scores did not reach significance. Similar to the finding of the first implementation, there was almost no difference between the groups in terms of immediate recall of the target words.

In order to find out if tasks with different TILLs had an effect on vocabulary retention over a two-week period, the groups were compared on their delayed post-test scores. Table 14 presents the mean scores of the groups on the delayed post-test.

Table 14. Vocabulary Retention Scores for the Second Implementation

	N	M	SD	Min.	Max.
FWG	33	4,742	2,2539	1,0	10,0
FBS	39	5,346	1,4830	2,5	8,5
RWG	33	5,576	1,1932	3,5	9,0
RBS	34	6,309	1,2310	3,0	8,0

As Table 14 displays, the initial gains of the groups were found to have decreased substantially during a two-week period, especially the lowest loaded groups were found to be suffering from the effect of time. As in the first text, the line-up of the mean scores were parallel to the total TILL at which the groups processed the target words. The groups completing the more loaded tasks were able to retain the initial vocabulary gains better than those completing a task with a lower TILL. The highest score was obtained by the RBS group (M=6.309, SD=1.231) which was followed by RWG (M=5.576, SD=1.193) and FBS (M=5.346, SD=1.483) groups. The lowest scoring group was, similar to the immediate post-test results, FWG (M=4.742, SD=2.253). One-way ANOVA was conducted to find out if these scores differed from each other significantly, the results of which are presented in Table 15.

Table 15. One-way ANOVA for Delayed Post-test Scores of the Second Implementation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	42,287	3	14,096	5,569	,001
Within Groups	341,705	135	2,531		
Total	383,993	138			

The results of one-way ANOVA showed that the participants completing different vocabulary tasks with different TILLs differed from each other significantly (F(3,135)=5.569, p<.05) in terms of their vocabulary retention scores. In order to detect where the significant difference was, Tamhane's T2 post-hoc test was exploited and the results of the statistical analysis demonstrated that the participants who completed RBS

task (*M*=6.309, *SD*=1.231) significantly outperformed their peers in the FBS (*M*=5.346, *SD*=1.483) and FWG groups (*M*=4.742, *SD*=2.253) with a moderate task effect.

Similar to the results on the delayed post-test of the first reading text, the results obtained in the second text demonstrated that vocabulary retention was affected by the increase in the total TILL positively. The groups that processed the TWs with a higher TILL retained more words. The RBS group especially, retained significantly more words than FBS and FWG groups respectively, providing support acknowledging the value of TILL in determining the retention of lexical items.

Another finding, which is also in line with the delayed post-test results of the first reading text, is that although inserting search component into tasks with equal level of evaluation brought about increases in vocabulary retention, these increases were not statistically significant. The FBS tasks which contained search did not result in significantly higher vocabulary retention than FWG task which lacked search. Similarly RBS task which induced search did not significantly outperform RWG task significantly.

4.2.2. Tasks with Equal Involvement Load Levels

The second purpose of the present inquiry was to uncover whether different tasks with equal involvement load level operationalised by a different distribution of search and evaluation components would yield similar vocabulary gain and retention. In order to fulfil this purpose, the fill-in by searching (FBS) and retelling with glossary (RWG) groups were compared on their immediate and delayed post-test scores for both implementations.

4.2.2.1. The First Implementation

Besides tasks with different TILLs, the present study also investigated tasks with equal involvement load levels. To this end, the two tasks, FBS and RWG were compared on their effect on immediate vocabulary recall and retention of words over time. As seen in the Table 8 above, the comparison of the groups on the immediate post-test

demonstrated that FBS (*M*=8.513, *SD*=1.574) and RWG (*M*=8.6, *SD*=1.493) groups with equal TILLs but differentiated distribution of search and evaluation components obtained nearly similar scores. Although RWG group obtained higher VG scores, these differences were not significant, which provides support for the assumption of TILH that tasks with equal involvement loads lead to similar amounts of VG.

These two equally loaded groups were also compared on the delayed post-test so as to measure the effect of TILL on long-term vocabulary retention. When Table 10 above is investigated, it can be observed that RWG (*M*=5.633, *SD*=2.046) and FBS (*M*=5.385, *SD*=1.411) groups did not differ significantly from each other since they retained nearly the same number of words as in the case of the immediate post-test results. In this sense, the results suggested that regardless of how the evaluation and search components are distributed in a task, the VG and VR will be similar when the total TILL is kept equal.

4.2.2.2. The Second Implementation

Similar statistical analyses to those conducted on the first reading text were exploited on the scores of the FBS and RWG groups who processed the TWs at the same involvement load level but with a different distribution of search and evaluation components. When the two groups were compared on their immediate post-test scores, as Table 12 above presents, the FBS (*M*=8.846, *SD*=1.107) and RWG (*M*=9.076, *SD*=1.146) tasks were found to be conducive to similar amounts of immediate vocabulary gains. Although the retelling group scored slightly higher, this difference was not significant.

Similarly, in order to find out the long term effect of equal TILLs on vocabulary scores, the participants completing the two equally loaded tasks were compared on their delayed post-test scores. As presented in Table 14, the comparison between the FBS (*M*=5.346, *SD*=1.483) and RWG (*M*=5.576, *SD*=1.193) groups demonstrated that there was nearly no difference with the latter retaining very slightly more words than the former. The results in this sense imply that when the total TILL is equal, the different distribution of search and evaluation components in a task does not affect VG and VR scores, which provides strong support for the TILH.

4.3. A Brief Summary of the Results

The comparison of the groups on immediate and delayed post-tests showed that the results were consistent across the two reading texts used in the study. The mean scores of the different groups on immediate and delayed post-tests across two implementations are presented in Table 16.

Table 16. Mean Scores of Groups across the Implementations

	Implementation 1		Implementation 2	
	Immediate	Delayed	Immediate	Delayed
Fill-in with Glossary	8,4	5,233	8,758	4,742
Fill-in by Searching	8,513	5,385	8,846	5,346
Retelling with Glossary	8,6	5,633	9,076	5,576
Retelling by Searching	9,172	6,406	9,294	6,309

Starting with the immediate VGs of the different loaded groups, it was found that an increase in total TILL, regardless of the components increased, resulted in an increase in the VG scores for both texts. The participants in the lowest loaded task group, FWG, got the lowest scores followed by FBS and RWG. The highest VG scores were obtained by the FBS group with the highest TILL. However these increases did not reach significance.

The results of the comparison on the delayed post-test scores also showed that increases in the total TILL affected VR positively. The participants who got the lowest retention scores were in the FWG group, followed by FBS and RWG. The highest scores belonged to the RBS group which completed the highest loaded tasks and significantly outperformed those in FWG and FBS groups. There were not any other significant differences between the groups, which showed that adding search to tasks with similar level of evaluation component and increasing the evaluation component without also inserting search did not lead to significant changes on VG and VR.

As for the effect of equal TILLs, the tasks with equal involvement loads were found to be resulting in similar VG and VR as the comparison between FBS and RWG groups demonstrated. The groups obtained similar scores on both immediate and

delayed post-tests, which showed that regardless of the distribution of the components in different tasks, the results will be similar when the TILL is equal.

4.4. Discussion of the Findings

In order to answer the research questions of the present study, four different tasks with varying TILLs were designed. The tasks under investigation were fill-in with glossary (FWG), fill-in by searching (FBS), retelling with glossary (RWG) and retelling by searching (RBS). Assigning the participants to such tasks made it possible to divide the participants into two broad groups as with-glossary and by-searching and then the other two groups as fill-in and retelling. Such groupings would be useful to control the search and evaluation components induced by the tasks. The two by-searching tasks, FBS and RBS, required the participants to look for the meanings of the words thus contained the search component. On the other hand, the with-glossary tasks, FWG and RWG lacked the search component. In this way, it would be possible to test the effect of inserting search into similar tasks. As for the evaluation component, fill-in tasks requiring the participants to compare a limited number of words induced moderate need while retelling tasks contained strong evaluation since the participants were required to refer to their existing vocabulary and syntactic knowledge to complete the task. Therefore, increasing the presence of evaluation component from moderate to strong would be investigated. Total TILLs operationalised by the tasks were 2, 3, 3, and 4 respectively. The tasks contained moderate need but differed in terms of how search and evaluation components were distributed. In the end, tasks with both differing and equal TILLS were achieved to provide answers for the research questions.

In order to answer the first research question seeking an answer to the inquiry whether tasks with differing TILLs significantly differed from each other in terms of their contribution to VG and VR, all the lower loaded tasks were compared to higher loaded tasks on the immediate and delayed post-tests. Similarly, the two tasks with equal TILLs were compared on the vocabulary post-tests so that the second research question addressing whether different tasks with equal TILLs would yield similar results could be answered. Similar comparisons were conducted for both texts. As a result of the multiple comparisons on vocabulary scores obtained on both texts, it was

found that the line-ups of the groups and the difference between the scores were consistent over two texts. This finding implies that involvement load level was a strong predictor of vocabulary gain and retention regardless of the different target words and texts chosen. Therefore, it will not be inappropriate to discuss the results of both implementations as a whole, and thus, the discussion in the next section applies to both texts.

4.4.1. The Effect of Tasks with Different Task-induced Involvement Loads

In the present study, the total TILLs of the tasks were adjusted by adding search component into a task, increasing the level of evaluation component from moderate to strong or inserting search and increasing the presence of evaluation at the same time. The with-glossary tasks lacked search component while by-searching tasks induced additional search as in FWG and FBS; and RWG and RBS groups. Similarly, fill-in groups lacked strong evaluation while retelling groups induced strong evaluation. Therefore, while evaluating the results obtained from tasks with different TILLs, the effect of adding search or increasing the evaluation component or both will be discussed.

The statistical analysis conducted on the immediate post-test scores for both of the reading texts showed that all the groups differed from each other with the higher loaded tasks yielding higher scores. The line-ups of the groups and the differences among them were the same across two texts. The groups were lined up parallel to their respective TILL. The further analysis to determine whether equipping a task with a higher TILL either by inserting search or increasing the level of evaluation was significantly beneficial yielded that the short term benefits may not be as much salient as expected.

When the groups completing the similar tasks with and without glossary were compared, it was found that consulting a dictionary for the TWs indeed made a difference. Even though most of the participants consulted their smart phones for the meanings of the TWs, the seemingly effortless act of typing the words on their screens brought about higher word gains. However, although the groups using a dictionary for the meanings of the TWs slightly outperformed their peers completing the tasks without

searching the TWs, the difference between them was not significant in the short term. The insignificant difference caused by the insertion of search component is something unforeseen by Laufer and Hulstijn (2001), but not new. Konno et al. (2009) found that search component had negative effects on immediate and long term retention of words. They attributed this effect to the fact that their participants were not good at exploiting external sources. Another possible explanation they provided was that the participants were too overwhelmed with finding the appropriate meaning of the TWs that they could not actually focus on the real meaning. Similarly, Marmol and Sanchez-Lafuente (2013) underlined the importance of the participants' proficiency levels when incorporating dictionary use. Their study with relatively lower level participants proved that search component may be counterproductive rather than beneficial. Another possible reason could be that the participants did not actually look up the meanings of the target words, as it was the case in Li (2014). Checking the online behaviours of the participants, Li (2014) found that search component was not effective in bringing about significant differences as the participants did not actually look up the meanings of the target words.

However, the present study did not yield results signalling that searching the meanings of the TWs was a counterproductive process but indeed it was proved that letting participants search the meanings led to higher scores, though slightly. In the present study, the participants were from English Language Teaching program with higher proficiency levels and more experience in consulting dictionaries and it was made sure that they used dictionaries. All the counterproductive factors mentioned in the former studies were eliminated as much as possible and this might have resulted in slight superiority of the by-searching groups. In this sense, the results resemble to Sarani et al. (2013) who collected data from English major university participants selected based on their TOEFL scores. The participants' level of proficiency and relative experience in exploiting external sources for meanings allowed the search component to take effect and reached significantly higher scores. The effect of the participants' proficiency and experience in dictionary use on incidental vocabulary acquisition as a result of consulting a dictionary was also proven by Yaqubi et al. (2010). They found that making students involved in a searching process was indeed helpful in bringing about significantly higher immediate word gains. To this end, it could be argued that how search component in the TILH functions may well depend on

the learners' proficiency level and their dictionary use habits. While participants from higher proficiency levels may pay more attention to the several meanings of the words and how they are used in context, less proficient learners may hold on to first meaning only without paying attention to the other meanings and broader contexts. Therefore, it can be suggested that when designing tasks inducing search component, the characteristics of the participants should be taken into consideration. All learners in different contexts may not benefit from using dictionaries. More seriously, lower level participants may just ignore looking up the meanings of the targeted words. All these factors preventing search from taking effect should be observed carefully.

In the present study, the other way of increasing the total TILL was strengthening the presence of evaluation from moderate to strong. The comparisons between the two fill-in tasks inducing moderate evaluation and retelling tasks with strong evaluation showed that the latter tasks resulted in higher VG scores. Keeping the search component equal for different tasks and increasing the total TILL through evaluation component resulted in more VGs as the comparisons between FBS and RBS, and FWG and RWG groups revealed. However, the results somehow were not as expected by Hulstijn and Laufer (2001) as the difference was not significant. In order to find out what might be the reason for this insignificant difference, looking into some other studies yielding inconclusive results proved to be helpful. Loading a task with strong evaluation was found to cause failure in promoting higher word gains due to learner deficiencies in meeting the requirements of the tasks. Walsh (2009) for example found that making students create original sentences (strong evaluation) did not lead to significantly higher scores than fill-in tasks and attributed this to deficiencies in the participants' vocabulary knowledge which may possibly prevent making strong connections with the TWs and existing vocabulary. Similarly, Van Polen (2014) suggested that requiring the participants to complete a reading comprehension task and a composition writing at the same time, as was the case in the current study, might be too overwhelming and consuming the learners' attentional sources thus causing blank-filling task with moderate evaluation to produce better results. Another possible explanation Van Polen (2014) came up with was that the participants may well have not met the requirements of the task appropriately due to neglecting to use some of the TWs in the composition. However, the insignificant results obtained in the present study are

no way close to these studies. First of all, increasing the evaluation from moderate to strong was actually beneficial as seen in the slight differences. Secondly, the participants who did not write a well-developed version of the texts with all the TWs incorporated were excluded from the data so as to determine the actual effect of thoroughly attending to the task. Beal (2007) also found that sentence writing group did not obtain higher scores than multiple choice group despite the superiority of evaluation component induced. The possible reason, according to the researcher, was that the participants did not meet the requirements of the vocabulary tests fully. It is again the learner-related factors interfering with the components thus preventing the tasks to take effect. Then it can be argued that learner-related factors such as the ability to write, attention span, vocabulary knowledge and proficiency level should be taken into consideration while designing tasks with varying TILLs. Moreover, close inspection is needed to make sure that the participants attend to the requirements of any task.

Another study proving that involving students in a sentence writing task was not significantly more beneficial than fill-in tasks was by Jahangiri and Abilipour (2014). However, this study used a very different methodology by involving the participants in fill-in task twice so as to equalise the time needed for different tasks. Therefore, the results cannot be taken for granted as there was clearly exposure interference, with the fill-in task exposing the participants to the target words twice. Although this may have some practical implications, the results cannot be evaluated against the assumptions of Laufer and Hulstijn (2001).

The aforementioned explanations for insignificant or even inferior results obtained from tasks with strong evaluation do not seem to be relevant to the present study. Conversely, the vocabulary gains under tasks with strong evaluation was actually slightly better than under tasks with moderate evaluation. Attending to the form and meaning associations of the target words by means of composition writing and dictionary look-ups was found to be contributing to the overall word gains. Considering that all the participants recalled nearly all the TWs, the most plausible explanation for the slight differences can be as Kim (2008) suggested that the effect of higher loaded tasks may not be salient in the short term but may well become evident as time elapses.

As for the delayed post-test-scores of the two texts, it was once again found that increases in the total TILL were accompanied by increases in vocabulary retention scores. The lowest loaded task resulted in the lowest retention scores and the most involving task yielded the highest retention scores. Regardless of the component the presence of which was adjusted to increase the total TILL, loading a task with more involvement level led to higher VR scores. Requiring the participants to look for the meanings of the TWs was more effective than providing them with a glossary of the TWs. The results acknowledging the importance of dictionary look-ups in immediate recall and retention of words corroborate the assumptions of Laufer and Hulstijn (2001) and the earlier studies by (Hulstijn et al., 1996; Knight, 1994; Luppescu & Day, 1993; Newton, 1995) who, as well, demonstrated that the words whose meanings were looked up in dictionaries were retained better. In this regard, introducing the element of consulting dictionaries can be argued to deepen the initial process of the words, thus leading to better retention.

Similar results were obtained by strengthening the presence of evaluation component in a task. The comparison between the groups completing fill-in tasks and creating a text by incorporating the TWs showed that the latter benefited from the task more in terms of VR scores. As suggested by Laufer and Hulstijn (2001), increasing the evaluation component from moderate to strong was effective in increasing the depth the target words were processed and led to better retention. Therefore, it can be posited that keeping some learner characteristics under control and making sure that the students meet the requirements of the task, requiring learners to write a text based on the reading passage using the target words will make the level the target words are processed deeper and therefore, lead to retention of more words.

Similar to the immediate post-test scores, adding search and increasing evaluation component at the same time was also associated with higher numbers of words retained. In this sense, the results of the delayed post-test were quite similar to those of the immediate post-test. Inserting search and increasing the evaluation component alone did not lead to significantly higher numbers of word retention. This can be attributed to the relative difficulty of the TWs as Kim (2008), Hulstijn and Laufer (2001), and Tu (2004) suggest, difficult words such as academic words and false cognates can be best retained through higher loaded tasks. It was the case in the

retelling by searching task including the highest involvement load among the groups. The results of the one-way ANOVA showed that the retelling by-searching task with the highest TILL retained significantly more words than fill-in with glossary and fill-in by-searching tasks. The fact that the groups were significantly differing from each other on the delayed post-test proves Kim (2008) right in the supposition that although immediate gains of higher loaded tasks may not seem to be worth the while, the long term effects can be noteworthy. When the distribution of the components of search and evaluation in the FBS task is investigated, it can be seen that search component was present, meaning that the participants completed a process of searching the meanings of the TWS. Additionally, strong evaluation existed, which meant that the participants had to connect the TWs to their existing vocabulary and syntactic knowledge for creating a text. This finding may provide some useful insights into the relationship between the components of involvement. The components of search and strong evaluation coming together, it can be suggested, resulted in significantly higher scores, which was not the case when they appeared separately. In this regard, there emerges another point to consider while adjusting the components of involvement load. The second experiment in Tu (2004), which investigated evaluation component only, showed that composition writing group lost their initial vocabulary gains over time and retained similar number of words as the blank filling group. However, the highest loaded group in this study did not lose their advantage. This might be due to the fact that in addition to strong evaluation, there was also search component present in retelling task. This result may indicate that the two components strongly support each other and lead to retention of more words. To this end, unlike the findings in Konno et al. (2009), Li (2014), Marmol and Sanchez-Lafuente (2013) who found dictionary use was overwhelming, and Soleimani et al. (2014), Walsh (2009), and Van Polen (2014) who suggested that requiring strong evaluation was counterproductive, this study proves that when the participants attend to the requirements of the tasks, dictionary use and composition writing lead to retention of more words. Similarly, it can be suggested that rather than trying to determine whether search or evaluation is more effective, their interaction in determining word gains in different tasks should be explored.

The answer to the first research question asking whether different TILLs would lead to differences in VG and VR scores is yes to a large extent since every step taken

to increase the total involvement brought about increases on both immediate and delayed post-test scores. However, there are two important concerns to keep in mind regarding the effect of increasing the involvement load of a task on incidental acquisition of vocabulary: designing the tasks according to the audience's proficiency levels and monitoring the students while they do the tasks. The current study trying to minimise the learner-related factors by means of two measures: a) selecting a more proficient sample with relatively larger vocabulary to carry out the demanding tasks, and experience in exploiting external sources like dictionaries and internet b) excluding the participants who did not attend to the requirements of the tasks appropriately, provided evidence that learners can actually benefit from higher TILLs on immediate and delayed recall of the target words. The groups involving in searching process outperformed the groups who were provided with the glossary. Similarly, the composition writing groups outperformed the fill-in groups. However, it was found that all the groups recalled nearly all of the TWs. This result may imply that the relative benefits of involving in tasks may depend on the learner characteristics such as proficiency levels and whether the requirements of the tasks are fulfilled.

Although the differences between the groups on the immediate post-test were insignificant, as Kim (2008) posited the overall effects may be salient in the long-term, there was a significant difference between the groups on the delayed post-test scores. The results obtained on the delayed post-test provided stronger support for Laufer and Hulstijn (2001) in that significantly higher number of words were retained by the participants completing the most involving task. An interesting finding, which was not found or investigated at all in the former studies, was that when participants took part in searching and passage production at the same time, the benefits were found to be more substantial. Similarly, using dictionaries without also creating a text and vice versa was found to be inefficient to bring about significant vocabulary retention, which could be attributed to the suggestion of Hulstijn and Laufer (2001), and Tu (2004) that challenging words may require more involving tasks for retention. This could have useful indications for practice.

By providing support for the TILH framework, the results of the study have much to contribute to the incidental acquisition of unknown words encountered during reading. As Eysenck (1982) put forward, the retention of the information is contingent

upon how deep this piece of information is processed rather than the intention to learn it. Also drawing on the framework of task-induced involvement load by Laufer and Hulstijn (2001), the study provided proof that involving the learners in tasks with more cognitive load, be it consulting dictionaries or retelling a text by forming connections between the new words and existing syntactic and semantic knowledge, will bring about retention of more words. Even though the participants did not pay special attention to commit the words into their long term memory for preparing for an upcoming test, or highly probably so, taking part in a vocabulary task which also serves reading comprehension checking purpose was effective in retaining large numbers of the words addressed by the tasks. In this regard, Craik and Lockhart (1972) and Laufer and Hulstijn (2001) were proven right in their assumptions that processing the words at higher levels of cognitive load will bring about more words incidentally acquired.

4.4.2. The Effect of Tasks with Equal Task-induced Involvement Loads

The present study also made an attempt to find out whether tasks with equal TILLs would yield similar results, regardless of how evaluation and search components were distributed. As Laufer and Hulstijn (2001) and Hulstijn and Laufer (2001) expressed their concerns that the search and evaluation components may not carry the same weight and further studies should test this issue with different distribution of these components, there was need for further verification. To this end, fill-in by-searching and retelling with glossary tasks were designed each with a TILL of 3. The former task contained moderate need, moderate evaluation and search while the latter induced moderate need and strong evaluation. In more practical terms the absence of creating a text was filled with a searching process in the first task and the absence of searching process was filled with creating a text in the second task. Neglected by most of the studies aiming at testifying the validity of task-induced involvement load hypothesis, measuring the effect of different distributions of evaluation search components can yield further verification of the framework as well as practical implications for classroom practice.

The comparison of the two groups on the immediate post-tests of the two texts indicted that the latter obtained slightly higher VG scores than the former. However, this superiority was not found to be significant. The results revealed that getting

involved in a searching process to choose the most appropriate words was not significantly different from creating a different version of the text without looking up the meanings of the TWs. Therefore, it can be inferred that both search and strong evaluation compensated for the lack of each other in contributing to immediate VG. The former studies aiming at exploring the effect of tasks with equal TILLs operationalised with a different distribution of search and evaluation by Ghabanchi et al. (2012), Pourakbari and Biria (2015) and Sarani et al. (2013) also found that when participants processed the TWs through tasks with equal TILLs, they gained similar number of words. Along with the results of these investigations, the results of the current inquiry suggest that a fill-in task can be made to yield as much word gain as a composition writing task does with the help of search component. Worded differently, an easier task requiring less time and mental effort to complete can be as beneficial as a more demanding task with the addition of a dictionary look-up element.

Similarly, the delayed post-test results of the two texts showed that the two groups with similar TILLs retained similar numbers of words. Although the participants who wrote their own versions of the text with the help of a glossary got slightly higher scores than those filling in the blanks by using a dictionary, the difference was not significant. It was seen that a lack of a specific component could be compensated for by the presence of another. To this end, the findings seem to be in line with the findings of Hazrat (2015), Pourakbari and Biria (2015), and Sarani et al. (2013) who compared tasks with equal TILL but containing different distribution of search and evaluation components in terms of their contribution to incidental retention.

These insignificant differences between the two equally loaded groups on both reading texts demonstrate that gaps in the total TILL in a task caused by a specific component can be filled by introducing another component. However, when designing equally loaded tasks with different distributions of search and evaluation, care should be taken as learner-related factors may prevent the components from taking effect. Based on the findings regarding the dictionary look-up element for example, when learners do not actually look up the words or they do not know how to exploit dictionaries at all, it is not likely that search component will compensate for the lack of strong evaluation. Similar caveats can be applied to creating a text, or a sentence as learners with limited capacity of word knowledge and writing skills may not benefit from strong evaluation.

Therefore, the answer to the second research question regarding the effects of two different tasks with equal TILLs on the VG and VR scores is a fully supporting yes. Both the immediate and the delayed post-test scores of the two groups on both reading texts were quite similar with the latter group obtaining slightly higher scores. The results regarding the effect of equally loaded tasks on VG and VR are in harmony with the suggestion of Laufer and Hulstijn (2001) that the total TILL is more effective than the nature of the task, be it an input or an output task. Similarly, it can be assumed that the components in the construct of involvement load carry the same weight as long as the effects of learner characteristics are controlled and the requirements of the tasks are met appropriately.

These results regarding the effect of different tasks with equal involvement load levels operationalised by a different distribution of search and evaluation components fully entertain the task-induced involvement load hypothesis. The results signal that Laufer and Hulstijn (2001) and others like Kim (2008), Yaqubi et al. (2010), Konno et al. (2009), Jahangard (2014), Li (2014), who hold doubts about the equality of the components, might be too prejudiced about the efficacy of search component. When exploited appropriately by taking into consideration the class dynamics, letting the students consult dictionaries for the meanings of the target words creates a deeper process and thus leads to as much retention as a composition writing task without consulting a dictionary brings about. Therefore, it can be stated that the search component should not be underestimated and used carefully to compensate for the lack of strong evaluation in a task.

CHAPTER 5: CONCLUSION AND IMPLICATIONS

5.1. Introduction

The present chapter is devoted to providing a summary of the findings of the current investigation and present various implications for classroom practice and for the literature of the framework tested.

5.2. An Overview of the Present Study

The present study using participants from an ELT program in Turkey attempted to investigate the effect of task-induced involvement load level (TILL) on incidental vocabulary gain (VG) and vocabulary retention (VR) by comparing four tasks designed within the framework of task-induced involvement load hypothesis (TILH). Two different reading texts containing 10 target words (TWs) were used and the texts were accompanied by different vocabulary tasks.

The tasks under investigation were fill-in the blanks with glossary (FWG), fill-in the blanks by searching (FBS), retelling with glossary (RWG) and retelling by searching (RBS). The overall TILLs of the tasks were 2, 3, 3 and 4 respectively. By means of implementing tasks with such variations of TILL, the present inquiry attempted to find out a) the effect of different tasks with differing TILLs and b) the effect of different tasks with equal TILLs operationalised by a different distribution of search and evaluation components on incidental VG and VR. Unannounced immediate and delayed post-tests were administered to measure the VG and VG of the participants. The scores the participants obtained on these post-test were compared to investigate the effect of TILL on vocabulary acquisition

The statistical analysis run on the data of both reading passages revealed that the results were consistent across different texts and a similar pattern emerged in terms of the differences of the groups based on their immediate VG and VR. Therefore, the results to be reported here apply for both of the texts.

5.2.1. Tasks with Different Task-induced Involvement Load Levels

In order to answer the first research question and determine whether different TILLs had effects on incidental VG and VR, the tasks with different TILLs were compared on their immediate and delayed post-test scores. It was found that the participants were able to remember nearly all of the targeted lexical items, regardless of the involvement load level at which they processed these words. The comparison between the groups in terms of their immediate VG showed that with the increase in the total TILL, there was also increase in the VG scores. However, the higher scoring groups did not differ from the lower scoring ones significantly. Introducing the component of search into similar tasks did not bring about the expected significant differences in VG scores. This was proven by the comparisons between FWG and FBS, and RWG and FWG which revealed that despite the inclusion of search, the FBS and RBS tasks did not result in significantly higher scores than FWG and RWG tasks lacking search. Similarly increased presence of evaluation did not result in significant differences on the immediate VG scores either, as the comparison between FWG and RWG, and FBS and RBS groups revealed. Although composition writing tasks yielded slightly higher scores than fill-in tasks, these differences were not found to be significant. Slight differences can be attributed to the fact that the participants were able to remember nearly all of the target words.

When the groups were compared on their delayed post-test scores, a similar pattern emerged as increasing the overall TILL resulted in increased levels of VR. However, unlike the immediate post-test results, it was found that the group completing the RBS task which induced the highest TILL significantly outperformed the FWG and FBS groups. This finding suggested that while strong evaluation and search components alone were not effective in retention of the TWs, they led to significantly higher retention scores when they existed in a task together. When the participants were required to look for the meanings of the TWs themselves and retell the text by using the TWs, they retained significantly more words. Considering the relative difficulty of the selected words, it can be concluded that participants have to get involved in searching and writing processes simultaneously for retaining difficult words. The assumptions of TILH regarding the effect of different TILLs on VG and VR were partially supported by these findings.

5.2.2. Tasks with Similar Task-induced Involvement Load Levels

As for the second research question regarding the equally loaded tasks, the two groups processing the TWs at the same TILL were compared on their immediate VG and VR. The comparison between the FBS and RWG which induced a total TILL of 3 would determine whether inserting search into or increasing the presence of already existing evaluation of a task was more effective in promoting incidental VG and retention. When the immediate post-test scores of the two groups were compared, it was found that there were not any statistically significant differences in terms of immediate VG although the RWG group obtained slightly higher scores. This finding suggests that using a dictionary to complete a fill-in task which includes comparing limited numbers of words to each other was equal to a retelling task with the help of a glossary which requires comparing the newly encountered words with already existing linguistic knowledge. Getting involved in a searching was found to be compensating for the lack of strong evaluation in task. Similarly, the mental effort invested into making several associations between the TWs and existing vocabulary knowledge compensated for the absence of search.

A similar pattern emerged in the delayed post-test scores. The participants assigned to two tasks with the same TILLs but a different distribution of evaluation and search components were found to be not significantly differing from each other in terms of retaining the words over a two-week period. Introducing search in the stead of strong evaluation and similarly filling in the gap caused from the absence of search by strong evaluation resulted in similar VR scores. Inserting search into a task with moderate need and evaluation proved to be as effective as increasing the level of evaluation from moderate to strong. Therefore, it can be concluded that the components in the TILH, however their presence is adjusted, similarly contribute to VG and VR.

5.2.3. Summary

The present study set out to investigate the effect of task-induced involvement load on incidental immediate vocabulary recall and vocabulary retention by means of four vocabulary tasks operationalising both different and equal TILLs. The groups assigned to these tasks were compared on their immediate and delayed post-test scores to

determine a) whether different tasks with different TILLs led to significant differences and b) whether different tasks with equal TILLs led to similar VG and VR scores.

Firstly, the results obtained from two different reading texts demonstrated that the effect of TILL was consistent over different texts and different target lexical items. Regarding this finding, it can be stated that the construct of involvement load has a great deal of predictive value when learner-related factors are kept under control. Implementing different texts and targeted words on the same participants by sticking to similar procedures will enable the construct to determine the potential benefits of different tasks.

Secondly, the immediate post-test comparison of the groups indicated that although higher loaded tasks brought about slightly higher mean scores, these differences were not significant. This is not a result expected by the TILH which assumes that regardless of the components being increased, tasks with higher TILLs will lead to significant differences. However, all the participants regardless of their groups were already able to retain nearly all the targeted words, which might have prevented the groups from differing significantly from each other. As Kim (2008) suggested, the effects of tasks with different involvement load levels may not necessarily be salient immediately but rather become evident over time.

Thirdly, the delayed post-test scores were highly contingent upon the TILLs of different tasks. The significant benefit was only achieved when the search and strong evaluation existed in a task at the same time. Worded differently, probably because of the relative difficulty of the target words, the retention required both dictionary look-ups and text creation, which presents the complementary relationship between the search and evaluation components.

Fourthly, the equally loaded tasks were found to be yielding similar results on both immediate and delayed post-tests. The assumption of the TILH that regardless of the mental efforts or the time different tasks require, it is the total TILL that will determine the overall vocabulary gain and retention over time. This was the case with the gap filling by searching and retelling with glossary tasks in this study. Although the latter required more time and mental effort to complete, the scores of the two groups on

both post-tests were similar, which may have some useful implications for classroom practice.

In a nutshell, the present study provided strong support for the predictive value of the task-induced involvement load hypothesis, which many studies could not achieve due to learner-related factors. It can be suggested that, keeping in mind the specific features of the learners in a specific context, the construct of involvement load can be exploited to design different tasks conducive to incidental vocabulary acquisition.

5.3. Implications

The present study was an attempt to test the construct of TILH proposed by Laufer and Hulstijn (2001) by designing different tasks with varying TILLs and find out the best vocabulary task in promoting incidental VG and retention. Therefore, the results of the study offer some implications for both the literature of TILH and classroom practices addressing the issue of vocabulary acquisition. Firstly, how the literature of TILH may benefit from the findings and then how teachers and material designers can exploit the results for classroom practices to promote vocabulary acquisition will be discussed.

The study aimed to verify the TILH through some tasks which differed in terms of TILL. The groups completing these tasks were compared on immediate word gains and retention of words over a two-week period. The results showed that the groups completing the tasks with higher TILLs were better at both immediate lexical gain and retention, which supports the assumption of TILH that tasks with higher TILLs lead to higher rates of vocabulary acquisition. Regardless of the components adjusted, the higher TILLs were found to be associated with higher vocabulary gains. This suggests that all the components in the construct carry the same weight. However, the interesting finding that only the retelling by searching task, which induced search and strong evaluation at the same time, yielded significantly higher vocabulary retention may indicate a strong interaction between look-ups and creating a text. When either one of these components of search and strong evaluation is absent, significance cannot be achieved. These components, when they were present in a task alone, did not bring about significantly higher vocabulary retention but when they appeared in a task together, they were effective in promoting vocabulary retention. In this sense, caution

should be taken when considering the relationship between the search and evaluation components of TILH. There may not be clear-cut levels and sometimes two components coming together can build up a "level". As suggested earlier, difficult words may consume all the attentional and cognitive sources and therefore, both dictionary look-ups and making connections between the new words and existing vocabulary and syntactic knowledge may be necessary to handle these words (Hulstijn & Laufer, 2001; Kim, 2008, & 2004).

The other finding that tasks with equal TILLs did not differ from each other although they contained different distribution of search and evaluation components, showed that these components can compensate for the absence of one another. There was not strong evaluation in the fill-in by searching task and similarly there was not search in the retelling with glossary task. However, the search component in the FBS task filled in the gap caused by the absence of strong evaluation component. Similarly, increasing the presence of evaluation component from moderate to strong in the RWG made up for the subtraction of search component. This finding suggests that the components are equally effective in determining the contribution of a task to the total amount of lexical gain and retention. Therefore, Laufer and Hulstijn (2001), Kim (2008), and Li (2014) may have underestimated the efficiency of search by stating that not all the components may carry the same weight. When it is ensured that the participants exploit external sources (designing tasks necessitating dictionary use), search component can be as effective as the evaluation component. However, certain distinctions between the types of search can be made as searching the words on different kinds of dictionaries may not be similar to each other. The present study did not find any difference caused by different types of dictionary use.

The present study while trying to verify the construct of task-induced involvement load, also compared some vocabulary tasks conducive to incidental vocabulary gain and retention. Firstly, the study found that even if the learners were unaware of the fact that they were to be tested on the TWs later, they incidentally acquired and retained most of the TWs by completing the tasks they were assigned to. Therefore, involving the learners in tasks where they are to use the words to accomplish some tasks can be a useful tool for drawing their attention to important words thus promoting incidental vocabulary gain and retention. As Eysenck (1982) underlined, the

storage of the new information in the memory does not depend on the willingness to learn but rather how deeply this knowledge is processed at the first encounter. Along with the previous research on the effectiveness of additional tasks over reading alone in promoting incidental vocabulary acquisition, the results strongly support this assumption as none of the participants were instructed to learn the target words or forewarned of an upcoming retention test. Yet most of the target words were found to be retained due to the fact that these words were processed through some tasks not specially designed for retention purposes but rather serving comprehension checking. Even the lowest loaded task was effective in bringing about substantial vocabulary gain, which underlines the importance of drawing the learners' attention to the selected words. Therefore, teachers should not be hesitant to exploit additional tasks as often as their curriculum permits.

Secondly, as the findings suggest, learners will benefit from vocabulary tasks with higher TILLs more when compared to those with lower TILLs. Designing tasks which include higher levels of evaluation and search components will help learners gain and retain more words. When learners are involved in searching the meanings of the TWs themselves and use these words to create a composition, they will be better at retaining these words. However, drawing on the results of the former studies with lower level participants, the relative effectiveness of involvement load will depend on the learner characteristics such as attention span, writing skills, vocabulary knowledge and dictionary use habits. These factors should be taken into consideration before implementing vocabulary tasks. Depending on the needs and characteristics of the students, preparing tasks which encourage original composition writing and consulting dictionaries could be beneficial for the acquisition of difficult words as Hulstijn and Laufer (2001) suggested. However, the relative efficacy of using dictionaries or composition writing will depend mostly on how the learners can handle these tasks. Consulting dictionaries, for example, will not make any difference if the learners do not know how to interpret the information the dictionary entries provide. Similarly, if the learners in specific context have not yet developed the required writing skills for creating texts, there will be no point in employing such kinds of tasks as the retention of the words depend on the success of the students in putting these target words in appropriate and well-connected contexts. If learners lack vocabulary knowledge and

skills to form strong connections with the new words and existing knowledge as found, they may just replicate the reading text and copy the sentences as they appear in the text. Therefore, the connections will remain superficial thus leading to forgetting the words immediately or not acquiring them at all. Moreover, another point while considering the classroom dynamics is whether the learners are actually doing what the tasks require them to do. As found by previous researchers (Li, 2014; Van Polen, 2014) students may from time to time be inefficient to catch up with the demands of a task as it is too overwhelming or due to lack of practice in the steps to take towards completing the tasks. Therefore, close inspection is of crucial importance while employing such tasks. If the participants do not meet the demands of the task, there will be no use in implementing them since the potential benefits will not take effect.

The third suggestion which can be useful for classroom practices is that two tasks with equal TILLs result in similar vocabulary gain and retention when the requirements of the tasks are fully met. The study found that a gap fill task by using a dictionary which required 40 minutes, and a retelling task with the help of a glossary lasting for 50 minutes respectively, were not significantly different from each other in terms of their contribution to immediate recall of vocabulary and retention of these words over a two-week period. The crucial point here is that whether making the learners look up the meanings of the target words and complete a gapped text by using these words or requiring them to produce a text by providing them with the meanings of the words are equally effective for vocabulary acquisition. Then it is the teacher's call to choose either one of them based on the classroom time they have or the skill they want to improve. As mentioned before, if the students' attention span and writing skills are limited, just encouraging dictionary use will be as beneficial as requiring them to produce a text, which may be challenging for certain contexts. As Hulstijn and Laufer (2001) suggest, less proficient students can be made to process the target words with search and moderate evaluation rather than strong evaluation. A balance between being involving and overwhelming can be achieved by distributing the search and evaluation components in tasks by considering the results of the present study. Considering the relative ease of fill-in tasks when compared to writing compositions and the fact that gap fill tasks can be reinforced with search component to result in similar vocabulary gains as writing a text, it will be logical to opt for the former task if learners are not

capable of writing. Similarly, based on the classroom dynamics, if the learners are capable of producing written texts, teachers can choose the latter to develop their students' writing skills further as well as their vocabulary knowledge and reading comprehension.

Similar suggestions can be provided for material designers. First of all, while preparing reading texts in course books, some tasks can be included such as gap-filling summary writing incorporating the target words, matching, multiple choice that address the specific vocabulary encountered in those texts. Reading texts accompanied by tasks directing the learners' attention to specific words have unanimously been found to contribute to incidental vocabulary acquisition more than reading only conditions (Folse, 2006; Keating, 2008; Kim, 2008; Sarbazi et al., 20114). Therefore, reading related tasks can be a part of the curriculum so as to prevent students from neglecting important words. For ensuring that the tasks are actually accomplished, some measures as including these tasks as a part of the examination process or including these tasks as homework that will be graded later can be taken by designing the books accordingly.

As suggested by the results, several distributions of evaluation and search can be operationalised to reach equal different involvement load levels. Therefore, it may be possible to prepare reading-based vocabulary tasks for different audience with different needs and characteristics. For example, for younger learners, less challenging tasks such as gap-fills and matching can be designed while writing an argumentative essay writing task incorporating specific vocabulary items encountered in a reading text can be prepared for more advanced learners. As Hulstijn and Laufer (2001) suggested, lower level students can benefit from a combination of moderate evaluation and search components. Rather than forcing the learners to produce written language when they are not fully ready to handle the mental requirements of this task, involving them in tasks that they can handle and are as affective as the more demanding tasks will be practical. This way, they can help teachers and students in their endeavour to build up a large vocabulary stock, necessary for competence in L2.

In sum, using the framework of task-induced involvement load and specifically the findings of the present study, both material designers and teachers can prepare reading materials accompanied by tasks which both serve checking reading

comprehension and direct the learners' attention to the meanings of specific words, which are assumed to be important for the learners. Especially, tasks which require learners to search the meanings of the target words and create a text by incorporating these words can be designed for promoting vocabulary gain and retention. Also considering the dynamics of the classes such as class time and learners' competence in writing, the most suitable task among equally loaded tasks can be preferred over the others. Designing tasks that will best suit the proficiency level and capabilities of the learners will be of crucial importance as overwhelming tasks will be no way beneficial as the participants cannot attend to the target words thoroughly. Similarly, employing tasks which are not challenging enough for specific audience will not bring about the expected vocabulary expansion.

5.4. Suggestions for Further Studies

The present study using the framework of Laufer and Hulstijn (2001) tested the effect of gap filling and retelling tasks. The most obvious limitation is at this point: using the framework, many other vocabulary tasks can be designed and put under investigation. The present results may be generalised for the tasks under investigation, but further studies are needed for further verification of the framework by means of different tasks. By making adjustments on the presence of search, need and evaluation components of task-induced involvement load framework and designing different vocabulary tasks, further studies can be conducted for further verification of the framework. Especially, tasks with similar involvement load with similar distribution of the components can be compared as there are not many studies conducted on the issue (e.g. fill-in the blanks of a text and fill in the blanks of separate sentences; writing a composition and writing separate sentences). The comparison between such tasks will be of crucial importance for testing whether the acquisition of words is contingent upon the total involvement load level or on the level of discourse.

Similarly, focusing on ELT students, who have a certain level of proficiency and capabilities in using dictionaries and writing skills, the results of the current study may not be generalised for lower proficiency learners. In order to investigate the effect of

task-induced involvement load on incidental vocabulary acquisition more thoroughly, further studies can focus on different participants with different proficiency levels.

Moreover, for eliminating the effect of word category on acquisition of the target words, only nouns were put under investigation in the current study. This creates the need for further studies investigating whether the involvement load level produces similar findings regarding different word categories.

Additionally, to reduce text effect, the reading texts employed for the study were selected based on the internal structure with both texts written in compare and contrast pattern. Therefore, further studies can explore the effect of involvement load through different texts written in different patterns. This will be especially useful for investigating the effect of text type on the success of the participants in retelling and blank filling.

Moreover, the present study did not investigate whether different sources can be exploited to strengthen the effect of search component. To this end, further studies can be conducted to find out the effects of advanced tools such as COCA and BNC which provide more sophisticated information on how certain words are used in real contexts.

The present study aimed at investigating whether task-induced involvement load had an effect on acquiring the meanings of the target words. Further studies can shed light on the acquisition of the forms as well.

REFERENCES

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425-438. http://dx.doi.org/10.1093/applin/24.4.425

Al Seyabi, F., & Tuzlukova, V. (2014). Writing Problems and Strategies: An Investigative Study in the Omani School and University Context. *Asian Journal of Social Sciences & Humanities*, *3*(4), 37-48.

Al-Nouh, N. A., Abdul-Kareem, M. M., & Taqi, H. A. (2015). EFL College Students' Perceptions of the Difficulties in Oral Presentation as a Form of Assessment. *International Journal of Higher Education*, *4*(1), 136-150. http://dx.doi.org/10.5430/ijhe.v4n1p136

Anova, C., Antoni, R., & Kasyulita, E. (2015). The correlation between students' vocabulary mastery and speaking skill at fifth semester of English study program in Pasir Pengaraian University. *Jurnal Ilmiah Mahasiswa FKIP Prodi Bahasa Inggris*, *1*(1), 1-9.

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: a comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading*, *37*(1), 22-47. http://dx.doi.org/10.1111/j.1467-9817.2012.01538.x

Bao, G. (2015). Task type effects on English as a Foreign Language learners' acquisition of receptive and productive vocabulary knowledge. *System*, *53*, 84-95. http://dx.doi.org/10.1016/j.system.2015.07.006

Beal, V. (2007). *The weight of involvement load in college level reading and vocabulary tasks* (Doctoral dissertation). Retrieved from Concordia University CLUES Library Catalogue (LE 3 C66E38M 2007 B43).

Begriche, F. (2014). *The role of teaching vocabulary to enhance foreign language learners' writing skill* (Master's Thesis). Retrieved from http://dspace.univ-biskra.dz:8080/jspui/handle/123456789/4741

Behbahani, S. M. K., Pourdana, N., Maleki, M., & Javanbakht, Z. (2011). EFL task-induced involvement and incidental vocabulary learning: succeeded or

surrounded. In *International Conference on Languages, Literature and Linguistics. IPEDR Proceedings*, *26*, 323-325.

Bouhitem, T. (2013). EFL first year students' writing problems and proposed remedial strategies the case of 1st year university students at the English Department University of Biskra. *Sciences Humaines*, *29*.

Brown, D. (2000). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*(3), 301-315.

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. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Sage Publications: Thousand Oaks.

Croll, C. (1971). Teaching vocabulary. *College Composition and Communication*, *22*(5), 378–380. http://dx.doi.org/10.2307/356214

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge 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. http://dx.doi.org/10.1177/1362168811431377

Eysenck, M.W. (1982). Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 197–228). New York: Academic Press.

Fan, M. Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The*

Modern Language Journal, *87*(2), 222-241. http://dx.doi.org/10.1111/1540-4781.00187

Feng, T. (2015). Involvement load in translation tasks and EFL vocabulary learning. *The New English Teacher*, *9*(1), 83-101.

Fhonna, R. (2014). The correlation between mastering vocabulary and speaking ability (case study at SMA 10 Fajar Harapan Banda Aceh). *Visipena*, *5*(1), 90-99.

Florit, E., Roch, M., & Levorato, M. C. (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. *Reading and Writing*, *27*(5), 793-817. http://dx.doi.org/10.1007/s11145-013-9464-1

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273-293. http://dx.doi.org/10.2307/40264523

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in second language acquisition*, *21*(2), 225-241.

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

Ghabanchi, Z., Davoudi, M., & Eskandari, Z. (2012). Vocabulary learning through input and output tasks: Investigating the involvement load hypothesis. *California Linguistic Notes*, *37*(1), 1-18.

Grasparil, T. A., & Hernandez, D. A. (2015). Predictors of Latino English learners' reading comprehension proficiency. *Journal of Educational Research*, *5*(1), 35-57. http://dx.doi.org/10.5590/JERAP.2015.05.1.03

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language learning*, *46*(4), 643-679. http://dx.doi.org/10.1111/j.1467-1770.1996.tb01355.x

Hamouda, A. (2013). An investigation of listening comprehension problems encountered by Saudi students in the EL listening classroom. *International Journal of Academic Research in Progressive Education and Development*, *2*(2), 113-155.

Haratmeh, M. S. (2012). Involvement load and task type in task effectiveness: Two aspects of vocabulary knowledge. *International Journal of Academic Research, 4*(4), 86-95. http://dx.doi.org/10.7813/2075-4124.2012/4-4/B.13

Harmon, J. M., Buckelew-Martin, E., & Wood, K. D. (2010). The cognitive vocabulary approach to word learning. *English Journal*, *100*(1), 100-107. Retrieved from http://www.jstor.org/stable/20787701

Hazrat, M. (2015). The effects of task type and task involvement load on vocabulary learning. *Waikato Journal of Education, 20*(2), 79-92. http://dx.doi.org/10.15663/wje.v20i2.189

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

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

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

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. http://dx.doi.org/10.1111/j.1540-4781.1996.tb01614.x

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, *51*(3), 539-558. http://dx.doi.org/10.1111/0023-8333.00164

Huy, N. T. (2015). Problems affecting learning writing skill of grade 11 at Thong Linh high school. *Asian Journal of Educational Research*, *3*(2), 53-69.

Jahangard, A. (2014). Task-induced involvement in L2 vocabulary learning: A case for listening comprehension. *Journal of English Language Teaching and Learning*, *5*(12), 43-62.

Jahangiri, K., & Abilipour, I. (2014). Effects of collaboration and exercise type on incidental vocabulary learning: Evidence against involvement load hypothesis. *Procedia-Social and Behavioral Sciences*, *98*, 704-712. http://dx.doi.org/10.1016/j.sbspro.2014.03.471

Jeon, E. H., & Yamashita, J. (2014). L2 Reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160-212. http://dx.doi.org/10.1111/lang.12034

Jing, L., & Jianbin, H. (2009). An empirical study of the involvement load hypothesis in incidental vocabulary acquisition in EFL listening. *Polyglossia*, *16*, 1-11.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*(3), 357-377. http://dx.doi.org/10.1093/applin/19.3.357

Kaneko, M. (2015). Vocabulary size required for the TOEFL IBT listening section. *Language Teacher*, *39*(1), 9-14.

Karbal, H. T. (1975). Keying in on vocabulary. *Elementary English*, *52*(3), 367–369. Retrieved from http://www.jstor.org/stable/41592626

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. http://dx.doi.org/10.1177/1362168808089922

Keyvanfar, A., & Badraghi, A. H. (2011). Revisiting task-induced involvement load and vocabulary enhancement: Insights from the EFL setting of Iran. *Žmogus ir Žodis*, *13*(3), 56-66.

Kharaghani, N., & Ghonsooly, B. (2015). The impact of vocabulary knowledge on reading comprehension ability of Iranian English learners receiving reciprocal teaching and cooperative grouping intervention program. *International Journal of Research Studies in Education*, *4*(3), 47-56. http://dx.doi.org/10.5861/ijrse.2015.1006

Khotimah, S. (2014). The use of problem based learning to improve students' speaking ability. *ELT Forum: Journal of English Language Teaching*, *3*(1), 50-56.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language learning*, *58*(2), 285-325. http://dx.doi.org/10.1111/j.1467-9922.2008.00442.x

Kim, Y. S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child development*, *86*(1), 128-144. http://dx.doi.org/10.1111/cdev.12293

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal*, *78*(3), 285-99.

Konno, K., Takanami, S., Okuyama, Y., & Hirai, A. (2009). Examining the effects of involvement load on Japanese EFL learners' vocabulary retention. *JLTA Journal*, (12), 46-64.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, *73*(4), 440–464. http://dx.doi.org/10.1111/j.1540-4781.1989.tb05325.x

Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *IRAL-International Review of Applied Linguistics in Language Teaching*, *28*(4), 293-308. http://dx.doi.org/10.1515/iral.1990.28.4.293

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26. http://dx.doi.org/10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, *16*(3), 307-322. http://dx.doi.org/10.1093/applin/16.3.307

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, *46*(1), 101-135. http://dx.doi.org/10.1111/j.1467-1770.1996.tb00642.x

Li, J. (2014). Effect of task-induced online learning behavior on incidental vocabulary acquisition by Chinese learners: Revisiting involvement load hypothesis. *Theory and Practice in Language Studies*, *4*(7), 1385-1394. http://dx.doi.org/10.4304/tpls.4.7.1385-1394

Ling, G. U. I. (2015). Predictability of vocabulary size on learners' EFL proficiency: Taking VST, CET4 and CET6 as instruments. *Studies in Literature and Language*, *10*(3), 18-23. http://dx.doi.org/10.3968/6679

Little, A., & Kobayashi, K. (2015). Vocabulary learning strategies of Japanese Life Science students. *TESOL Journal*, *6*(1), 81-111. http://dx.doi.org/10.1002/tesj.141

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*(2), 263-279. http://dx.doi.org/10.1111/j.1467-1770.1992.tb00717.x

Ma, Y. H., & Lin, W. Y. (2015). A study on the relationship between English reading comprehension and English vocabulary knowledge. *Education Research International*, *2015*. http://dx.doi.org/10.1155/2015/209154

Mallett, J. J. (1975). Add Spice to Vocabulary Study. *Language Arts*, 52(6), 843-846. Retrieved from http://www.jstor.org/stable/41961128

Mármol, G. A., & Sánchez-Lafuente, Á. A. (2013). The involvement load hypothesis: the effect on vocabulary learning in primary education. *Revista Española de Lingüística Aplicada*, (26), 11-24.

Min, H. T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, *58*(1), 73-115. http://dx.doi.org/10.1111/j.1467-9922.2007.00435.x

Mokhtari, K., & Niederhauser, D. S. (2013). Vocabulary and syntactic knowledge factors in 5th grade students' reading comprehension. *International Electronic Journal of Elementary Education*, *5*(2), 157-170.

Nagy, W.E., & Herman, P.A. (1987). Depth and breadth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–35). Hillsdale, NJ: Erlbaum.

Namaghi, S. A., & Malekpur, A. (2015). Vocabulary learning strategies from the bottom-pp: A grounded theory. *The Reading Matrix: An International Online Journal*, *15*(2).

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

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, *11*(2), 159-176. http://dx.doi.org/10.1177/026765839501100207

Nowrouzi, S., Tam, S. S., Zareian, G., & Nimehchisalem, V. (2015). Iranian EFL students' listening comprehension problems. *Theory and Practice in Language Studies*, *5*(2), 263-269. http://dx.doi.org/10.17507/tpls.0502.05

Nurfuah, E. (2014). *Correlational study between students' vocabulary mastery, habit of watching English movie, and speaking skill of the eighth grade of SMP Islam Diponegoro Surakarta in the 2012/2013 academic year* (Doctoral dissertation). Retrieved from http://eprints.uns.ac.id/id/eprint/16181

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.174-200). New York: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in second language acquisition*, *21*(2), 195-224.

Pérez Manzanilla, I. S., & Díaz Cabrera, K. M. (2014). *Factors that may have an impact on advanced EFL students' speaking ability* (Master's Thesis). Retrieved from http://cdigital.uv.mx/handle/123456789/35250

Peters, E. (2012). The differential effects of two vocabulary instruction methods on EFL word learning: A study into task effectiveness. *International Review of Applied Linguistics in Language Teaching*, *50*(3), 213-238. http://dx.doi.org/10.1515/iral-2012-0009

Pflaum, S. W. (1973). Expansion of Meaning Vocabulary: Strategies for Classroom Instruction. *Elementary English*, 50(1), 89-93. Retrieved from http://www.jstor.org/stable/41388367

Pichette, F., De Serres, L., & Lafontaine, M. (2011). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*(1), 66-82. http://dx.doi.org/10.1093/applin/amr037

Pourakbari, A. A., & Biria, R. (2015). Efficacy of task-induced involvement in incidental lexical development of Iranian senior EFL students. *English Language Teaching*, *8*(5), 122-131. http://dx.doi.org/10.5539/elt.v8n5p122

Price, K. W., Meisinger, E. B., Louwerse, M. M., & D'Mello, S. (2015). The contributions of oral and silent reading fluency to reading comprehension. *Reading Psychology*, *37*(2). http://dx.doi.org/10.1080/02702711.2015.1025118

Putra, A. R. (2014). *Using picture series to improve the writing skill on recount of 8th graders SMP Muhammadiyah 3 Depok in the 2013–2014 Academic Year* (Doctoral dissertation). Retrieved from http://eprints.uny.ac.id/id/eprint/18475

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child development*, *86*(1), 159-175. http://dx.doi.org/10.1111/cdev.12292

Rassaei, E. (2015). Effects of three forms of reading-based output activity on L2 vocabulary learning. *Language Teaching Research*, 1-20. http://dx.doi.org/10.1177/1362168815606160

Readence, J. E., & Searfoss, L. W. (1980). Teaching strategies for vocabulary development. *English Journal*, *69*(7), 43-46. http://dx.doi.org/10.2307/817410

Richards, J., Platt, J., Weber, H., & Inman, P. (1986). *Longman dictionary of applied linguistics*. Longman: UK.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(04), 589-619.

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research*, *6*(3), 183-222. http://dx.doi.org/10.1191/1362168802lr108oa

Rudy, M. (2013). *EFL Writing strategies of the second year students of SMPIT Daarul 'Ilmi Kemiling Bandar Lampung*. Paper presented at International Conference on Education and Language (ICEL), Bandar Lampung University.

Sagarra, N., & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The Modern Language Journal*, *90*(2), 228-243. http://dx.doi.org/10.1111/j.1540-4781.2006.00394.x

Sarani, A., Negari, G. M., & Ghaviniat, M. (2013). The role of task type in L2 vocabulary acquisition: a case of involvement load hypothesis. *Acta Scientiarum Language and Culture*, *35*(4), 377-386. http://dx.doi.org/10.4025/actascilangcult.v35i4.21135

Sarbazi, M. R. (2014). Involvement load hypothesis: Recalling unfamiliar words meaning by adults across genders. *Procedia-Social and Behavioral Sciences*, *98*, 1686-1692. http://dx.doi.org/10.1016/j.sbspro.2014.03.594

Sawaki, Y., Quinlan, T., & Lee, Y. W. (2013). Understanding learner strengths and weaknesses: assessing performance on an integrated writing task. *Language Assessment Quarterly*, *10*(1), 73-95. http://dx.doi.org/10.1080/15434303.2011.633305

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

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

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

Shimizu, S, (2008). v8.an -revised web edition [On-line concordance program] http://www.tcp-ip.or.jp/~shim/J8LevelMarker/j8lm.cgi

Sidek, H. M., & Rahim, H. A. (2015). The role of vocabulary knowledge in reading comprehension: A cross-linguistic study. *Procedia-Social and Behavioral Sciences*, *197*, 50-56. http://dx.doi.org/10.1016/j.sbspro.2015.07.046

Silverman, R. D., Proctor, C. P., Harring, J. R., Hartranft, A. M., Doyle, B., & Zelinke, S. B. (2015). Language skills and reading comprehension in English monolingual and Spanish–English bilingual children in grades 2–5. *Reading and Writing*, *28*(9), 1381-1405. http://dx.doi.org/10.1007/s11145-015-9575

Sofian, H., & Salam, U. (2015). Improving students' writing ability in narrative text by using picture series in SMA. *Jurnal Pendidikan dan Pembelajaran*, *4*(4), 1-12.

Solak, E., & Altay, F. (2014). Prospective EFL teachers' perceptions of listening comprehension problems in Turkey. *Journal of International Social Research*, *7*(30).

Soleimani, H., Rahmanian, M., & Sajedi, K. (2015). A revisit to Vocabulary Acquisition in involvement load hypothesis. *Procedia-Social and Behavioral Sciences*, *192*, 388-397. http://dx.doi.org/10.1016/j.sbspro.2015.06.055

Sonbul, S., & Schmitt, N. (2009). Direct teaching of vocabulary after reading: is it worth the effort? *ELT Journal*, *64*(3), 253-360. http://dx.doi.org/10.1093/elt/ccp059

Strasser, K., & Río, F. D. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly*, *49*(2), 169-187. http://dx.doi.org/10.1002/rrq.68

Tahir, S. Z. (2015). Improving students' speaking skill through voice chat at University of Iqra Buru. *Journal of Modern Education Review*, *5*(3), 296-306. http://dx.doi.org/10.15341/jmer(2155-7993)/03.05.2015/009

Tighe, E. L., & Schatschneider, C. (2014). A quantile regression approach to understanding the relations among morphological awareness, vocabulary, and reading comprehension in adult basic education students. *Journal of Learning Disabilities*, *2014,* 1-13. http://dx.doi.org/10.1177/0022219414556771

Tu, H. F., & Su, I. R. (2004). *Effects of task-induced involvement on incidental vocabulary learning in a second language*. (Unpublished master's thesis). National Tsing Hua University, Taiwan.

Van Polen, Z. (2014). *The effects of task-induced involvement load and word exposure frequency on L2 incidental vocabulary learning through reading* (Doctoral dissertation). Universiteit Van Amsterdam, Netherlands.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*(2), 390-416. http://dx.doi.org/10.1111/lang.12105

Wahyuni, W. (2014). The influence of English day towards students' speaking skill at the eighth graders. *ELTS*, *2*(4), 1-10.

Walsh, M. I. (2009). *The involvement load hypothesis applied to high school learners in Japan: Measuring the effects of evaluation* (Unpublished master's thesis). Birmingham University, Edgbaston, United Kingdom.

Wang, C., Xu, K., & Zuo, Y. (2014). The Effect of evaluation factor on the incidental vocabulary acquisition through reading. *International Journal of English Linguistics*, *4*(3), 59-66. http://dx.doi.org/10.5539/ijel.v4n3p59

Wang, S. (2015). An empirical study on the role of vocabulary knowledge in EFL listening comprehension. *Theory and Practice in Language Studies*, *5*(5), 989-995. http://dx.doi.org/10.17507/tpls.0505.14

Watanabe, Y. (1997). Input, intake, and retention. *Studies in Second Language Acquisition*, *19*(03), 287-307.

Webb, S. (2013). Depth of vocabulary knowledge. In C. A. Chapelle (Eds.), *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing Ltd.

Wesche, M. B., & Paribakht, T. S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *Modern Language Journal*, *84*(2), 196-213. Retrieved from http://www.jstor.org/stable/330485

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, *11*(2), 87-95. http://dx.doi.org/10.1080/00461527409529129

Xu, X. (2010). An empirical study on the effect of task on L2 incidental vocabulary acquisition through reading. *Asian Social Science*, *6*(7), 126-131. http://dx.doi.org/10.5539/ass.v6n7p126

Yang, Y. I. (2015). An investigation of Chinese junior high school teachers' and students' attitudes towards EFL writing. *International Journal of Research Studies in Education*, *5*(2). http://dx.doi.org/10.5861/ijrse.2015.1209

Yaqubi, B., Rayati, R. A., & Gorgi, N. (2012). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *Journal of Teaching Language Skills*, *2*(1), 145-163.

Zhang, J., McBride-Chang, C., Wong, A. M. Y., Tardif, T., Shu, H., & Zhang, Y. (2014). Longitudinal correlates of reading comprehension difficulties in Chinese children. *Reading and Writing*, *27*(3), 481-501. http://dx.doi.org/10.1007/s11145-013-9453-4

APPENDICES

Appendix I. The First Text

The first of our three brains to evolve is what scientists call the reptilian cortex. This brain sustains the elementary activities of animal survival such as respiration, adequate rest and a beating heart. We are not required to consciously "think" about these activities. The reptilian cortex also houses the "startle centre", a mechanism that facilitates swift reactions to unexpected occurrences in our surroundings. That panicked lurch you experience when a door slams shut somewhere in the house, or the heightened awareness you feel when a twig cracks in a nearby bush while out on an evening walk are both examples of the reptilian cortex at work. When it comes to our interaction with others, the reptilian brain offers up only the most basic impulses: aggression, mating, and territorial defence. There is no great difference, in this sense, between a crocodile defending its spot along the river and a turf war between two urban gangs.

Although the lizard may stake a claim to its habitat, it exerts total indifference toward the well-being of its young. Listen to the anguished squeal of a dolphin separated from its pod or witness the sight of elephants mourning their dead, however, and it is clear that a new development is at play. Scientists have identified this as the limbic cortex. Unique to mammals, the limbic cortex impels creatures to nurture their offspring by delivering feelings of tenderness and warmth to the parent when children are nearby. These same sensations also cause mammals to develop various types of social relations and kinship networks. When we are with others of "our kind" – be it at soccer practice, church, school or a nightclub – we experience positive sensations of togetherness, solidarity and comfort. If we spend too long away from these networks, then loneliness sets in and encourages us to seek companionship.

Only human capabilities extend far beyond the scope of these two cortexes. Humans eat, sleep and play, but we also speak, plot, rationalise and debate finer points of morality. Our unique abilities are the result of an expansive third brain – the neocortex which engages with logic, reason and ideas. The power of the neocortex comes from its ability to think beyond the present, concrete moment. While other mammals are mainly restricted to impulsive actions, humans can think about the "big picture". We can string together simple lessons (for example, an apple drops

downwards from a tree; hurting others causes unhappiness) to develop complex theories of physical or social phenomena (such as the laws of gravity and a concern for human rights). The neocortex is also responsible for the process by which we decide on and commit to particular courses of action. Strung together over time, these choices can accumulate into feats of progress unknown to other animals. Anticipating a better grade on the following morning's exam, a student can ignore the limbic urge to socialise and go to sleep early instead. Over three years, this ongoing sacrifice translates into a first class degree and a scholarship to graduate school; over a lifetime, it can mean ground-breaking contributions to human knowledge and development. The ability to sacrifice our drive for immediate satisfaction in order to benefit later is a product of the neocortex.

Understanding the triune brain can help us appreciate the different natures of brain damage and psychological disorders. The most devastating form of brain damage, for example, is a condition in which someone is understood to be brain dead. In this state a person appears merely unconscious – sleeping, perhaps – but this is illusory. Here, the reptilian brain is functioning on autopilot despite the permanent loss of other cortexes.

Disturbances to the limbic cortex are registered in a different manner. Pups with limbic damage can move around and feed themselves well enough but do not register the presence of their littermates. Scientists have observed how, after a limbic lobotomy, "one impaired monkey stepped on his outraged peers as if treading on a log or a rock". In our own species, limbic damage is closely related to sociopathic behaviour. Sociopaths in possession of fully-functioning neocortexes are often shrewd and emotionally intelligent people but lack any ability to relate to, empathise with or express concern for others. One of the neurological wonders of history occurred when a railway worker named Phineas Gage survived an incident during which a metal rod skewered his skull, taking a considerable amount of his neocortex with it. Though Gage continued to live and work as before, his fellow employees observed a shift in the equilibrium of his personality. Gage's animal propensities were now sharply pronounced while his intellectual abilities suffered; garrulous or obscene jokes replaced his once quick wit. New findings suggest, however, that Gage managed to soften these abrupt changes over time and rediscover an appropriate social manner. This would indicate that reparative

therapy has the potential to help patients with advanced brain trauma to gain an improved quality of life.

<u>Part 1</u>: Write the correct letter, A, B or C in boxes by classifying the following as typical of:

A: the reptilian cortex

B: the limbic cortex

C: the neocortex

1. giving up short-term happiness for future gains	
2. maintaining the bodily functions necessary for life	
3. experiencing the pain of losing another	
4. forming communities and social groups	
5. making a decision and carrying it out	
6. guarding areas of land	
7. developing explanations for things	
8. looking after one's young	
9. responding quickly to sudden movement and noise	

<u>Part 2</u>: Complete the sentences below.

Write NO MORE THAN TWO WORDS from the passage for each answer.

10. A person with only a functioning reptilian cortex is known as _____.

11. _____in humans is associated with limbic disruption.

12. An industrial accident caused Phineas Gage to lose part of his _____

13. After his accident, co-workers noticed an imbalance between Gage'_____ and higher-order thinking.

Appendix II. The Second Text

Dr Rajendra Persaud explains how practical intelligence is linked to success.

This year, record numbers of high school students obtained top grades in their final exams, yet employers complain that young people still lack the basic skills to succeed at work. The only explanation offered is that exams must be getting easier. But the real answer could lie in a study just published by Professor Robert Sternberg, an eminent psychologist at Yale University in the USA and the world's leading expert on intelligence. His research reveals the existence of a totally new variety: practical intelligence.

Professor Sternberg's astonishing finding is that practical intelligence, which predicts success in real life, has an inverse relationship with academic intelligence. In other words, the more practically intelligent you are, the less likely you are to succeed at school or university. Similarly, the more paper qualifications you hold and the higher your grades, the less able you are to cope with problems of everyday life and the lower your score in practical intelligence.

Many people who are clearly successful in their place of work do badly in standard IQ (academic intelligence) tests. Entrepreneurs and those who have built large businesses from scratch are frequently discovered to be high school or college drop-outs. IQ as a concept is more than 100 years old. It was supposed to explain why some people excelled at a wide variety of intellectual tasks. But IQ ran into trouble when it became apparent that some high scorers failed to achieve in real life what was predicted by their tests.

Emotional intelligence (EQ), which emerged a decade ago, was supposed to explain this deficit. It suggested that to succeed in real life, people needed both emotional as well as intellectual skills. EQ includes the abilities to motivate yourself and persist in the face of frustrations; to control impulses and delay gratification; to regulate moods and keep distress from swamping the ability to think; and to understand and empathize with others. While social or emotional intelligence was a useful concept in explaining many of the real-world deficiencies of super intelligent people, it did not go any further than the IQ test in measuring success in real life. Again, some of the most successful people in the business world were obviously lacking in social charm.

Not all the real-life difficulties we face are solvable-with just good social skills - and good social acumen in one situation may not translate to another. The crucial problem with academic and emotional intelligence scores is that they are both poor predictors of success in real life. For example, research has shown that IQ tests predict only between 4% and 25% of success in life, such as job performance.

Professor Sternberg's group at Yale began from a very different position to traditional researchers into intelligence. Instead of asking what intelligence was and investigating whether it predicted success in life, Professor Sternberg asked what distinguished people who were thriving from those that were not. Instead of measuring this form of intelligence with mathematical or verbal tests, practical intelligence is scored by answers to real-life dilemmas such as: 'If you were travelling by car and got stranded on a motorway during a blizzard, what would you do?' An important contrast between these questions is that in academic tests there is usually only one answer, whereas in practical intelligence tests - as in real life - there are several different solutions to the problem.

The Yale group found that most of the really useful knowledge which successful people have acquired is gained during everyday activities - but typically without conscious awareness. Although successful people's behaviour reflects the fact that they have this knowledge, high achievers are often unable to articulate or define what they know. This partly explains why practical intelligence has been so difficult to identify.

Professor Sternberg found that the best way to reach practical intelligence is to ask successful people to relate examples of crucial incidents at work where they solved problems demonstrating skills they had learnt while doing their jobs. It would appear that one of the best ways of improving your practical intelligence is to observe master practitioners at work and, in particular, to focus on the skills they have acquired while doing the job. Oddly enough, this is the basis of traditional apprentice training. Historically, the junior doctor learnt by observing the consultant surgeon at work and the junior lawyer by assisting the senior barrister.

Another area where practical intelligence appears to resolve a previously unexplained paradox is that performance in academic tests usually declines after formal education ends. Yet, most older adults contend that their ability to solve practical

problems increases over the years. The key implication for organizations and companies is that practical intelligence may not be detectable by conventional audits and performance measuring procedures. Training new or less capable employees to become more practically intelligent will involve learning from the genuinely practically intelligent rather than from training manuals or courses.

Perhaps the biggest challenge is in recruitment, as these new studies strongly suggest that paper qualifications are unlikely to be helpful in predicting who will be best at solving your company's problems. Professor Sternberg's research suggests that we should start looking at companies in a completely different way - and see them as places where a huge number of problems are being solved all the time but where it may take new eyes to see the practical intelligence in action.

<u>Part 1</u>: Choose the correct answer, A, B, C or D

1) Professor Sternberg's study showed that

A. qualifications are a good indicator of success at work.
B. education can help people cope with real-life problems.
C. intelligent people do not always achieve well at school.
D. high grades can indicate a lack of practical intelligence

2) What is the 'deficit' referred to in the fourth paragraph?

A. People with high IQ scores could not score well in EQ tests.
B. EQ tests were unable to predict success at work.
C. High IQ scores did not always lead to personal success.
D. People with high EQ scores could not cope with real life.

3) Professor Sternberg's research differed from previous studies because

A. he used verbal testing instead of mathematics.
B. he began by establishing a definition of intelligence.
C. he analysed whether intelligence could predict success in real life.
D. he wanted to find out what was different about successful people.

4) Part of the reason why practical intelligence had not been identified before Professor Sternberg's study is that

A. the behaviour of successful people had never been studied.

B. successful people are too busy with their everyday lives.

C. successful people cannot put their knowledge into words.

D. successful people are unaware of their own abilities.

5) In order to increase the practical intelligence of employees, companies need to

A. adopt an apprentice-style system.

B. organise special courses.

C. devise better training manuals.

D. carry out an audit on all employees

<u>Part 2</u>: Write A, B or C in the boxes by classifying the following characteristics as belonging to:

A: academic intelligence (IQ) tests

B: emotional intelligence (EQ) tests

C: practical intelligence tests

6) measures skills which are likely to improve with age	
7) assesses people's social skills	
8) measures the ability to deal with real-life difficulties	
9) the oldest of the three tests	
10) high scorers learn from their actions	
11) high scorers are more likely to stay calm in difficult	
12) questions have more than one possible answer	

Appendix III. Fill-in by Searching for the First Text

Number:

Please read the following passage. Fill in the blanks with the most appropriate words from the list. Based on the text, answer the questions that follow.

The Triune Brain

The first of our three brains to evolve is what scientists call the reptilian cortex. This brain sustains the elementary activities of animal survival such as 1) _____, adequate rest and a beating heart. We are not required to consciously "think" about these activities. The reptilian cortex also houses the "startle centre", a mechanism that facilitates swift reactions to unexpected occurrences in our surroundings. That panicked lurch you experience when a door slams shut somewhere in the house, or the heightened awareness you feel when a twig cracks in a nearby bush while out on an evening walk are both examples of the reptilian cortex at work. When it comes to our interaction with others, the reptilian brain offers up only the most basic impulses: 2) _____, mating, and territorial defence. There is no great difference, in this sense, between a crocodile defending its spot along the river and a turf war between two urban gangs.

Although the lizard may fight for its habitat, it exerts total 3) _____ toward the well-being of its young. Listen to the anguished squeal of a dolphin separated from its pod or witness the sight of elephants mourning their dead, however, and it is clear that a new development is at play. Scientists have identified this as the limbic cortex. Unique to mammals, the limbic cortex impels creatures to care for their 4) _____ by delivering feelings of tenderness and warmth to the parent when children are nearby. These same sensations also cause mammals to develop various types of social relations and 5) _____ networks. When we are with others of "our kind" – be it at soccer practice, church, school or a nightclub – we experience positive sensations of togetherness, 6) _____ and comfort. If we spend too long away from these networks, then loneliness sets in and encourages us to seek companionship.

Only human capabilities extend far beyond the scope of these two cortexes. Humans eat, sleep and play, but we also speak, plot, rationalise and debate finer points of 7) _____. Our unique abilities are the result of an expansive third brain – the

neocortex which engages with logic, reason and ideas. The power of the neocortex comes from its ability to think beyond the present, concrete moment. While other mammals are mainly restricted to impulsive actions, humans can think about the "big picture". We can string together simple lessons (for example, an apple drops downwards from a tree; hurting others causes unhappiness) to develop complex theories of physical or social phenomena (such as the laws of gravity and a concern for human rights). The neocortex is also responsible for the process by which we decide on and commit to particular courses of action. Strung together over time, these choices can accumulate into feats of progress unknown to other animals. Anticipating a better grade on the following morning's exam, a student can ignore the limbic urge to socialise and go to sleep early instead. Over three years, this ongoing sacrifice translates into a first class degree and a scholarship to graduate school; over a lifetime, it can mean ground-breaking contributions to human knowledge and development. The ability to sacrifice our drive for immediate satisfaction in order to benefit later is a product of the neocortex.

Understanding the triune brain can help us appreciate the different natures of brain damage and psychological disorders. The most devastating form of brain damage, for example, is a condition in which someone is understood to be brain dead. In this state a person appears merely unconscious – sleeping, perhaps – but this is illusory. Here, the reptilian brain is functioning on autopilot despite the permanent loss of other cortexes.

Disturbances to the limbic cortex are registered in a different manner. Pups with limbic damage can move around and feed themselves well enough but do not register the presence of their littermates. Scientists have observed how, after a limbic lobotomy, "one impaired monkey stepped on his outraged peers as if treading on a log or a rock". In our own species, limbic damage is closely related to sociopathic behaviour. Sociopaths in possession of fully-functioning neocortexes are often shrewd and emotionally intelligent people but lack any ability to relate to, empathise with or express concern for others. One of the neurological wonders of history occurred when a railway worker named Phineas Gage survived an incident during which a metal rod skewered his skull, taking a considerable amount of his neocortex with it. Though Gage continued to live and work as before, his fellow employees observed a shift in the 8) _____ of

his personality. Gage's animal 9) ________ were now sharply pronounced while his intellectual abilities suffered; garrulous or obscene jokes replaced his once quick 10) _____. New findings suggest, however, that Gage managed to soften these abrupt changes over time and rediscover an appropriate social manner. This would indicate that reparative therapy has the potential to help patients with advanced brain trauma to gain an improved quality of life.

Word List

propensity	solidarity	kinship	indifference
respiration	morality	wit	equilibrium
offspring		aggression	pudency

Appendix IV. Fill-in by Searching for the Second Text

Number:

Please read the following passage. Fill in the blanks with the most appropriate words from the list. Based on the text, answer the questions that follow.

Practical intelligence lends a hand

Dr Rajendra Persaud explains how practical intelligence is linked to success.

This year, record numbers of high school students obtained top grades in their final exams, yet employers complain that young people still lack the basic skills to succeed at work. The only explanation offered is that exams must be getting easier. But the real answer could lie in a study just published by Professor Robert Sternberg, an eminent psychologist at Yale University in the USA and the world's leading expert on intelligence. His research reveals the existence of a totally new variety: practical intelligence.

Professor Sternberg's astonishing finding is that practical intelligence, which predicts success in real life, has an inverse relationship with academic intelligence. In other words, the more practically intelligent you are, the less likely you are to succeed at school or university. Similarly, the more paper qualifications you hold and the higher your grades, the less able you are to cope with problems of everyday life and the lower your score in practical intelligence.

Many people who are clearly successful in their place of work do badly in standard IQ (academic intelligence) tests. 1) _____ and those who have built large businesses from scratch are frequently discovered to be high school or college drop-outs. IQ as a concept is more than 100 years old. It was supposed to explain why some people excelled at a wide variety of intellectual tasks. But IQ ran into trouble when it became apparent that some high scorers failed to achieve in real life what was predicted by their tests.

Emotional intelligence (EQ), which emerged a decade ago, was supposed to explain this deficit. It suggested that to succeed in real life, people needed both emotional as well as intellectual skills. EQ includes the abilities to motivate yourself and persist in the face of 2) _____ to control impulses and delay 3) _____; to

regulate moods and keep 4) _____ from swamping the ability to think; and to understand and empathize with others. While social or emotional intelligence was a useful concept in explaining many of the real-world 5) _____ of super intelligent people, it did not go any further than the 10 test in measuring success in real life. Again, some of the most successful people in the business world were obviously lacking in social charm.

Not all the real-life difficulties we face are solvable-with just good social skills - and good social 6) _____ in one situation may not translate to another. The crucial problem with academic and emotional intelligence scores is that they are both poor predictors of success in real life. For example, research has shown that IQ tests predict only between 4% and 25% of success in life, such as job performance.

Professor Sternberg's group at Yale began from a very different position to traditional researchers into intelligence. Instead of asking what intelligence was and investigating whether it predicted success in life, Professor Sternberg asked what distinguished people who were thriving from those that were not. Instead of measuring this form of intelligence with mathematical or verbal tests, practical intelligence is scored by answers to real-life 7) _____ such as: “If you were travelling by car and got stranded on a motorway during a blizzard, what would you do?” An important contrast between these questions is that in academic tests there is usually only one answer, whereas in practical intelligence tests - as in real life - there are several different solutions to the problem.

The Yale group found that most of the really useful knowledge which successful people have acquired is gained during everyday activities - but typically without conscious awareness. Although successful people's behaviour reflects the fact that they have this knowledge, high achievers are often unable to articulate or define what they know. This partly explains why practical intelligence has been so difficult to identify.

Professor Sternberg found that the best way to reach practical intelligence is to ask successful people to relate examples of crucial incidents at work where they solved problems demonstrating skills they had learnt while doing their jobs. It would appear that one of the best ways of improving your practical intelligence is to observe master practitioners at work and, in particular, to focus on the skills they have acquired while

doing the job. Oddly enough, this is the basis of traditional 8) _____ training. Historically, the junior doctor learnt by observing the consultant surgeon at work and the junior lawyer by assisting the senior barrister.

Another area where practical intelligence appears to resolve a previously unexplained paradox is that performance in academic tests usually declines after formal education ends. Yet most older adults contend that their ability to solve practical problems increases over the years. The key implication for organizations and companies is that practical intelligence may not be detectable by conventional 9) _____ and performance measuring procedures. Training new or less capable employees to become more practically intelligent will involve learning from the genuinely practically intelligent rather than from training manuals or courses.

Perhaps the biggest challenge is in 10) _____, as these new studies strongly suggest that paper qualifications are unlikely to be helpful in predicting who will be best at solving your company's problems. Professor Sternberg's research suggests that we should start looking at companies in a completely different way - and see them as places where a huge number of problems are being solved all the time but where it may take new eyes to see the practical intelligence in action.

Word List

recruitment	entrepreneur	audit	gratification
acumen	apprentice	dilemma	caveat
frustration	distress		deficiency

Appendix V. Fill-in with Glossary for the First Text

Number:

Please read the following passage. Fill in the blanks with the most appropriate words from the list. Based on the text, answer the questions that follow.

The Triune Brain

The first of our three brains to evolve is what scientists call the reptilian cortex. This brain sustains the elementary activities of animal survival such as 1) _____, adequate rest and a beating heart. We are not required to consciously "think" about these activities. The reptilian cortex also houses the "startle centre", a mechanism that facilitates swift reactions to unexpected occurrences in our surroundings. That panicked lurch you experience when a door slams shut somewhere in the house, or the heightened awareness you feel when a twig cracks in a nearby bush while out on an evening walk are both examples of the reptilian cortex at work. When it comes to our interaction with others, the reptilian brain offers up only the most basic impulses: 2) _____, mating, and territorial defence. There is no great difference, in this sense, between a crocodile defending its spot along the river and a turf war between two urban gangs.

Although the lizard may fight for its habitat, it exerts total 3) _____ toward the well-being of its young. Listen to the anguished squeal of a dolphin separated from its pod or witness the sight of elephants mourning their dead, however, and it is clear that a new development is at play. Scientists have identified this as the limbic cortex. Unique to mammals, the limbic cortex impels creatures to care for their 4) _____ by delivering feelings of tenderness and warmth to the parent when children are nearby. These same sensations also cause mammals to develop various types of social relations and 5) _____ networks. When we are with others of "our kind" – be it at soccer practice, church, school or a nightclub – we experience positive sensations of togetherness, 6) _____ and comfort. If we spend too long away from these networks, then loneliness sets in and encourages us to seek companionship.

Only human capabilities extend far beyond the scope of these two cortexes. Humans eat, sleep and play, but we also speak, plot, rationalise and debate finer points of 7) _____. Our unique abilities are the result of an expansive third brain – the

neocortex which engages with logic, reason and ideas. The power of the neocortex comes from its ability to think beyond the present, concrete moment. While other mammals are mainly restricted to impulsive actions, humans can think about the "big picture". We can string together simple lessons (for example, an apple drops downwards from a tree; hurting others causes unhappiness) to develop complex theories of physical or social phenomena (such as the laws of gravity and a concern for human rights). The neocortex is also responsible for the process by which we decide on and commit to particular courses of action. Strung together over time, these choices can accumulate into feats of progress unknown to other animals. Anticipating a better grade on the following morning's exam, a student can ignore the limbic urge to socialise and go to sleep early instead. Over three years, this ongoing sacrifice translates into a first class degree and a scholarship to graduate school; over a lifetime, it can mean ground-breaking contributions to human knowledge and development. The ability to sacrifice our drive for immediate satisfaction in order to benefit later is a product of the neocortex.

Understanding the triune brain can help us appreciate the different natures of brain damage and psychological disorders. The most devastating form of brain damage, for example, is a condition in which someone is understood to be brain dead. In this state a person appears merely unconscious – sleeping, perhaps – but this is illusory. Here, the reptilian brain is functioning on autopilot despite the permanent loss of other cortexes.

Disturbances to the limbic cortex are registered in a different manner. Pups with limbic damage can move around and feed themselves well enough but do not register the presence of their littermates. Scientists have observed how, after a limbic lobotomy, "one impaired monkey stepped on his outraged peers as if treading on a log or a rock". In our own species, limbic damage is closely related to sociopathic behaviour. Sociopaths in possession of fully-functioning neocortexes are often shrewd and emotionally intelligent people but lack any ability to relate to, empathise with or express concern for others. One of the neurological wonders of history occurred when a railway worker named Phineas Gage survived an incident during which a metal rod skewered his skull, taking a considerable amount of his neocortex with it. Though Gage continued to live and work as before, his fellow employees observed a shift in the 8) _____ of

his personality. Gage's animal 9) _____ were now sharply pronounced while his intellectual abilities suffered; garrulous or obscene jokes replaced his once quick 10) _____. New findings suggest, however, that Gage managed to soften these abrupt changes over time and rediscover an appropriate social manner. This would indicate that reparative therapy has the potential to help patients with advanced brain trauma to gain an improved quality of life.

Word List

aggression: noun
(saldırganlık) feelings of anger or antipathy resulting in hostile or violent behaviour; readiness to attack or confront

equilibrium: noun
(denge) a state in which opposing forces or influences are balanced

indifference: noun
(ilgisizlik) lack of interest, concern, or sympathy

kinship: noun
(akrabalık) blood relationship

morality: noun
(ahlak) principles concerning the distinction between right and wrong or good and bad behaviour

offspring: noun
(yavru) an animal's young

propensity: noun
(eğilim) an inclination or natural tendency to behave in a particular way

pudency: noun
(mahcupluk) embarrassment

respiration: noun
(solunum) the action of breathing

solidarity: noun
(birlik, beraberlik) unity or agreement of feeling or action, especially among individuals with a common interest; mutual support within a group

wit: noun
(ince espri) a natural aptitude for using words and ideas in a quick and inventive way to create humour

Appendix VI. Fill-in with Glossary for the Second Text

Number:

Please read the following passage. Fill in the blanks with the most appropriate words from the list. Based on the text, answer the questions that follow.

Practical intelligence lends a hand

Dr Rajendra Persaud explains how practical intelligence is linked to success.

This year, record numbers of high school students obtained top grades in their final exams, yet employers complain that young people still lack the basic skills to succeed at work. The only explanation offered is that exams must be getting easier. But the real answer could lie in a study just published by Professor Robert Sternberg, an eminent psychologist at Yale University in the USA and the world's leading expert on intelligence. His research reveals the existence of a totally new variety: practical intelligence.

Professor Sternberg's astonishing finding is that practical intelligence, which predicts success in real life, has an inverse relationship with academic intelligence. In other words, the more practically intelligent you are, the less likely you are to succeed at school or university. Similarly, the more paper qualifications you hold and the higher your grades, the less able you are to cope with problems of everyday life and the lower your score in practical intelligence.

Many people who are clearly successful in their place of work do badly in standard IQ (academic intelligence) tests. 1) _____ and those who have built large businesses from scratch are frequently discovered to be high school or college drop-outs. IQ as a concept is more than 100 years old. It was supposed to explain why some people excelled at a wide variety of intellectual tasks. But IQ ran into trouble when it became apparent that some high scorers failed to achieve in real life what was predicted by their tests.

Emotional intelligence (EQ), which emerged a decade ago, was supposed to explain this deficit. It suggested that to succeed in real life, people needed both emotional as well as intellectual skills. EQ includes the abilities to motivate yourself and persist in the face of 2) _____ to control impulses and delay 3) _____; to

regulate moods and keep 4) _____ from swamping the ability to think; and to understand and empathize with others. While social or emotional intelligence was a useful concept in explaining many of the real-world 5) _____ of super intelligent people, it did not go any further than the 10 test in measuring success in real life. Again, some of the most successful people in the business world were obviously lacking in social charm.

Not all the real-life difficulties we face are solvable-with just good social skills - and good social 6) _____ in one situation may not translate to another. The crucial problem with academic and emotional intelligence scores is that they are both poor predictors of success in real life. For example, research has shown that IQ tests predict only between 4% and 25% of success in life, such as job performance.

Professor Sternberg's group at Yale began from a very different position to traditional researchers into intelligence. Instead of asking what intelligence was and investigating whether it predicted success in life, Professor Sternberg asked what distinguished people who were thriving from those that were not. Instead of measuring this form of intelligence with mathematical or verbal tests, practical intelligence is scored by answers to real-life 7) _____ such as: “If you were travelling by car and got stranded on a motorway during a blizzard, what would you do?” An important contrast between these questions is that in academic tests there is usually only one answer, whereas in practical intelligence tests - as in real life - there are several different solutions to the problem.

The Yale group found that most of the really useful knowledge which successful people have acquired is gained during everyday activities - but typically without conscious awareness. Although successful people's behaviour reflects the fact that they have this knowledge, high achievers are often unable to articulate or define what they know. This partly explains why practical intelligence has been so difficult to identify.

Professor Sternberg found that the best way to reach practical intelligence is to ask successful people to relate examples of crucial incidents at work where they solved problems demonstrating skills they had learnt while doing their jobs. It would appear that one of the best ways of improving your practical intelligence is to observe master practitioners at work and, in particular, to focus on the skills they have acquired while

doing the job. Oddly enough, this is the basis of traditional 8) _____ training. Historically, the junior doctor learnt by observing the consultant surgeon at work and the junior lawyer by assisting the senior barrister.

Another area where practical intelligence appears to resolve a previously unexplained paradox is that performance in academic tests usually declines after formal education ends. Yet most older adults contend that their ability to solve practical problems increases over the years. The key implication for organizations and companies is that practical intelligence may not be detectable by conventional 9) _____ and performance measuring procedures. Training new or less capable employees to become more practically intelligent will involve learning from the genuinely practically intelligent rather than from training manuals or courses.

Perhaps the biggest challenge is in 10) _____, as these new studies strongly suggest that paper qualifications are unlikely to be helpful in predicting who will be best at solving your company's problems. Professor Sternberg's research suggests that we should start looking at companies in a completely different way - and see them as places where a huge number of problems are being solved all the time but where it may take new eyes to see the practical intelligence in action.

Word List

acumen:noun
(keskin zeka) the ability to make good judgements and take quick decisions

apprentice: noun
(çırak) a person who is learning a trade from a skilled employer, having agreed to work for a fixed period at low wages

audit: noun
(denetim) an official inspection of an organization's accounts, typically by an independent body

caveat: noun
(uyarı) an explanation or warning that should be remembered when you are doing or thinking about something

deficiency: noun
(eksiklik) a lack or shortage

dilemma: noun
(ikilem) a situation in which a difficult choice has to be made between two or more alternatives, especially ones that are equally undesirable

distress: noun
(ızdırap, dert) extreme anxiety, sorrow, or pain

entrepreneur: noun
(girişimci) a person who sets up a business or businesses, taking on financial risks in the hope of profit

frustration: noun
(hüsran) the feeling of being upset or annoyed as a result of being unable to change or achieve something

gratification: noun
(haz) pleasure, especially when gained from the satisfaction of a desire

recruitment: noun
(işe alım) the action of finding new people to join an organization or support a cause

Appendix VII. Retelling by Searching for the First Text

Number:

Please read the following passage. Based on the text, answer the questions that follow.

The Triune Brain

The first of our three brains to evolve is what scientists call the reptilian cortex. This brain sustains the elementary activities of animal survival such as **respiration**, adequate rest and a beating heart. We are not required to consciously "think" about these activities. The reptilian cortex also houses the "startle centre", a mechanism that facilitates swift reactions to unexpected occurrences in our surroundings. That panicked lurch you experience when a door slams shut somewhere in the house, or the heightened awareness you feel when a twig cracks in a nearby bush while out on an evening walk are both examples of the reptilian cortex at work. When it comes to our interaction with others, the reptilian brain offers up only the most basic impulses: **aggression**, mating, and territorial defence. There is no great difference, in this sense, between a crocodile defending its spot along the river and a turf war between two urban gangs.

Although the lizard may fight for its habitat, it exerts total **indifference** toward the well-being of its young. Listen to the anguished squeal of a dolphin separated from its pod or witness the sight of elephants mourning their dead, however, and it is clear that a new development is at play. Scientists have identified this as the limbic cortex. Unique to mammals, the limbic cortex impels creatures to care for their **offspring** by delivering feelings of tenderness and warmth to the parent when children are nearby. These same sensations also cause mammals to develop various types of social relations and **kinship** networks. When we are with others of "our kind" – be it at soccer practice, church, school or a nightclub – we experience positive sensations of togetherness, **solidarity** and comfort. If we spend too long away from these networks, then loneliness sets in and encourages us to seek companionship.

Only human capabilities extend far beyond the scope of these two cortexes. Humans eat, sleep and play, but we also speak, plot, rationalise and debate finer points of **morality**. Our unique abilities are the result of an expansive third brain – the neocortex which engages with logic, reason and ideas. The power of the neocortex

comes from its ability to think beyond the present, concrete moment. While other mammals are mainly restricted to impulsive actions, humans can think about the "big picture". We can string together simple lessons (for example, an apple drops downwards from a tree; hurting others causes unhappiness) to develop complex theories of physical or social phenomena (such as the laws of gravity and a concern for human rights). The neocortex is also responsible for the process by which we decide on and commit to particular courses of action. Strung together over time, these choices can accumulate into feats of progress unknown to other animals. Anticipating a better grade on the following morning's exam, a student can ignore the limbic urge to socialise and go to sleep early instead. Over three years, this ongoing sacrifice translates into a first class degree and a scholarship to graduate school; over a lifetime, it can mean ground-breaking contributions to human knowledge and development. The ability to sacrifice our drive for immediate satisfaction in order to benefit later is a product of the neocortex.

Understanding the triune brain can help us appreciate the different natures of brain damage and psychological disorders. The most devastating form of brain damage, for example, is a condition in which someone is understood to be brain dead. In this state a person appears merely unconscious – sleeping, perhaps – but this is illusory. Here, the reptilian brain is functioning on autopilot despite the permanent loss of other cortexes.

Disturbances to the limbic cortex are registered in a different manner. Pups with limbic damage can move around and feed themselves well enough but do not register the presence of their littermates. Scientists have observed how, after a limbic lobotomy, "one impaired monkey stepped on his outraged peers as if treading on a log or a rock". In our own species, limbic damage is closely related to sociopathic behaviour. Sociopaths in possession of fully-functioning neocortexes are often shrewd and emotionally intelligent people but lack any ability to relate to, empathise with or express concern for others. One of the neurological wonders of history occurred when a railway worker named Phineas Gage survived an incident during which a metal rod skewered his skull, taking a considerable amount of his neocortex with it. Though Gage continued to live and work as before, his fellow employees observed a shift in the **equilibrium** of his personality. Gage's animal **propensities** were now sharply pronounced while his

intellectual abilities suffered; garrulous or obscene jokes replaced his once quick **wit**. New findings suggest, however, that Gage managed to soften these abrupt changes over time and rediscover an appropriate social manner. This would indicate that reparative therapy has the potential to help patients with advanced brain trauma to gain an improved quality of life.

Using the words written in bold in the text, rewrite the text in your own words:

a) **explain the differences between the reptilian cortex, the limbic cortex and the neocortex;**
b) **provide examples for different functions of these three cortexes;**
c) **explain what problems may occur in case of any damage to these three cortexes.**

(You can use a dictionary to look up the meanings of the words)

Appendix VIII. Retelling by Searching for the Second Text

Number:

Please read the following passage. Based on the text, answer the questions that follow.

Practical Intelligence Lends a Hand

Dr Rajendra Persaud explains how practical intelligence is linked to success.

This year, record numbers of high school students obtained top grades in their final exams, yet employers complain that young people still lack the basic skills to succeed at work. The only explanation offered is that exams must be getting easier. But the real answer could lie in a study just published by Professor Robert Sternberg, an eminent psychologist at Yale University in the USA and the world's leading expert on intelligence. His research reveals the existence of a totally new variety: practical intelligence.

Professor Sternberg's astonishing finding is that practical intelligence, which predicts success in real life, has an inverse relationship with academic intelligence. In other words, the more practically intelligent you are, the less likely you are to succeed at school or university. Similarly, the more paper qualifications you hold and the higher your grades, the less able you are to cope with problems of everyday life and the lower your score in practical intelligence.

Many people who are clearly successful in their place of work do badly in standard IQ (academic intelligence) tests. **Entrepreneurs** and those who have built large businesses from scratch are frequently discovered to be high school or college drop-outs. IQ as a concept is more than 100 years old. It was supposed to explain why some people excelled at a wide variety of intellectual tasks. But IQ ran into trouble when it became apparent that some high scorers failed to achieve in real life what was predicted by their tests.

Emotional intelligence (EQ), which emerged a decade ago, was supposed to explain this deficit. It suggested that to succeed in real life, people needed both emotional as well as intellectual skills. EQ includes the abilities to motivate yourself and persist in the face of **frustrations**; to control impulses and delay **gratification**; to

regulate moods and keep **distress** from swamping the ability to think; and to understand and empathize with others. While social or emotional intelligence was a useful concept in explaining many of the real-world **deficiencies** of super intelligent people, it did not go any further than the IQ test in measuring success in real life. Again, some of the most successful people in the business world were obviously lacking in social charm.

Not all the real-life difficulties we face are solvable-with just good social skills - and good social **acumen** in one situation may not translate to another. The crucial problem with academic and emotional intelligence scores is that they are both poor predictors of success in real life. For example, research has shown that IQ tests predict only between 4% and 25% of success in life, such as job performance.

Professor Sternberg's group at Yale began from a very different position to traditional researchers into intelligence. Instead of asking what intelligence was and investigating whether it predicted success in life, Professor Sternberg asked what distinguished people who were thriving from those that were not. Instead of measuring this form of intelligence with mathematical or verbal tests, practical intelligence is scored by answers to real-life **dilemmas** such as: 'If you were travelling by car and got stranded on a motorway during a blizzard, what would you do?' An important contrast between these questions is that in academic tests there is usually only one answer, whereas in practical intelligence tests - as in real life - there are several different solutions to the problem.

The Yale group found that most of the really useful knowledge which successful people have acquired is gained during everyday activities - but typically without conscious awareness. Although successful people's behaviour reflects the fact that they have this knowledge, high achievers are often unable to articulate or define what they know. This partly explains why practical intelligence has been so difficult to identify.

Professor Sternberg found that the best way to reach practical intelligence is to ask successful people to relate examples of crucial incidents at work where they solved problems demonstrating skills they had learnt while doing their jobs. It would appear that one of the best ways of improving your practical intelligence is to observe master practitioners at work and, in particular, to focus on the skills they have acquired while doing the job. Oddly enough, this is the basis of traditional **apprentice** training.

Historically, the junior doctor learnt by observing the consultant surgeon at work and the junior lawyer by assisting the senior barrister.

Another area where practical intelligence appears to resolve a previously unexplained paradox is that performance in academic tests usually declines after formal education ends. Yet, most older adults contend that their ability to solve practical problems increases over the years. The key implication for organizations and companies is that practical intelligence may not be detectable by conventional **audits** and performance measuring procedures. Training new or less capable employees to become more practically intelligent will involve learning from the genuinely practically intelligent rather than from training manuals or courses.

Perhaps the biggest challenge is in **recruitment**, as these new studies strongly suggest that paper qualifications are unlikely to be helpful in predicting who will be best at solving your company's problems. Professor Sternberg's research suggests that we should start looking at companies in a completely different way - and see them as places where a huge number of problems are being solved all the time but where it may take new eyes to see the practical intelligence in action.

Using the words written in bold in the text, rewrite the text in your own words:

a) **explain the differences between academic intelligence, emotional intelligence and practical intelligence;**
b) **give examples of how different intelligence types function;**
c) **provide the suggestions for companies to improve practical intelligence.**

(<u>You can use a dictionary to look up the meanings of the words)</u>

Appendix IX. Retelling with Glossary for the First Text

Number:

Please read the following passage. Based on the text, answer the questions that follow.

The Triune Brain

The first of our three brains to evolve is what scientists call the reptilian cortex. This brain sustains the elementary activities of animal survival such as **respiration**[1], adequate rest and a beating heart. We are not required to consciously "think" about these activities. The reptilian cortex also houses the "startle centre", a mechanism that facilitates swift reactions to unexpected occurrences in our surroundings. That panicked lurch you experience when a door slams shut somewhere in the house, or the heightened awareness you feel when a twig cracks in a nearby bush while out on an evening walk are both examples of the reptilian cortex at work. When it comes to our interaction with others, the reptilian brain offers up only the most basic impulses: **aggression**[2], mating, and territorial defence. There is no great difference, in this sense, between a crocodile defending its spot along the river and a turf war between two urban gangs.

Although the lizard may fight for its habitat, it exerts total **indifference**[3] toward the well-being of its young. Listen to the anguished squeal of a dolphin separated from its pod or witness the sight of elephants mourning their dead, however, and it is clear that a new development is at play. Scientists have identified this as the limbic cortex. Unique to mammals, the limbic cortex impels creatures to care for their **offspring**[4] by delivering feelings of tenderness and warmth to the parent when children are nearby. These same sensations also cause mammals to develop various types of social relations and **kinship**[5] networks.

1. **respiration**: noun (solunum) the action of breathing

2. **aggression**: noun (saldırganlık) feelings of anger or antipathy resulting in hostile or violent behaviour; readiness to attack or confront

3. **indifference**: noun (ilgisizlik) lack of interest, concern, or sympathy

4. **offspring**: noun (yavru) an animal's young.

When we are with others of "our kind" – be it at soccer practice, church, school or a nightclub – we experience positive sensations of togetherness, **solidarity**[6] and comfort. If we spend too long away from these networks, then loneliness sets in and encourages us to seek companionship.

Only human capabilities extend far beyond the scope of these two cortexes. Humans eat, sleep and play, but we also speak, plot, rationalise and debate finer points of **morality**[7]. Our unique abilities are the result of an expansive third brain – the neocortex which engages with logic, reason and ideas. The power of the neocortex comes from its ability to think beyond the present, concrete moment. While other mammals are mainly restricted to impulsive actions, humans can think about the "big picture". We can string together simple lessons (for example, an apple drops downwards from a tree; hurting others causes unhappiness) to develop complex theories of physical or social phenomena (such as the laws of gravity and a concern for human rights).

The neocortex is also responsible for the process by which we decide on and commit to particular courses of action. Strung together over time, these choices can accumulate into feats of progress unknown to other animals. Anticipating a better grade on the following morning's exam, a student can ignore the limbic urge to socialise and go to sleep early instead. Over three years, this ongoing sacrifice translates into a first class degree and a scholarship to graduate school; over a lifetime, it can mean ground-breaking contributions to human knowledge and development. The ability to sacrifice our drive for immediate satisfaction in order to benefit later is a product of the neocortex.

Understanding the triune brain can help us appreciate the different natures of brain damage and psychological disorders. The most devastating form of brain damage, for example, is a

5. **kinship**: noun (akrabalık) blood relationship
6. **solidarity**: noun (birlik, beraberlik) unity or agreement of feeling or action, especially among individuals with a common interest; mutual support within a group.
7. **morality**: noun (ahlak) principles concerning the distinction between right and wrong or good and bad behaviour.

condition in which someone is understood to be brain dead. In this state a person appears merely unconscious – sleeping, perhaps – but this is illusory. Here, the reptilian brain is functioning on autopilot despite the permanent loss of other cortexes.

Disturbances to the limbic cortex are registered in a different manner. Pups with limbic damage can move around and feed themselves well enough but do not register the presence of their littermates. Scientists have observed how, after a limbic lobotomy, "one impaired monkey stepped on his outraged peers as if treading on a log or a rock". In our own species, limbic damage is closely related to sociopathic behaviour. Sociopaths in possession of fully-functioning neocortexes are often shrewd and emotionally intelligent people but lack any ability to relate to, empathise with or express concern for others. One of the neurological wonders of history occurred when a railway worker named Phineas Gage survived an incident during which a metal rod skewered his skull, taking a considerable amount of his neocortex with it. Though Gage continued to live and work as before, his fellow employees observed a shift in the **equilibrium**[8] of his personality. Gage's animal **propensities**[9] were now sharply pronounced while his intellectual abilities suffered; garrulous or obscene jokes replaced his once quick **wit**[10]. New findings suggest, however, that Gage managed to soften these abrupt changes over time and rediscover an appropriate social manner. This would indicate that reparative therapy has the potential to help patients with advanced brain trauma to gain an improved quality of life.

8. **equilibrium**: noun (denge) a state in which opposing forces or influences are balanced.
9. **propensity**: noun (eğilim) an inclination or natural tendency to behave in a particular way.
10. **wit**: noun (ince zaka) a natural aptitude for using words and ideas in a quick and inventive way to create humour

Using the words in the word list next to the text, rewrite the text in your own words:

a) **explain the differences between the reptilian cortex, the limbic cortex and the neocortex;**
b) **provide examples for different functions of these three cortexes;**
c) **explain what problems may occur in case of any damage to these three cortexes.**

(You can use a dictionary to look up the meanings of the words)

Appendix X. Retelling with Glossary for the Second Text

Number:

Please read the following passage. Based on the text, answer the questions that follow.

Practical intelligence lends a hand

Dr Rajendra Persaud explains how practical intelligence is linked to success.

This year, record numbers of high school students obtained top grades in their final exams, yet employers complain that young people still lack the basic skills to succeed at work. The only explanation offered is that exams must be getting easier. But the real answer could lie in a study just published by Professor Robert Sternberg, an eminent psychologist at Yale University in the USA and the world's leading expert on intelligence. His research reveals the existence of a totally new variety: practical intelligence.

Professor Sternberg's astonishing finding is that practical intelligence, which predicts success in real life, has an inverse relationship with academic intelligence. In other words, the more practically intelligent you are, the less likely you are to succeed at school or university. Similarly, the more paper qualifications you hold and the higher your grades, the less able you are to cope with problems of everyday life and the lower your score in practical intelligence.

Many people who are clearly successful in their place of work do badly in standard IQ (academic intelligence) tests. **Entrepreneurs**[1] and those who have built large businesses from scratch are frequently discovered to be high school or college drop-outs. IQ as a concept is more than 100 years old. It was supposed to explain why some people excelled at a wide variety of intellectual tasks. But IQ ran into trouble when it became apparent that some

1. **entrepreneur:** noun (girişimci) a person who sets up a business or businesses, taking on financial risks in the hope of profit

high scorers failed to achieve in real life what was predicted by their tests.

Emotional intelligence (EQ), which emerged a decade ago, was supposed to explain this deficit. It suggested that to succeed in real life, people needed both emotional as well as intellectual skills. EQ includes the abilities to motivate yourself and persist in the face of **frustrations**[2]; to control impulses and delay **gratification**[3]; to regulate moods and keep **distress**[4] from swamping the ability to think; and to understand and empathize with others. While social or emotional intelligence was a useful concept in explaining many of the real-world **deficiencies**[5] of super intelligent people, it did not go any further than the IQ test in measuring success in real life. Again, some of the most successful people in the business world were obviously lacking in social charm.

Not all the real-life difficulties we face are solvable-with just good social skills - and good social **acumen**[6] in one situation may not translate to another. The crucial problem with academic and emotional intelligence scores is that they are both poor predictors of success in real life. For example, research has shown that IQ tests predict only between 4% and 25% of success in life, such as job performance.

Professor Sternberg's group at Yale began from a very different position to traditional researchers into intelligence. Instead of asking what intelligence was and investigating whether it predicted success in life, Professor Sternberg asked what distinguished people who were thriving from those that were not. Instead of measuring this form of intelligence with mathematical or verbal tests, practical intelligence is scored by answers to real-life **dilemmas**[7] such as: "If you were travelling by car and got stranded on a motorway during a blizzard, what would you do?" An important contrast between these questions is that in academic tests there is usually only one answer,

2. **frustration**: noun (hüsran) The feeling of being upset or annoyed as a result of being unable to change or achieve something

3. **gratification**: noun (haz) Pleasure, especially when gained from the satisfaction of a desire

4. **distress**: noun (ızdırap, dert) Extreme anxiety, sorrow, or pain

5. **deficiency:** noun (eksiklik) a lack or shortage

6. **acumen**: noun (keskin zeka) The ability to make good judgements and take quick decisions

7. **dilemma**: noun (ikilem) A situation in which a difficult choice has to be made between two or more alternatives, especially ones that are equally undesirable

whereas in practical intelligence tests - as in real life - there are several different solutions to the problem.

The Yale group found that most of the really useful knowledge which successful people have acquired is gained during everyday activities - but typically without conscious awareness. Although successful people's behaviour reflects the fact that they have this knowledge, high achievers are often unable to articulate or define what they know. This partly explains why practical intelligence has been so difficult to identify.

Professor Sternberg found that the best way to reach practical intelligence is to ask successful people to relate examples of crucial incidents at work where they solved problems demonstrating skills they had learnt while doing their jobs. It would appear that one of the best ways of improving your practical intelligence is to observe master practitioners at work and, in particular, to focus on the skills they have acquired while doing the job. Oddly enough, this is the basis of traditional **apprentice**[8] training. Historically, the junior doctor learnt by observing the consultant surgeon at work and the junior lawyer by assisting the senior barrister.

Another area where practical intelligence appears to resolve a previously unexplained paradox is that performance in academic tests usually declines after formal education ends. Yet most older adults contend that their ability to solve practical problems increases over the years. The key implication for organizations and companies is that practical intelligence may not be detectable by conventional **audits**[9] and performance measuring procedures. Training new or less capable employees to become more practically intelligent will involve learning from the genuinely practically intelligent rather than from training manuals or courses.

Perhaps the biggest challenge is in **recruitment**[10], as these new studies strongly suggest that paper qualifications are unlikely to be helpful in predicting who will be best at solving your company's

8. **apprentice**: noun (çırak) A person who is learning a trade from a skilled employer, having agreed to work for a fixed period at low wages

9. **audit**: noun (denetim) an official inspection of an organization's accounts, typically by an independent body

10. **recruitment**: noun (işe alım) the action of finding new people to join an organization or support a cause

problems. Professor Sternberg's research suggests that we should start looking at companies in a completely different way - and see them as places where a huge number of problems are being solved all the time but where it may take new eyes to see the practical intelligence in action.

Using the words in the word list next to the text, rewrite the text in your own words:

a) **explain the differences between academic intelligence, emotional intelligence and practical intelligence;**

b) **give examples of how different intelligence types function;**

c) **provide your suggestions for companies to improve practical intelligence.**

Appendix XI. Vocabulary Tests for the First Text

(Pre-test, Immediate Post-test, Delayed Post-test)

Number: **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**	**Number:** **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**	**Number:** **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**
aggression	solidarity	offspring
equilibrium	wit	indifference
indifference	indifference	solidarity
kinship	equilibrium	propensity
morality	morality	wit
offspring	propensity	respiration
propensity	kinship	aggression
respiration	aggression	morality
solidarity	offspring	kinship
wit	respiration	equilibrium

Appendix XII. Vocabulary Tests for the Second Text

(Pre-test, Immediate Post-test, Delayed Post-test)

Number: **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**	**Number:** **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**	**Number:** **Write the Turkish equivalents or English explanations of the following words. If you don't know the meaning, put "X".**
acumen	gratification	recruitment
apprentice	apprentice	deficiency
audit	recruitment	acumen
deficiency	dilemma	apprentice
dilemma	frustration	gratification
distress	acumen	entrepreneur
entrepreneur	audit	frustration
frustration	deficiency	dilemma
gratification	distress	audit
recruitment	entrepreneur	distress

Appendix XIII. Consent Form for the Instructors

ARAŞTIRMA GÖNÜLLÜ KATILIM FORMU

Bu çalışma, "Görev kaynaklı katılım düzeyinin İngilizce Öğretmenliği bölümü öğrencilerinin rastlantısal kelime öğrenimi ve akılda tutmaları üzerindeki etkisi" başlıklı bir araştırma çalışması olup rastlantısal kelime öğrenimi üzerinde değişik görevlerin etkisini bulma amacını taşımaktadır. Çalışma, Araştırma Görevlisi Tuncay Karalık tarafından yürütülmekte ve sonuçları ile kelime öğretiminde farklı görevlerin kendine has faydaları ortaya konacaktır.

- Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır.
- Çalışmanın amacı doğrultusunda, okuma parçalarının yanında değişik görevler verilmesi suretiyle edinilen kelime sayılarının ölçümü yapılarak öğrencilerinizden veriler toplanacaktır.
- Araştırmada katılımcıların isimleri gizli tutulacaktır.
- Araştırma kapsamında toplanan veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmanın amacı dışında ya da bir başka araştırmada kullanılmayacaktır.
- İstemeniz halinde öğrencilerinizden toplanan verileri inceleme hakkınız bulunmaktadır.
- Öğrencilerinizden toplanan veriler sanal depolama yöntemi ile korunacak ve araştırma bitiminde arşivlenecek veya imha edilecektir.
- Veri toplama sürecinde/süreçlerinde size rahatsızlık verebilecek herhangi bir soru/talep olmayacaktır. Yine de katılımınız sırasında herhangi bir sebepten rahatsızlık hissederseniz çalışmadan istediğiniz zamanda ayrılabileceksiniz. Çalışmadan ayrılmanız durumunda toplanan veriler çalışmadan çıkarılacak ve imha edilecektir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı Anadolu Üniversitesi Yabancı Diller Eğitimi bölümünden Tuncay Karalık'a yöneltebilirsiniz.

Bu çalışmaya tamamen kendi rızamla katılacağımı, istediğim takdirde çalışmadan ayrılabileceğimi ve öğrencilerimden toplanan bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.

(Lütfen bu formu doldurup imzaladıktan sonra veri toplayan kişiye veriniz.)

Katılımcı Ad ve Soyadı:
İmza:
Tarih: