Author:
Xie, Haoran

Title:
Digital Game-based Vocabulary Learning

Revisiting the Involvement Load Hypothesis in Digital Game Environments

School of Education

Faculty of Social Sciences and Law

Digital Game-based Vocabulary Learning: Revisiting the Involvement Load Hypothesis in Digital Game Environments

XIE Haoran

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Education in the Faculty of Social Sciences and Law, School of Education

April 2021

Word Count: 43,290

Abstract

The involvement load hypothesis (ILH) proposed by Laufer and Hulstijn (2001) is one of the most widely acknowledged hypotheses about vocabulary learning (Hu & Nassaji, 2016). It suggests that the effectiveness of a task in promoting vocabulary learning is contingent upon the involvement load of the task, which is composed of the amount of need, search, and evaluation it imposes (Hulstijn & Laufer, 2001). Tasks with greater involvement loads tend to be more effective than tasks with lower loads, and tasks with similar involvement loads tend to promote vocabulary learning similarly (Hulstijn & Laufer, 2001). There is rich empirical evidence in support of the ILH in conventional learning environments, however, few studies have investigated the ILH in digital game environments. To fill this research gap, I conducted the present study to investigate whether the ILH can, in its present form, predict differences between digital and non-digital environments. This main research question consists of two sub-questions. (1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one? (2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form?

A total of 135 students participated in the study, and they were randomly divided into three groups. Students in Group A were asked to complete Task 1: Reading a text and answering reading comprehension questions on paper. Students in Group B were asked to complete Task 2: Reading a text and answering reading comprehension questions in a digital game. Students in Group C were asked to complete Task 3: Reading a text and inferring meanings of the underlined words in a digital game. Essentially, there were two pairs of comparisons. The first one was between Group A and Group B, which examined whether learners performed differently in a digital game environment and in a conventional one while learning with the same vocabulary learning task. The second one was between Group B and Group C, which examined whether learners performed differently on two tasks with the same involvement load in a digital game-based environment.

The participants were students from a Hong Kong university with similar educational backgrounds (first year Master of Education students), language proficiency levels (intermediate

English learners), and prior knowledge of the target vocabulary (they knew no more than two target words). Their first languages were either Mandarin or Cantonese. Before the project, their prior knowledge of the 10 target words was measured using Folse's (2006) modified vocabulary knowledge scale (MVKS). After completing the learning session, all participants were tested immediately using the MVKS. Subsequently, 10 from each group were interviewed to investigate their learning experiences and perceptions of the learning approaches. One week later, the remaining 35 students from each group were post-tested using the MVKS.

I also interviewed 30 students to investigate their learning experience. At the interviews, I asked them to reflect on their learning process. Questions included what the students found useful or useless for their vocabulary learning and what they thought about the learning task and game. The semi-structured interview was conducted to triangulate the quantitative data collected from the post-tests. It aimed to identify what features of the learning task and game motivated the students, engaged them in the learning process, directed their attention to the target vocabulary knowledge, and were considered useful by them. Analysis of these features could contribute to the better understanding of the experimental results concerning the effectiveness of the three tasks.

In the context of this study, I examined task effectiveness in terms of promoting vocabulary learning from two perspectives, the initial learning of the target vocabulary knowledge, which was evaluated by an immediate posttest, and the retention of the target vocabulary knowledge, which was evaluated by a delayed posttest. The results indicated that for the same task of reading a text and answering reading comprehension questions, learners performed better in both immediate learning and retention of vocabulary knowledge in a digital game environment than in a conventional one. The research results also indicated that, for Task 2: Reading a text and answering reading comprehension questions in a digital game and Task 3: Reading a text and inferring meanings of the underlined words in a digital game, learners performed differently. Although the two tasks induce the same involvement load (moderate need, search, and evaluation) and tend to promote vocabulary learning with similar effectiveness according to the ILH, the research results showed that Task 3 was significantly more effective than Task 2.

The interview results indicated that the superiority of digital game environments over non-game environments was likely attributable to the digital game learning environment that was

conducive to the increase in learning motivation. Moreover, as reading comprehension did not specifically direct or guide students to focus on the target vocabulary, they rarely attempted to remember the forms of the target words or figure out the exact meanings of the target words. Thus, they did not spend additional efforts on building clear form-meaning links for the target words, which might be the main reason why the reading comprehension task was less effective than the reading plus inferencing task.

Thus, I suggested that the application of ILH to digital game-based vocabulary learning may consider adding one more degree of prominence to "need" when evaluating the involvement load of a task in digital game environments. I also suggested that when learners are required to infer exact meanings of target words based on the contexts, the involvement load of search should be strong, rather than moderate.

Dedication

To Dr. Di Zou, my loving wife.

Acknowledgements

I would like to express my sincere gratitude to my EdD supervisor Prof. Guoxing Yu for his guidance and suggestions over the years. I am extremely thankful for his inspiring discussions and valuable comments during this study. I would like to thank Prof. Leon Tikly, Prof. Roger Dale, Prof. Michael Crossley, Prof. Sally Barnes, and Dr. Lisa Lucas for their patient teaching by traveling half away from the earth. I would like to thank all participants in this study. Last but not least, my deepest gratitude to my loving wife, Dr. Di Zou, for her unconditional support and care.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted to any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signature: signature removed due to permissions issues Date: 15 April 2021

Table of Contents

Abstract 1
Dedication 4
Acknowledgements 5
Author's Declaration 6
Table of Contents 7
List of Tables 10
List of Figures 11
Glossary 12
CHAPTER 1: INTRODUCTION 13
1.1 Research Background 13
1.2 Motivation 14
1.3 Significance 16
1.4 Organization 16
CHAPTER 2: LITERATURE REVIEW 18
2.1 Introduction 18
2.2 Second Language Vocabulary Acquisition 18
2.2.1 Vocabulary knowledge 19
2.2.2 Vocabulary learning 20
2.2.3 Different Types of Vocabulary Learning 21
2.3 Vocabulary Learning Theories 22
2.3.1 Motivation, Noticing, Frequency of Exposure, Generative Use, and Engagement with vocabulary 23
2.3.2 Levels of Processing Theory 26
2.3.3 Involvement Load Hypothesis 26
2.3.4 Technique Feature Analysis 29
2.3.5 Comparing Involvement Load Hypothesis to other Theories 32
2.4 Computer Assisted Language Learning 33
2.5 Digital Game-Based Vocabulary Learning 35
2.5.1 Theories Related to Digital Game-Based Vocabulary Learning 37
2.5.2 Research Trends of Digital Game-Based Vocabulary Learning 42
2.6 Summary 56
CHAPTER 3: METHODOLOGY 58
3.1 Introduction 58
3.2 Task Design 59
3.3 Research Design 62
3.3.1 Pre-test 64
3.3.2 Learning Processes 65

3.3.3 Immediate Post-test 66
3.3.4 In-depth Interview 67
3.3.5 Delayed Post-test 68
3.4 Participants 69
3.5 Game Design 72
3.5.1 Selection of the Game Type 72
3.5.2 Selection of the Game Development Tool 73
MonoGame 74
3.5.3 Selection of the Digital Tool for Playing the Game 75
3.5.4 Game Flow 76
3.6 Learning Materials 82
3.6.1 The Reading Text 82
3.6.2 The Target Vocabulary 84
3.7 Measurement Tool: The Modified Vocabulary Knowledge Scale 89
3.8 Research Ethics 92
3.9 Data Analysis 93
3.10 Summary 96
CHAPTER 4: EXPERIMENTAL RESULTS 99
4.1 Introduction 99
4.2 Overall Experimental Results 99
4.3 Test of Normality 102
4.4 Test of Homogeneity of Variance 103
4.5 Task Effectiveness in Terms of Promoting Initial Learning and Retention of Target Vocabulary Knowledge 103
4.5.1 Comparing Task 1 and Task 2 in Terms of Promoting Initial Learning 104
4.5.2 Comparing Task 2 and Task 3 in Terms of Promoting Initial Learning 105
4.5.3 Comparing Task 1 and Task 2 in Terms of Promoting Retention 106
4.5.4 Comparing Task 2 and Task 3 in Terms of Promoting Retention 107
4.6 Loss of Vocabulary Knowledge 108
4.7 Interview Results 110
4.8 Summary 113
CHAPTER 5: DISCUSSION 116
5.1 Introduction 116
5.2 Effectiveness of Digital Game-Based Vocabulary Learning 116
5.3 Digital Game Environment and Motivation 117
5.4 Digital Game Environment and Anxiety 118
5.5 Engagement with Vocabulary 119
5.6 Flow theory 120
5.7 Noticing and Form-focused Instruction 121

5.8 Revisiting ILH and TFA from the Perspectives of Digital Game-Based Vocabulary Learning 122
5.8.1 Applying ILH to Digital Game-Based Vocabulary Learning 122
5.8.2 Applying TFA to Digital Game-Based Vocabulary Learning 123
5.9 Summary 128
CHAPTER 6: IMPLICATIONS 130
6.1 Introduction 130
6.2 Theoretical Implications 130
6.3 Pedagogical Implications 131
6.4 Summary 134
CHAPTER 7: CONCLUSION AND LIMITATIONS 135
7.1 Conclusion 135
7.2 Limitations 137
REFERENCES 140
APPENDICES 161
Appendix 1. Task 1 - Reading a text and answering reading comprehension questions on paper (adapted from Zou, 2012) 161
Appendix 2. Task 2 - Reading a text and answering reading comprehension questions in a digital game (adapted from Zou, 2012) 164
Appendix 3. Task 3 - Reading a text and inferring meanings of the underlined words in a digital game (adapted from Zou, 2012) 167
Appendix 4. The Modified Vocabulary Knowledge Scale Used in the Pre-test, Immediate Post-Test, and Delayed Post-Test (Folse, 2006; Zou, 2012) 170
Appendix 5. Entries of the Target Words 171
Appendix 6. The Ethical Review Approval Letter from The Education University of Hong Kong 177
Appendix 7. The Ethical Review Approval Letter from University of Bristol 178
Appendix 8. SoE Research Ethics Form - University of Bristol 179
Appendix 9. Test Sample 183
Appendix 10. Student Interview Sample Questions 184
Appendix 11. Student Consent Form 185
Appendix 12. Information Sheet 186

List of Tables

Table 1: Four psycholinguistic factors that are important for vocabulary learning 25
Table 2: Nine factors that lead to engagement with vocabulary 26
Table 3: The dimensions, components, and degrees of the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Xie, Zou, Lau, Wang, & Wong, 2016) 27
Table 4: All Components and Criteria of Techniques Feature Analysis (Nation, & Webb, 2020) 30
Table 5: The scale of generativity (Joe, 1998; Nation & Webb, 2011) 31
Table 6: The extracted learning theoretical frameworks from Zou et al. (2019, p.13) in digital game-based vocabulary learning 37
Table 7: Foci of the questionnaires used in game-based vocabulary learning research 46
Table 8: Game types 47
Table 9: Game features 49
Table 10: Game devices and their affordances and limitations 51
Table 11: Details of the three tasks 61
Table 12: The details of three learning tasks and their relationships to the research questions 63
Table 13: The number of participants in different experimental stages 71
Table 14: The most popular free game development tools (Gamedesigning, 2019) 74
Table 15: Survey results of the topic familiarity and level of difficulty 84
Table 16: A sample of modified vocabulary levels test (Zou, 2012, p. 116) 87
Table 17: A sample questionnaire measuring word familiarity (Zou, 2012) 88
Table 18: The five scale of VKS of vocabulary knowledge (Paribakht and Wesche, 1997; Zou 2012) 90
Table 19: Folse's (2006) Modified Vocabulary Knowledge Scale (MVKS) 90
Table 20: The Modified Vocabulary Knowledge Scale (Folse, 2006) further adapted by Zou (2012) 91
Table 21: The purposes of two comparisons among the three groups 94
Table 22: The objectives and comparisons 95
Table 23: Descriptive statistics of the participants' pre-test, immediate and delayed post-test scores 100
Table 24: Results of the independent samples t-test of the participants' scores in the immediate posttest (Task 1 vs. Task 2) 105
Table 25: Results of the independent samples t-test of the participants' scores in the immediate posttest (Task 2 vs. Task 3) 106
Table 26: Results of the independent samples t-test of the participants' scores in the delayed posttest (Task 1 vs. Task 2) 107
Table 27: Results of the independent samples t-test of the participants' scores in the delayed posttest (Task 2 vs. Task 3) 108
Table 28: Descriptive statistics of the attritions 109
Table 29: Test of Homogeneity of Variances 109
Table 30: ANOVA Results 109
Table 31: Results of Multiple Comparisons 110
Table 32: Summary of the quantitative and qualitative results 114
Table 33: My suggested TFA scores 125

List of Figures

Figure 1: Publication Years 44
Figure 2: Journal titles and numbers of publications in the journals 44
Figure 3: The education levels of the participants in the early years (gray) and the recent years (blue) 52
Figure 4: The research framework of this study 62
Figure 5: The designed experimental procedures 64
Figure 6: The results of the survey for the preferred tool for playing the digital game 75
Figure 7: The detailed game flow of the learning task 77
Figure 8: The game interface of the first scenario 78
Figure 9: The game interface of the first scenario with the instructions 78
Figure 10: The game interface of the second scenario with the task card 79
Figure 11: The game interface of the second scenario of question items 79
Figure 12: The game interface of the second scenario of completion of all questions 80
Figure 13: The overall MVKS scores of the three groups of participants in the three tests 101

Glossary

Term	Definition
Checklist	In this dissertation, the checklist refers to Nation and Webb's (2011) checklist of the technique feature analysis. It consists of five main components: motivation, noticing, retrieval, generation, and retention.
Engagement	Engagement refers to the behavioural, cognitive, and emotional state of a student when s/he concentrates on learning (Elmaadaway, 2018). In this dissertation, I focused on the engagement with vocabulary. It involves frequency of exposure; attention; noticing; intention, requirements, and need to learn; manipulation of the vocabulary; and amount of time and interaction with it (Schmitt, 2008).
Frequency of Exposure	Frequency of exposure refers to the number of times that a learner encounters a word within a particular period (Schmitt, 2008).
Generative Use	Generative use of vocabulary refers to encountering a word in new contexts different from previous input or using it in original contexts (Nation, 2001).
Hypothesis	A hypothesis is an assumption proposed for the sake of argument, the successful execution of which will cause the expected consequences.
Motivation	Motivation at the language level is orientation toward the language, the speakers, and their cultures; motivation at the learner level is the desire for self-achievement; and motivation at the learning situation level is related to factors such as the learning activities, materials, contexts, and goals (Dörnyei, 1994).
Need	In this dissertation, need is the drive to learn. It is the motivational dimension of involvement as in Laufer and Hulstijn's (2001) involvement load hypothesis.
Noticing	Noticing refers to a learner's attention and conscious perception of target knowledge (Schmitt, 2000).
Strategy	In this dissertation, the strategy refers to the learning methods and conscious thoughts and actions that learners take to complete a task or achieve a learning goal.
Theory	A theory includes definitions of key variables, the domain where the theory applies, the relationships among variables, and the predictions (Wacker, 1998).

CHAPTER 1: INTRODUCTION

1.1 Research Background

English has the largest number of speakers and the third-largest number of native speakers in the world according to the statistics from Ethnologue (2019a), which is the annual publication data source for living languages collected by SIL international (SIL International, 2019), an international non-profit organization. Specifically, the total number of English speakers globally, including 379 million native (L1) speakers and 753 million second language (L2) speakers, is 1.132 billion (Ethnologue, 2019b). According to the three-circles model of English-speaking countries proposed by Kachru (2006), the countries in the world are categorized as (i) "inner circle" countries, such as the United States and the United Kingdom, which have a large number of L1 speakers; (ii) "outer circle" countries, such as India and Nigeria, which have a small number of L1 speakers and use English as their L2 in education, government or media; and (iii) "expanding circle" countries, such as China and Brazil, which have a large number of English as a foreign language (EFL) learners. Moreover, using a loose criterion of language proficiency, the total number of English speakers is more than two billion (Crystal, 2003). By 2019, there were 27 non-sovereign and 55 sovereign entities, using English as their official language (Ethnologue, 2019b). In sum, English is now widely acknowledged as the global lingua franca (Jenkins, 2006; Seidlhofer, 2005).

With the trend of globalization and the development of information technology, learning English has becomes a necessity for non-English speakers to enhance their communicative and intercultural competence in modern society (Ahmad, 2016; Sharifian, 2013). One of the most fundamental requirements for learning English is vocabulary acquisition as words are the basic units of the English language (Alqahtani, 2015; August, Carlo, Dressler, & Snow, 2005; Landau, Smith, & Jones, 1988). Although the importance of vocabulary acquisition to English language learning has been commonly acknowledged in academic and educational communities, there does not seem to have any agreement on "which vocabulary learning strategy is the most effective" (Hulstijn, 1992; Konopak et al., 1987).

Generally, research on vocabulary learning strategies falls into two categories. The first category is incidental vocabulary acquisition, which refers to "the process of acquiring vocabulary with the focus on something other than target words" (Paribakht & Wesche, 1999, p. 196). The other category is intentional vocabulary acquisition, which refers to a process of acquiring target words

deliberately without involving other irrelevant information (Nation, 2006). This second learning strategy is usually considered an "explicit" or "decontextualized" process (Elgort & Warren, 2014). In many previous studies (Nation & Newton, 1997; Schmitt, 2000), intentional vocabulary acquisition has been criticized as this strategy is quite time-consuming and least effective even if it is a necessary step before using incidental vocabulary acquisition. In contrast, incidental vocabulary acquisition embeds target words in various contexts so that the learner can implicitly acquire these words without focusing on them (Laufer & Hulstijn, 2001). As the target words are acquired in various contexts, the incidental learning strategy can facilitate the development of various aspects of vocabulary knowledge and improve learning motivation (Zou, Xie, Li, Wang, & Chen, 2014).

The rapid development of digital games (Barbosa, & Madeira, 2019; Hodgson, Man, & Leung, 2010) in recent years has provided an excellent opportunity for further facilitating incidental vocabulary acquisition. Many research studies (Chen, Tseng & Hsiao, 2018; Neville, Shelton & McInnis, 2009; Yip & Kwan, 2006; Zou, Huang & Xie, 2019) have attempted to gamify vocabulary teaching and learning in digital game environments. Digital game-based vocabulary learning (DGBVL), supported by various language learning theories, has become a popular learning approach to incidental vocabulary acquisition (Zou, Huang & Xie, 2019). For example, the contextual vocabulary learning theory states that greater mental efforts are needed for the contextual learning paradigm than conventional learning ones (Hulstijn, 1992). Therefore, better retention can be achieved by the contextual learning paradigm, which has been aptly facilitated and supported by immersive digital gaming environments (Dickey, 2011; Hamari et al., 2016; Sung & Hwang, 2013). Another supporting theory for DGBVL is multimedia learning theory (Mayer, 2003), which argues that the simultaneous information processing in rich media (e.g., video, images, and animation) can promote the effectiveness of word retention. Due to the power of game development engines and the fast development of graphic-processing hardware in recent years, digital game-based learning (DGBL) can be considered an ideal form for providing rich media for language learners (Chang, Liang, Chou, & Lin, 2017).

1.2 Motivation

The preliminary motivation for the present study is to investigate whether the theories about vocabulary learning in conventional environments (i.e., paper-based environments) can also account for the results of vocabulary learning in digital game environments. Among a range of vocabulary learning theories, the Involvement Load Hypothesis (ILH) developed by Laufer and Hulstijn (2001) is

one of the most widely acknowledged hypotheses about vocabulary learning (Hu & Nassaji, 2016). It suggests that the effectiveness of a task in promoting vocabulary learning is contingent upon the involvement load of the task, which is composed of the amount of need, search, and evaluation it imposes (Hulstijn & Laufer, 2001).

Specifically, need is the drive to learn the target vocabulary; search refers to the attempt to find the meanings or forms of the target vocabulary; and evaluation involves comparing different word meanings and forms, as well as creating new contexts for the target vocabulary. Need is moderate when learners are imposed by the task to learn the target vocabulary and is strong when learners are self-imposed to learn. Search has only one degree of prominence. Evaluation is moderate when learners compare words to decide which ones are appropriate for the contexts and strong when learners create original contexts using the target vocabulary (Laufer & Hulstijn, 2001). The sum of the degrees of the three components of the ILH is the involvement load of a certain task. According to the ILH, the involvement load of a task is associated with the cognitive load required by the task for learners' completion of it (Laufer & Hulstijn, 2001). Tasks with higher involvement load are more effective than the tasks with lower involvement load (Laufer & Hulstijn, 2001).

The ILH has been widely employed for estimating the task-induced involvement loads and the task effectiveness in promoting vocabulary learning in conventional learning environments (Eckerth & Tavakoli, 2012; Laufer, & Rozovski-Roitblat, 2015; Zou, 2012, 2017), however, few studies have investigated whether the ILH could be applied to estimate vocabulary learning effectiveness in digital game environments. Because the characteristics of digital game environments, including media richness and immersive contexts, can facilitate word retention according to multimedia and contextual learning theories (Hulstijn, 1992; Mayer, 2003), it is uncertain whether effective factors for facilitating word retention are characteristics of digital game environments or learning tasks alone. Thus, research on the application of ILH in game environments is important.

There also have been some inconsistent empirical results for the ILH in non-game tasks. For example, Zou (2017) found that composition writing tasks can be more effective in promoting word retention than sentence-writing tasks, whereas Laufer (2001) found similar levels of vocabulary acquisition for both tasks. Thus, further investigation of the ILH in terms of its applicability in estimating effectiveness of vocabulary learning tasks is needed.

To fill the research gap, this study investigates the application of the ILH in digital game environments. Specifically, based on the previous research which examined whether tasks with higher involvement load promoted more effective vocabulary learning than those with lower involvement load in conventional learning environments, I examine whether the involvement loads of different tasks are similarly associated with the task effectiveness in digital game environments. In this way, I aim to generalize the findings and empirical results attained from conventional learning tasks when measured against the results reached in tasks in digital game environments. Moreover, as there has been limited research on how to estimate involvement loads of vocabulary learning tasks in digital game environments, I aim to augment the ILH in the context of digital game-based vocabulary learning.

1.3 Significance

The main significance of this dissertation is as follows. Firstly, I examine whether the ILH, which is typically employed for estimating involvement loads of vocabulary learning tasks, is still valid for tasks in digital game environments. In this way, I generalize the application of the ILH from conventional learning settings to digital game-based learning environments. I also compare the effectiveness of vocabulary learning tasks in the conventional learning environment and the digital game environments, through which, I investigate the potential effects of learning environments on learning performance. Thirdly, I employ both qualitative and quantitative methods for data collection to provide solid evidence for the empirical results. Fourthly, I propose possible modifications to the ILH, extending it from a hypothesis that is mainly used in conventional learning settings to one that can also be applied to estimate the effectiveness of vocabulary learning tasks in digital game environments and guide digital game-based vocabulary teaching and learning. Additionally, I discuss the implications for how to integrate digital game-based vocabulary learning into teaching and learning practices. I focus specifically on task-induced involvement load and task effectiveness in conventional and digital game-based vocabulary learning settings, aiming to bridge conventional and digital game-based vocabulary learning from the perspective of the ILH.

1.4 Organization

The remaining chapters of this dissertation are organized as follows: In Chapter 2, a comprehensive literature review is presented by examining research studies about second language vocabulary acquisition, digital game-based vocabulary acquisition, and the ILH. In Chapter 3, the research methodology, including the main research questions, participants, experimental design, learning task materials, and effectiveness measurement, adopted in this study, is explained. Chapter 4

reports the empirical results corresponding to the main research questions. Chapter 5 discusses the empirical results, possible reasons for such results, and recommendations concerning how to augment the ILH and apply it in digital game environments. Chapter 6 elaborates on the implications for digital game-based vocabulary acquisition in terms of methodology and pedagogy, aiming to bridge conventional and digital game-based vocabulary learning from the perspective of the ILH. Finally, Chapter 7 summarizes the findings and limitations to provide insights into future research directions.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The general field of this research is second language vocabulary acquisition, which involves main questions concerning what vocabulary knowledge is, how vocabulary learning occurs, and what main theories about vocabulary learning are. Within this general field, I investigated game-based vocabulary learning, which also involves questions concerning what it is and based on what theories, game-based vocabulary learning is founded. So, I reviewed in this chapter two main research areas – second language vocabulary acquisition and digital game-based vocabulary learning, which can be further divided into several sub-areas (i) vocabulary knowledge, vocabulary learning, and vocabulary learning theories, including the ILH; and (ii) game-based vocabulary learning and game-based learning related theories.

Specifically, Section 2.2: *Second Language Vocabulary Acquisition* (SLVA) presents the importance of vocabulary knowledge, explains what aspects of knowledge are involved in vocabulary knowledge, and reviews various types of vocabulary learning.

Section 2.3: *Vocabulary Learning Theories* introduces main vocabulary learning theories and then elaborates on two most widely acknowledged theories, the ILH and the checklist of technique feature analysis (TFA).

Section 2.4: *Computer Assisted Language Learning* reviews the development of computer assisted language learning and typical research issues of the field.

Section 2.5: *Digital Game-Based Vocabulary Learning* (DGBVL) begins with an overview of the different paradigms for game-based vocabulary acquisition and then discusses the theoretical frameworks from the linguistic and education dimensions, which have often been used as the supporting theories for DGBVL. Furthermore, as the current study is closely related to DGBVL, a comprehensive review about the recent research studies about DGBVL are conducted.

2.2 Second Language Vocabulary Acquisition

In the literature of second language vocabulary acquisition, the importance of vocabulary knowledge has been commonly agreed upon in both academic and educational communities, although

there has been a long-term inconsistency over definitions and measurements for vocabulary knowledge (Barcroft, 2004; Landauer, Kireyev, & Panaccione, 2011; Lightbown & Spada, 2006; Singh, Hui, Chan, & Golinkoff, 2014). However, this inconsistency implies there is a need for a framework with various aspects and measurements of vocabulary knowledge. Accordingly, this section will first examine the different aspects and measurements of vocabulary knowledge, and then different vocabulary learning methods will be elaborated.

2.2.1 Vocabulary knowledge

Vocabulary acquisition is one of the most important skills necessary for learning a foreign language. David Wilkins (1972) identified the significance of words by arguing that "without grammar little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). People need to use words to express themselves in any language. Therefore, building a deep and rich vocabulary can give L2 students a solid language knowledge preparation for using English to convey their message more efficiently.

Vocabulary knowledge can be divided into a few aspects. For example, Webb (2005) divided vocabulary knowledge into five aspects – orthography, syntax, association, grammatical functions, and meaning and forms – which can be measured by receptive and productive tests. Nation (2001, p. 31) presented "a list of the vocabulary knowledge types that native-speakers typically possess and the list including (1) a word's spoken form, (2) a word's written form, (3) a words' part-of-speech, derivative forms, and grammatical patterns, (4) a word's collocations, (5) how frequently a word is used in a language, (6) the many stylistic constraints which determine if a word is appropriate in a particular context, (7) a word's conceptual meaning(s), and (8) a word's semantic network of associations." Furthermore, the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) assumed that the knowledge about meanings and forms was the most important for developing reading skills.

Additionally, vocabulary is the foundation for developing all other language skills, including reading, listening, speaking, writing, and translation. Words are the main resources used by learners when they are using English. Language learners use words in virtually all communicative contexts: when they are reading a text in English; when they are writing a letter to a friend who can only read English; when they are communicating with native English speakers, when they watch a movie without subtitles, or when they are listening to a favorite English song.

However, learning new words is difficult for students, particularly in the K-12 EFL/ESL context. One of the most difficult aspects of learning a foreign language is how to retain newly learned words (Alemi, 2010), as they often disappear from the memory if the learners simply store the words without using them. Vocabulary learning difficulties are often caused by certain features of the English language. For instance, English has many idiomatic expressions natural to native speakers. However, they require extra effort from EFL/ESL learners to master their usage. Students need to understand how multiple words whose meaning is often opaque are used together to produce meaning. Further, the spelling of English words is often inconsistent making pronunciation difficult. Words that share similar spellings may have completely different sounds, for example, the words "dough" and "tough." In addition, English vocabulary can sometimes be used differently according to its context. Learners may have many chances to use day-to-day English in informal situations; however, what they have learned in the classrooms is formal academic English, which means that there is a gap between learner's vocabulary knowledge and their language usage. Thus, they often do not hear or use what they have learned (Nation, 2001).

All these challenges can make it difficult to learn new words in English. Therefore, besides emphasizing the importance of vocabulary learning, English learners also need to actively explore efficient vocabulary learning strategies and utilize many resources to increase the effectiveness of their vocabulary acquisition. Students can become better readers and writers if they actively pursue their vocabulary acquisition. Essentially, the vocabulary learning principle is straightforward: effective retrieval of words comes from effective storage of words because effective input always precedes effective output (Nation, 2001, 2013).

2.2.2 Vocabulary learning

There are various strategies for vocabulary learning. According to Schmitt and Schmitt (1993), vocabulary learning strategies can generally be categorized into two types. The first type is when a new word is learned for the first time, such as when a dictionary is used to check the meaning of the word. The second strategy is to remember the word. Common ways of remembering words include written and verbal repetition. Sanaoui (1995) identified two vocabulary learning approaches in second languages from the perspective of how much vocabulary learning is organized by the learners. These vocabulary learning approaches include structured and unstructured approaches. Compared with the structured approach, learners who adopt an unstructured approach are more likely to engage in self-initiated activities, independent study, recording vocabulary items, reviewing their records, and

practicing the words they have learned. In another study, Gu and Johnson (1996) divided 91 vocabulary learning strategies into cognitive strategies and metacognitive regulation. Cognitive strategies include behaviors such as the use of the dictionary, contextual guessing, and note-taking. Metacognitive regulation comprises Selective Attention and Self-Initiation. Even though scholars have classified vocabulary learning strategies differently due to the difference in focus (Fan, 2003), those classifications still provide basic understandings of the strategy types.

Furthermore, some research has focused on the specific activities that help the learners to acquire new words. Day (1991) found that students can learn new words incidentally through reading. Students perform even better in vocabulary tests if they use a bilingual dictionary while reading (Luppescu & Day, 1993). In addition to dictionary, which can help to define unknown words for learners, it reveals that students learn new words more effectively while listening to stories explaining the target words (Brett, Rothlein, & Hurley, 1996). Swanborn and De Glopper (1999) similarly identified the factors that affect students' learning of unknown words when reading. Factors include learners' reading ability, text target word ratio, and vocabulary knowledge. To make vocabulary classes more interesting for learners, researchers tend to develop vocabulary learning strategies based on digital games. Nguyen and Nga (2003) demonstrated that students can learn new words that appear in games. Furthermore, Yip and Kwan (2006) showed that students perform better when using online vocabulary games as a learning tool.

From the perspective of Hong Kong university students, Fan (2003), who investigated students' use of dictionaries, found that the most frequent vocabulary learning strategy was guessing at the meaning of unknown words. In line with of Fan's findings, the most frequently used vocabulary learning strategies are contextual guessing and dictionary use for Chinese university students (Gu, 2010). Asgari and Mustapha (2011), conducting research in a Malaysian university, also investigated the common use of strategies in learning vocabulary. The researchers found that learning through reading, the use of the dictionaries and media use were the most common strategies for ESL learners.

2.2.3 Different Types of Vocabulary Learning

The academic community generally believe that the recognition and production aspects of vocabulary learning are developed through implicit vocabulary learning, while the learning of vocabulary meanings and meditational aspects is mainly through explicit vocabulary learning (Nation, 2001). Implicit vocabulary learning, a natural learning process without conscious operations, is

suitable for the learning of vocabulary forms and grammar collocations (Ellis, 2004). However, learners may be impervious to certain knowledge aspects, and thus input fails to become intake, explicit vocabulary learning with conscious attention and noticing of linguistic cues is therefore important (Schmidt, 2000). Explicit vocabulary learning induces learners' conscious involvement in selecting and processing linguistics features, and such forms of elaboration and deliberate inferencing are conducive to the learning of vocabulary meanings and use (Nation, 2001). Nevertheless, explicit vocabulary learning activities impose heavy mental processing and tend to be laborious and demanding in terms of learner concentration, so a balance between implicit and explicit vocabulary learning is crucial.

Incidental vocabulary learning happens when learners acquire vocabulary knowledge from contexts without intending to do so, typical examples of which include extensive reading and game playing (Nation, 2001). Intentional vocabulary learning happens when learners acquire vocabulary knowledge by deliberately committing lexical information to memory, typical examples of which include doing spelling exercises and memorizing a list of target words (Barcroft, 2004). In sum, learners' focal attention in incidental vocabulary learning is on something other than learning the language itself, while their focal attention in intentional vocabulary learning is the language. Incidental vocabulary learning is meaning-focused, while intentional vocabulary learning is language-focused (Zou, 2012).

The literature suggests that incidental vocabulary learning with explicit focus on vocabulary knowledge is conducive to effective vocabulary knowledge retention (Nation, 2001). This is likely because incidental vocabulary learning normally involves frequent exposures, and explicit vocabulary learning through focused exercises promises great chances of information memorization by virtue of direct attention and concentration on target knowledge (Schmitt, 2000, 2008). Therefore, incidental vocabulary learning with explicit word-focused exercises can supplement each other and lead to effective learning. Game-based vocabulary learning is a typical type of incidental vocabulary learning with explicit word-focused exercises (Zou, Huang, & Xie, 2019).

2.3 Vocabulary Learning Theories

The recognition that some ways of vocabulary learning are more conducive to effective acquisition of vocabulary knowledge than others has engendered studies that aim to identify factors promoting effective vocabulary learning. It is generally believed that motivation, noticing of key

features of vocabulary knowledge, frequency of exposure, and generative use of target vocabulary are essential factors (Nation, 2001; Schmitt, 2008).

2.3.1 Motivation, Noticing, Frequency of Exposure, Generative Use, and Engagement with vocabulary

Motivation refers to the need or drive to learn. Both Gardner and MacIntyre's model of integrative and instrumental motivation (1991) and Bagnole's model of extrinsic and intrinsic motivation (1993) believe that external reward for learning is equally important as learner interest in the learning activity itself. Concerning language learning motivation, Dörnyei (1994) proposed that motivation at the language level is orientation toward the language, the speakers, and their cultures; motivation at the learner level is the desire for self-achievement; and motivation at the learning situation level is related to factors such as the learning activities, materials, contexts, and goals. Language learning motivation influences language learners in a multitude of perspective, including their learning behaviors, attitudes, participation in learning activities, and engagement in learning processes, all of which place influences on their learning performance and persistence in learning (Gardner, 2007).

The noticing hypothesis holds that conscious noticing of key features of vocabulary knowledge is essential for learners to successfully change input into intake (Schmidt, 1995). Noticing is also referred to as attention and learners' conscious perception of selected knowledge features in some contexts (Schmitt, 2000). The chance of a word being noticed by a learner is associated with various factors, including the salience of the word in the input, the learner's previous encounters with the word, and her/his awareness of the importance of the word (Mackey, 2006; Nation, 2001). Based on the noticing hypothesis and numerous studies on effective approaches to vocabulary teaching and learning, Ellis proposed a concept form-focused instruction, which refers to "any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic forms" and argued that form-focused instruction is important for successful vocabulary learning (Ellis, 2001, p.1). Results of many empirical studies have provided support to the effectiveness of form-focused instruction and indicated that word-focused tasks are good approaches to implementing form-focused instruction (e.g., Hulstijn & Laufer, 2001; Kim, 2008; Zou, 2016).

Many researchers also believe that the frequency of the word influences the chance of a word being noticed to a large extent, and frequent exposure to the target vocabulary is conducive to

successful learning (Schmitt, 2008). As complete mastery of a word involves learning a few component knowledge aspects, vocabulary learning is incremental in nature and a cumulative process with subsequent encounters building on previous ones (Nation, 2001). The first encounter with a new word may inform a learner how the word looks or sounds like and what it means in that context. The learner may notice the word class through subsequent encounters, but consolidation of these knowledge aspects demands a few more encounters, and even more for the enhancement of contextual knowledge aspects, such as the collocational and semantic-grammatical constraints on word use (Schmitt, 2008). The relationship between word knowledge and reading is reciprocal. Reading leads to word knowledge development, while word knowledge is indispensable for reading comprehension. Some researchers believe that 95% of the running words in a text need to be known for thorough understanding; and knowing 98-99% of the words in a text is desirable for incidental vocabulary learning to take place while reading for pleasure (Nation, 2001). Thus, extensive reading with graded readers, the syntax and lexis of which are controlled to make the content accessible to learners, is widely recommended (Nation & Anthony, 2013).

Generative use of vocabulary, which refers to the meeting or using a word in new contexts, is of low degree when the context where the word is used is slightly different from the original input but high when the context is substantially different (Nation, 2001). High degrees of generative use indicate that a learner has started to integrate the vocabulary knowledge into her/his language system (Nation, 2001). The results of a number of empirical studies show that higher degrees of generative use (i.e., filling target words in blanks based on given contexts, creating original contexts for target words, etc.) are more effective than lower degrees of generative use (i.e., encountering target words in new contexts while reading or listening, etc.) in terms of promoting the learning of vocabulary knowledge (Joe, 1998; Nation, 2001; Schmitt, 2008). Generative use has direct implications for pedagogy as it promotes effective learning through promoting learners to re-conceptualize and consolidate their word knowledge (Joe, 1998; Newton, 1995; Ebrahimzadeh, 2017).

These four psycholinguistic factors, motivation, noticing of key features of vocabulary knowledge, frequency of exposure, and generative use of target vocabulary all play important roles in effective vocabulary learning. Motivation views learning from the perspective of learners' willingness to learn; noticing is a prerequisite for transforming input into intake; frequency of use strengthens and consolidates prior knowledge; and generative use indicates learners' reconceptualization of vocabulary knowledge (Nation, 2001; Schmitt, 2008). Effective vocabulary learning ought to involve all four

factors, and those which involve more factors tend to promote better learning than those which involve fewer factors (Zou, 2012). A summary of these four factors and their importance for vocabulary learning is presented in Table 1.

Motivation	Motivation influences students' learning behaviors, attitudes, participation in learning activities, and engagement in learning processes, all of which relate to their learning performance and persistence in learning.
Noticing	Conscious noticing of key features of vocabulary knowledge is essential for learners to successfully change input into intake
Frequency of exposure	Frequent exposure to the target vocabulary is essential for the learning of different aspects of word knowledge.
Generative use	Generative use promotes learners to re-conceptualize and consolidate their word knowledge.

Table 1: Four psycholinguistic factors that are important for vocabulary learning

Additionally, Schmitt (2008) summarized nine factors that would lead to engagement with vocabulary. Firstly, Schmitt believed that increased frequency of exposure could increase the chance of remembering a target word. Secondly, attention on the target vocabulary knowledge is essential, and increased attention can lead to successful learning of the target knowledge. Thirdly, noticing of the target knowledge aspect plays an important role, and increased noticing is conducive to effective learning. Fourthly, learners ought to have intention to learn, and when their intention increases, the learning effectiveness increases as well. Fifthly, when learners feel the requirement to learn, possible causes of which might be learners' awareness that the target vocabulary knowledge could fill a knowledge gap for them or teachers' requirements, their learning engagement increases. Sixthly, when learners feel a need to learn or use the target vocabulary, a possible cause of which might be the need to complete a learning task, their engagement increases. Seventhly, with increased manipulation of the target vocabulary and its properties, learners' engagement increases. Eighthly, increasing amount of time on learning leads to increasing engagement with vocabulary. Lastly, increased amount of interaction with the target vocabulary can enhance learning engagement and lead to successful learning (Schmitt, 2008). These nine factors summarize the main features of effective vocabulary learning; however, they seem overlap with each other sometimes. Thus, it is not easy for teachers and teachers to evaluate the exact number of factors that a task involves. Consequently, an easier to apply theory or model concerning effective vocabulary learning features is necessary. A summary of the nine factors

is presented in Table 2.

1)	Increased frequency of exposure
2)	Increased attention on the lexical item
3)	Increased noticing of the lexical item
4)	Increased intention to learn the item
5)	A requirement to learn the item (by a teacher or syllabus)
6)	A need to learn/use the item (for a task or a personal goal)
7)	Increased manipulation of the item and its properties
8)	Increased amount of time spent engaging with the item
9)	Amount of interaction with the item

Table 2: Nine factors that lead to engagement with vocabulary

2.3.2 Levels of Processing Theory

The levels of processing theory argues that the chance of remembering a piece of information in long-term memory depends on the depth with which it is processed, not the length of time that it is held in short-term memory (Craik & Lockhart, 1972). Compared to superficial information processing (e.g., answering questions about structural and phonemic aspects of a word), meaningful information processing (e.g., answering questions about whether a word fits a context or to what category it belongs) leads to deeper processing and better retention (Craik & Tulving, 1975).

Specifically, the depth of processing qualitatively refers to various types of processing, including the shallowing processing of structural aspects of vocabulary knowledge and the deep processing of word meanings (Brown & Craik, 2000). While the elaboration refers to the richness or extensiveness of processing within one level (Lockhart & Craik, 1990).

Numerous empirical results in the literature can serve as evidence in support of this theory of levels of processing. However, it is limited in that it does not state explicitly what constitutes a level of processing, which level is deeper than another, and why (Hulstijn & Laufer, 2001). It is generally a framework of heuristic value.

2.3.3 Involvement Load Hypothesis

Acknowledging the significance of the levels of processing theory, the involvement load

hypothesis argues that deep information processing is essential for successful information memorization and realizes the necessity of operationalizing the cognitive notions (i.e., depth of processing and elaboration on vocabulary information) from the perspective of vocabulary learning tasks (Hulstijn & Laufer, 2001). Thus, among several hypotheses which have been proposed to explain vocabulary learning and task effectiveness in promoting vocabulary learning, the ILH has been widely recognized by the academic community.

The ILH can be considered as a framework for investigating incidental vocabulary learning (Reynolds, 2017). Specifically, need refers to the momentum of completing the task. It has two degrees of prominence. It is moderate when the learning drive is imposed by external factors and strong when it is imposed by the learners themselves. Search is the attempt to look for the meaning or form of a word and has only one degree of prominence. Evaluation involves either comparing different meanings or forms to select one that is most suitable for the given context or creating original contexts using target words (Laufer & Hulstijn, 2001). Evaluation is strong when original contexts are created for the target words while moderate when not. Table 3 shows the three components of ILH and their degrees of prominence. Search and Evaluation are from the cognitive dimension, and Need is from the motivational dimension (Hulstijn & Laufer, 2001).

Cognitive dimension		Motivational dimension
Search	Evaluation	Need
Moderate when learners attempt to look for the meaning or form of a word	Moderate when learners compare different meanings or forms and select one that is most suitable for the given context	Moderate when it is externally imposed.
	Strong when original contexts are created using target words	Strong when it is self-imposed by the learners themselves.

Table 3: The dimensions, components, and degrees of the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Xie, Zou, Lau, Wang, & Wong, 2016)

The ILH has been widely adopted by researchers in measuring the effectiveness of vocabulary learning. For example, Reynolds (2017) applied this hypothesis as a framework in his research. He selected a mobile game called Draw Something, which asks players to draw and guess a given word. This study investigated how involvement loads were induced while students played the mobile game.

Jung and Graf (2008) also argued that personalized learning helps learners acquire new words more effectively. The Personalized Learning Theory suggests that the instructional method should be designed to meet learner needs. Specifically, the learning objectives can be adjusted based on learners' knowledge gap and intended learning outcomes; the instructional approaches can be varied according to learners' preferences and habits; and the instructional content can be customized based on learners' prior knowledge. Xie et al. (2016) applied the ILH in personalized vocabulary learning and developed a framework, which could generate word-focused tasks via involvement load-based profiles and topic-based profiles obtained from social media. The results showed that the ILH-guided personalized vocabulary learning system promoted more effective and enjoyable vocabulary learning than intentional vocabulary learning. This indicates that appropriate use of the ILH in accordance with the learning contexts and needs has important practical impact in pedagogy.

Hazrat and Read (2021) critically reviewed the ILH and identified two categories of issues. Concerning its assumptions, the ILH is limited in terms of its "uncertainty about the relative weight of the components, lack of evidence regarding the effect of the distribution of the components, and the limited range of scores available to assign to tasks" (p. 1). Equal degrees of need, search, and evaluation do not necessarily contribute to vocabulary learning equally (Kim, 2011), and tasks with the same involvement load but different distribution of the components are not equally effective (Hu & Nassaji, 2016). Kim (2011) suggested that strong evaluation might be more important than strong need or search. Laufer (2019) also argued that evaluation and search are not of equal importance in promoting vocabulary learning. Concerning its predictive ability, firstly, the ILH is influenced by time on task (Hazrat & Read, 2021). Hulstijn and Laufer (2001) argued that time on task was different from time on target vocabulary, so the ILH did not consider time on task. However, Keating (2008) found that when time on task was controlled, tasks with higher involvement load were not more effective than those with lower load. Huang et al.'s (2012) meta-analysis also indicated that longer time on task led to better vocabulary learning. Secondly, the predictive ability of the ILH is influenced by students' level of proficiency (Hazrat & Read, 2021; Kim, 2011). Keating (2008) doubted that the ILH only applied to beginning learners. Thirdly, the predictive ability of the ILH is influenced by the frequency of students' exposure to target words (Hazrat & Read, 2021). Flose (2006) compared one sentence writing task and three cloze exercises and found that the former less effective than the latter though its involvement load was greater, so it seemed that frequency of exposure played a more influential role than involvement load in determining task effectiveness. Eckerth and Tavakoli (2012) also argued that frequency of exposure and task involvement load were both important for vocabulary learning,

nevertheless, the effect of frequency of exposure did not seem to last over time, but that of involvement load did.

Another limitation of the ILH is that most studies on it were conducted in non-game environments, so it is uncertain whether the ILH can, in its present form, predict differences between digital and non-digital environments. Thus, I conducted the present study to investigate whether the ILH can predict differences between digital and non-digital environments. I also aimed to discuss the issues concerning the relative weight of the three components of the ILH: need, search, and evaluation in response to Hazrat and Read's (2021) call for research in this direction.

2.3.4 Technique Feature Analysis

The checklist of Technique Feature Analysis (TFA) is another commonly used vocabulary learning theory. It was developed by Nation and Webb (2011) and examines the involvement of five main categories (i.e., motivation, noticing, retrieval, generation, and retention). Each category further includes three to five items. There are altogether 18 questions in the checklist, and point values are used to evaluate different techniques (Nation & Webb, 2011). The scores for each criterion question are in binary mode (either 0 or 1), so the score of TFA ranges from 0 to 18. The full list is presented in the Table 4.

Component	Criteria	Scores	
Motivation	Is there a clear vocabulary learning goal?	0	1
	Does the activity motivate learning?	0	1
	Do the learners select the words?	0	1
Noticing	Does the activity focus attention on the target words?	0	1
	Does the activity raise awareness of new vocabulary	0	1
	learning?	0	1
	Does the activity involve negotiation?		
Retrieval	Does the activity involve retrieval of the word?	0	1
	Is it productive retrieval?	0	1
	Is it recall?	0	1
	Are there multiple retrievals of each word?	0	1
	Is there spacing between retrievals?	0	1
Generation	Does the activity involve generative use?	0	1
	Is it productive?	0	1
	Is there a marked change that involves the use of other	0	1
	words?		
Retention	Does the activity ensure successful linking of form and	0	1
	meaning?	0	1
	Does the activity involve instantiation?	0	1
	Does the activity involve imaging?	0	1
	Does the activity avoid interference?		
Total score			18

Table 4: All Components and Criteria of Techniques Feature Analysis (Nation, & Webb, 2020)

Motivation: this component contains three criteria regarding the motivational factors of the learning tasks. They ask whether the task has a clear vocabulary learning goal, whether it motivates learning, and whether learners could select the words to learn by themselves.

Noticing: this component contains three criteria about whether the learning activities involve noticing in the three ways. Specifically, they ask whether the task focuses learner attention on the target words, whether it raises learners' awareness of word learning, and whether it involves negotiation.

Retrieval: this component contains five criteria regarding whether the learning activities

involve the retrieval of the target words. They ask whether the task involves retrieval of the word meaning or form, whether it is productive or recall, whether there are multiple retrievals, and whether they are spaced.

Generation: this component contains three criteria regarding whether the learning activities involve the generative use of the target words from three aspects. They ask whether the learning activity involves generative use, whether it is productive, and whether there is a marked change that involves the use of other words. Th last criterion is a further follow-up to the previous one and is only applicable for productive generative use. Joe (1998) proposed a scale for measuring different degrees of generativity, the details of which are shown in Table 5. Generally, there are four degrees of prominence of generativity: (i) *no generation* refers to repeating/copying/reading the text from the materials directly; (ii) *low generation* refers to slight changes in the text in terms of grammar or inflection; (iii) *reasonable generation* refers to substantial changes in the text in terms of grammar or collocation*;* and (iv) *high generation* refers to elaborating or stretching the meaning or changing the derivational affixes. For this criterion, the involvement of the use of other words indicates that the degree of prominence of generativity is high. For example, if the learning activity involves sentence production based on the target words, the score of this criterion will be given one point as the sentence production involves the use of other words in the productive generative use.

Generativity Scale	Descriptions
0 No generation	repeating/copying/reading the text from the materials directly.
	slight changes in the text in terms of grammar or inflection.
1 Low generation	substantial changes in the text in terms of grammar or collocation.
2 Reasonable generation	
3 High generation	elaborating/stretching the meaning or changing the derivational affixes

Table 5: The scale of generativity (Joe, 1998; Nation & Webb, 2011)

Retention: this component contains four criteria regarding whether the learning activities involve different methods for promoting the retention of target words. The first one asks whether the activity ensures successful form and meaning linkage, the second one asks whether it involves learners noticing an instance of using the target words in a meaningful context or not. The third criterion checks whether the learning activity involves the learner seeing visual images of the meanings of target words or deliberately establishing a linkage between visual images and the meanings of target words or not.

The last one checks whether the learning activity involves the learning of several unknown words that share similar/opposite meanings, or not.

2.3.5 Comparing Involvement Load Hypothesis to other Theories

Compared to ILH, the TFA seems more accurate in terms of evaluating the effectiveness of word-focused tasks, as it includes more items which represent a wider range of specific features that play important roles in retention of target vocabulary knowledge. Nation and Webb (2020) compared the ILH and TFA and found that the use of TFA can be more distinguishable than ILH for estimating the effectiveness of vocabulary learning tasks. Zou and Xie (2018) also adopted TFA in a personalized vocabulary learning system and found it useful for estimating task effectiveness.

However, the ILH is an easier approach than TFA to conceptualizing the mental efforts induced by certain tasks. TFA contains five components (i.e., motivation, noticing, retrieval, generation, and retention), and each component includes three to five sub-categories. If one wants to check the TFA of a vocabulary learning task, s/he needs to answer 18 questions in total. However, the ILH is comprised of three components (i.e., need, search, and evaluation). If one wants to check the involvement load of a vocabulary learning task, s/he only needs to answer three sets of questions: (1) Is need involved? What is the degree of it? (2) Is search involved? What is the degree of it? (3) Is evaluation involved? What is the degree of it? Thus, the evaluation of a task in terms of the TFA is more complicated than the ILH, which includes only three dimensions. However, some researchers compared ILH and TFA and suggested that TFA was more accurate than ILH (e.g., Chaharlang & Farvardin, 2018; Gohar, Rahmanian & Soleimani, 2018; Hu & Nassaji, 2016; Zou, Wang, Xie, Cheng, Wang, & Lee, 2020). This is likely because TFA includes more items than ILH, because of which, it appears a bit redundant, and some criteria seem overlapped.

Nevertheless, I select ILH as the general theoretical framework for this study because it is the most widely adopted theory in the current literature on task-based vocabulary learning. It is easy to apply and has solid foundation on empirical evidence in support of its reliability in terms of estimating the effectiveness of vocabulary learning tasks. Many studies confirm that, cloze exercises induce lower involvement load (moderate need, no search, and moderate evaluation) than writing tasks (moderate need, no search, and strong evaluation) but higher load than reading tasks (moderate need, no search, and no evaluation), and cloze exercises are significantly less effective than writing tasks but more effective than reading tasks (e.g., Beal, 2007; Eckerth & Tavakoli, 2012; Huang, Willson, & Eslami,

2012; Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Mármol & Sánchez-Lafuente, 2013; Zou, 2017).

Compared to the ILH, many theories, assumptions, and psycholinguistic factors concerning vocabulary learning are less comprehensive, for example, the noticing hypothesis, the concept of motivation, frequency of exposure, and the hypothesis of generative use. Moreover, the theory of levels of processing is less explicit than the ILH, while the TFA is more demanding for the operationalization and application. Therefore, ILH is selected as the theoretical framework of this study. Based on this theoretical framework, I designed the digital game-based vocabulary learning tasks of this study and analyzed the results from the perspective of task-induced involvement load. However, I am also interested to know whether it is ILH or TFA that is better to explain the research findings, so I discuss the research results from the perspective of the technique features of the tasks as well. Therefore, in this literature review section, I reviewed both ILH and TFA.

2.4 Computer Assisted Language Learning

Computer assisted language learning (CALL) has developed for over 50 years. Warschauer reviewed the field in 2000, summarized that CALL had emerged from Structural CALL (from 1970s to 1980s) to Communicative CALL (from 1980s to 1990s), and believed that CALL had become Integrated CALL in the 21st century (Warschauer, 2000). Structural CALL considered language as a formal structural system and focused on grammar-translation and audio-lingual learning. Drill and practice were common, aiming to improve students' language accuracy. Communicative CALL regarded language as a mentally constructed system from the cognitive perspective. It focused on communicate language teaching, and communicative exercises were common in Communicative CALL, aiming to improve students' language accuracy and fluency. The main digital devices were personal computers. Integrated CALL viewed language from the socio-cognitive perspective and considered social interaction important for language knowledge and skill development. Content-based language learning, English for special purposes, and English for academic purposes were the main learning foci, and authentic discourse was essential. Multimedia and internet were applied in integrated CALL, aiming to promote students' agency in language learning (Warschauer, 2000).

Based on Warschauer (2000), Bax clarified the differences among the three stages and re-defined them as Restricted CALL (from 1960s to 1980), Open CALL (from 1980s until 2003), and Integrated CALL (after 2003) (Bax, 2003). Restricted CALL language systems include mainly closed drills and

quizzes. Open CALL systems are more advanced and include simulations, games, and computer-mediated communication. Integrated CALL focused on integrated language skills and its unique features include frequent interaction with other students and some interaction with computer through the lesson. Also, different from Restricted CALL and Open CALL, the feedback of Integrated CALL is in the format of interpreting, evaluating, commenting, and stimulating thought.

Golonka, et al. (2014) systematically analyzed 350 papers on CALL and identified four main types of technology: schoolhouse- or classroom-based technologies, individual study tools, network-based social computing, and mobile and portable devices. Typical examples of schoolhouse- or classroom-based technologies include course management systems (CMS), interactive white board, and ePortfolio. Typical examples of individual study tools include corpus, electronic dictionaries, intelligent tutoring systems, and Grammar checker. Typical examples of network-based social computing include virtual world or serious game, discussion forum, tools for instant messages, social networking sites (SNS), blogs, and Wiki. Typical examples of mobile and portable devices include tablets, personal digital assistants (PDA), and smartphones. Among different approaches to CALL, learners showed generally positive attitude toward virtual world and games (Golonka, et al., 2014).

Gillespie (2020) conducted a bibliometric analysis of 777 articles on CALL published from 2006 to 2016. He found that the most studied CALL topics included computer-mediated communication, natural language processing, and Web 2.0. The less studied topics included mobile-assisted language learning, multimedia, virtual reality, blended learning, and games. The scarcely studied topics are web and virtual learning environments. The least studied topics include interactive whiteboard and Massive Open Online Courses. Digital game-based language learning was a popular topic in CALL though it was not one of the most investigated topics (Gillespie, 2020).

I also investigated 1295 articles on CALL with my collaborators using topic modeling and bibliometrics and found that the most popular CALL topics include digital multimodal composing, mobile devices, multimedia, captions and subtitles, audiovisual resources, computer-mediated communication, digital games, glosses and annotations, wiki, digital books, automatic speech recognition, virtual world, and virtual reality (Chen et al., 2021). Consistent with the findings of Golonka, et al. (2014), we found that digital games were widely applied in CALL. Moreover, our trend test indicated that digital games had been increasingly used in recent years and were generally regarded as effective means of language learning. This is perhaps because many research results indicated that

digital games could promote learners' engagement and interactions and enhance their language development (Chen et al., 2021).

2.5 Digital Game-Based Vocabulary Learning

Digital game-based vocabulary learning is an active sub-field of computer assisted vocabulary learning. Digital games are popularly applied by practitioners and researchers to enhance students' learning performance and increase their learning motivation. Many language learners consider vocabulary learning boring and feel frustrated as they often forget what they have learned several days ago (Zou, 2016). Games appear to be an effective means of vocabulary learning as they are normally interesting, and students are generally more willing to repeat learning and practices in the format of game playing (Erhel & Jamet, 2013; Yang, Chang, Hwang, & Zou, 2020; Zou et al., 2019).

Learners' recall of newly learned words can also be improved if the targeted words are being used in gaming tasks, although there are differences in the learning efficiency of different types of tasks. For output-oriented activities, writing tasks can facilitate word retention more effectively than reading tasks (Jahangard & Movassagh, 2011), and sentence scramble exercises have been shown to be more effective than gap-fill tasks (Haratmeh, 2012). For input-oriented tasks, vocabulary acquisition occurs better when learners have the chance to perform an action according to the meaning of the input materials (Shintani, 2012). Accordingly, more "depth of processing" from the learners is desired when language learning is involved in task performance (Craik & Lockhart, 1972). As deeper processing involves more extensive cognitive processing and tends to appear when people are conducting tasks of information searching and decision making, these types of tasks have better potential to facilitate learning and retention. Situated in the context of a game-based vocabulary learning environment, the deep processing of concepts that are represented by the targeted words are facilitated when learners are taking proactive roles in the games.

Furthermore, learners are more willing to recall newly learned words if they need to judge the concepts that are represented by the words to reach success in compelling problem-solving scenarios. For instance, based on research by Nairne et al. (2007), it is more effective for learners to learn novel words based on a survival scenario in the wilderness than learning the vocabulary by moving to a new house in the city. Because of the "survival processing effect," students will feel a sense of urgency and actively put more effort into exploring multiple roles for the learning items (Toyota & Kikuchi, 2005).

Strong affective factors, including fear or excitement, caused by natural predators or human attackers in an imagined survival scenario can help facilitate memory and recall in vocabulary learning (Soderstrom & McCabe, 2011). Gameplayers' emotions can be stimulated using various strategies. Usually, game developers design obstacles, problems, and decision-making tasks for players to manage under different situations so that learners will be motivated to explore possible solutions to the problems. Additionally, if the players are interested in designed situations, they may generate strong emotions. The attraction levels of the challenges or scenarios presented by the games for are different for each learner. In this case, how the games facilitate learners' vocabulary learning can be influenced to a great extent by the players' interest in learning the subject matter represented by the games.

Prior research has investigated the efficacy of using computer games and simulations to facilitate vocabulary learning, with mixed results being reported. For example, some studies have reported positive results of the digital game, Sims, in helping university students in their English language learning (Miller & Hegelheimer, 2006; Ranalli, 2008). In contrast, the research results of game-based vocabulary learning in other studies have been less promising. DeHaan and colleagues (2010) designed a laboratory-based experiment with a music videogame. Participants in this study were required to learn the lyrics in videogames in pairs, where one person played while the other watched. Researchers tested the participants' vocabulary acquisition twice, once immediately after the task, and the other two weeks later. The results revealed the game players recalled less vocabulary than the watchers; and there was a dramatic decrease in vocabulary retention for all the participants in the second test. The authors attempted to explain the result using the cognitive load theory: the game players took multiple tasks of finishing the game tasks and learning the lyrics simultaneously, possibly generating extraneous cognitive load during the learning process. As a result, participants' learning outcomes were negatively influenced by the extraneous cognitive load.

Several principles regarding the use of games to facilitate vocabulary learning in SLA have been generated by research studies; however, the efficiency of game-based vocabulary learning remains uncertain. Therefore, it seems timely to design digital games under the guidance of the existing theoretical frameworks to assess their effectiveness for vocabulary acquisition.

2.5.1 Theories Related to Digital Game-Based Vocabulary Learning

Most of the theories or theoretical frameworks adopted in game-based vocabulary studies fall into two dimensions, the linguistic and the educational dimension, according to a recent review study about game-based vocabulary learning (Zou et al., 2019). A summary of all relevant theories, theoretical concepts, and instructional approaches is presented in Table 6.

Dimension	Theories, theoretical concepts, and instructional approaches
Linguistics	Involvement load hypothesis/levels of processing hypothesis Extramural English Input hypothesis CALL task appropriateness
Education	Game-based learning Interaction hypothesis/collaborative dialog/ competitive learning Zone of proximal development/Scaffolding Generative theory of multimedia learning Situated learning/contextual learning Motivation Activity theory Connectivity theory Personalized learning Flow theory

Table 6: The extracted learning theoretical frameworks from Zou et al. (2019, p.13) in digital game-based vocabulary learning

(1) *The Involvement Load Hypothesis* was applied by Ranalli (2008), Sandberg et al. (2014), and Ali Mohsen (2016) to design the research and explain the empirical results.

(2) *Extramural English* refers to the type of out-of-classroom learning when learners learn and use language in digital settings based on personal interests (Sundqvist & Wikström, 2015). It is initiated and undertaken by learners independently without teacher evaluation or guidance. Typical extramural English learning activities include listening to English songs, watching English movies, posting English comments on social media, or chatting with other game players in English (Sundqvist & Wikström, 2015; Sylvén & Sundqvist, 2012).

(3) *Input Hypothesis* argues that learning occurs when learners receive input that is slightly higher than their current level of language proficiency (Krashen, 1982). Sylven and Sundqvist (2012) applied this hypothesis to guide their research on game-based extramural English learning for young learners.

(4) *Computer-Assisted Language Learning Task Appropriateness* is a framework developed by Chapelle (2001) to evaluate whether a task is appropriate for learning or not by examining six components: learner fit, positive meaning focus, language learning potential, impact, authenticity, and practicality.

In the education dimension, a total of 10 theories are employed to guide the design of research or explain the results. The details of these theories are listed as follows.

(1) *Game-based learning* theories in general focus on the positive learning outcomes of digital games for education (Boyle et al., 2016; Connolly, Boyle, Hainey, McArthur, & Boyle, 2012; Müller, Son, Nozawa, & Dashtestani, 2018; Yu, 2018), the social and psychological factors of digital games for facilitating language learning (Peterson, 2010; Peterson, 2013), features that make learning attractive (Gee, 2009), and the student-centered perception of digital games (Blake, 2011).

(2) *The Interaction hypothesis* (Long, 1981), which is further developed as collaborative dialog by Swain (2000) and social interaction (Verga & Kotz, 2017), claims that collaboration, interaction, and/or competition can result in effective learning, which has been supported by several empirical studies (Karakostas & Demetriadis, 2011; Whittemore, 1924). Creation of rich opportunities for interaction is a key feature of digital games, which lead to successful learning effectively.

(3) *Zone of proximal development* proposed by Vygotsky (1980) is a well-known theory claiming that a reasonable development process of learning is to fill the closest knowledge gaps between the learners' actual levels with guidance. The zone of proximal development can be viewed as a kind of scaffolding, which provides actions or assistance according to the learner's requirement (Wesiak et al., 2014). Many digital games applied this theory to guide their design for effective learning.

(4) *The generative theory of multimedia learning* (Mayer, 1997; Mayer, 1999a; Mayer 2001; Mayer & Moreno, 2002) makes use of the multi-modality of learning materials including text, images, and videos for promoting the learning effectiveness of learners as the cognitive load will increase if learners receive information from multiple channels. There have been several

digital game-based vocabulary learning studies (Ali Mohsen, 2016; Sandberg et al., 2014; Smeets & Bus, 2015; Wu & Huang, 2017; Young & Wang, 2014), which use the multimedia theory as a theoretical foundation. Also, many digital games integrate multimedia into their design and promote learning effectively.

(5) *Situated learning* (Brown, Collins, & Duguid, 1989), also known as contextual learning, integrates language learning into real or simulated contexts and situations that facilitates learners' understanding of the actual use of the language. Recently, researchers have implemented situated/contextual language learning in digital game environments using mobile technologies (Hwang & Wang, 2016; Sandberg et al., 2014). The results showed that digital games are an effective means of providing vivid and rich environments for situated learning.

(6) *Motivation* is a significant factor in improving learning effectiveness (Dörnyei, 1994). Some researchers (Sylvén & Sundqvist, 2012; Young & Wang, 2014) have investigated the ways digital games help to increase motivation for vocabulary learning. The literature generally indicate that digital games are effective for all students in terms of increasing learner motivation (Hung, Huang, & Hwang, 2014; Papastergiou, 2009), as they normally provide rich opportunities for interaction, immediate reward for achievement, and fun activities for learning (Zou et al., 2019). Further, students may have improved intrinsic motivation in digital game (Proulx, Romero, & Arnab, 2018), especially when the game involve peer-peer collaboration (Liao, Chen, & Shih, 2019).

(7) *Activity theory* describes a transformative process that derives outcomes between the interaction of subjects and objects by using tools (Nardi, 2003; Peña-Ayala, Sossa, & Ménde, 2014). Activity theory has been used to guide many related studies, including personalized learning environments (Buchem, Attwell, & Torres, 2011), learning support (Daniels, Edwards, Engeström, Gallagher, & Ludvigsen, 2009), teaching practices (Blina & Munrob, 2008), learning objects (Hansson, 2012), and game-based vocabulary learning (Pan, 2017).

(8) *Connectivity theory*, proposed by Klimesch (1994), assumes that vocabulary knowledge can be virtually represented as a connected network, the higher complexity of which indicates a greater possibility of vocabulary acquisition. Franciosi (2017) applied this theory to guide game design and explain vocabulary learning results.

(9) *Personalized learning* is defined as "instruction in which the pace of learning and the instructional approach are optimized for the needs of each learner. Learning objectives, instructional approaches, and instructional content (and its sequencing) may all vary based on learner needs. In addition, learning activities are meaningful and relevant to learners, driven by

their interests, and often self-initiated." (U.S. Department of Education, 2017, p. 9). According to this learning strategy, Wei et al. (2018) investigated personalized strategies for game-based vocabulary learning and found that learning effectiveness can be enhanced by personalized game-based learning.

(10) *Flow theory*, first proposed by Csikszentmihalyi (1975), describes an ideal mental state under which the learner is fully involved and immersed in performing a certain task. Flow theory has been used for guiding the development of digital educational games (Kiili, 2005) for providing an immersive and engaging gaming experience. For the achievement of flow in games, Salen and Zimmerman (2004) highlighted the importance of challenge, clear goals and feedback, and a sense of control for game design. Schell (2008) also regarded challenge, clear goals, few distractions, and direct feedback important. Song and Zhang (2008) similarly suggested that goals, feedback, and balance between challenge and students' skills are crucial for well-designed games. Digital educational games, which provide learners with flow experiences, have been shown to lead to effective learning and engagement (Wang & Chen, 2010; Wei et al., 2018). The results showed that the three dimensions of engagement (i.e., behavioral, cognitive, and emotional engagement) are interrelated, and emotional engagement alone does not necessarily lead to learning if consequent cognitive engagement is absent (Ge & Ifenthaler, 2017). A number of previous studies indicate that game environments are conducive to the increase of learner engagement from all three dimensions (Zou et al., 2019).

Concerning the effectiveness of digital game in terms of promoting vocabulary learning, studies have demonstrated that the presence of social and cultural interaction leads to successful learning. The situated cognition theory argues that social interaction and conversation are essential for learning since all knowledge cannot be separated from the activities and situations in which they are produced (Brown et al., 1989). Moreover, Jonassen (1997) further showed that problems related to communities are complex and ill-structured, which facilitate the development of schematized knowledge in the cultural and linguistic context. Based on the theories mentioned above, Neville et al. (2009) designed an interactive fictional digital game. This game aimed to engage learners by using their existing schemata in their local culture and then applying them to a foreign culture. The results indicated that the game indeed helped students to learn. Further, Cognitive Load Theory (Sweller, 1994) is often used in assessing interaction with instructional technology in learning. It claims that the learning process requires mental effort to process information. New knowledge is produced by working memory when individuals access new information. This theory was further developed as a

common measurement for instructional design to determine whether learning materials contain too much complexity (Paas, Renkl, & Sweller, 2003). For instance, DeHaan, Reed, & Kuwada (2010) adopted Cognitive Load Theory as a framework when they developed a music video game for vocabulary learning. The purpose of their research was to investigate whether there is too much cognitive load on second language learners when interactivity factors are included in games. In this study, the participants were divided into (i) watchers who only observing games-playing procedures and (ii) players who played games and had more interactions with games than watchers. The results demonstrated that cognitive loads of players were higher than watchers significantly. Another second language acquisition theory, namely the sociocultural theory, focuses on the importance of minimizing the gap between a learner's current and potential language level (Lantolf & Thorne, 2008). Stimulated by this theory, Sylvén and Sundqvist (2012) found that there is a positive relationship between the frequency of playing online role-playing games and language proficiency development. The contextual vocabulary learning theory states that contextual learning leads to higher retention of newly learned words since it requires learners to put in more mental effort than translation learning (Hulstijn, 1992). This theory implies that digital games can help facilitate vocabulary learning because they are able to provide immersive and contextual environments for learners (Dickey, 2011; Hamari et al., 2016; Sung & Hwang, 2013). Multimedia learning theory (Mayer, 2014) also demonstrated the crucial role of media in learning. This theory claims that simultaneous processing of information in multiple modalities (e.g., texts, voices, images, and so on) is conducive to successful retention.

Multimedia learning theory is not only established based on three existing theories, i.e., dual coding theory (Clark & Paivio, 1991; Paivio, 1990), cognitive load theory (Baddeley, 1992; Chandler & Sweller, 1991; Sweller, 2002), and constructivist learning theory (Mayer, 1996; Mayer, 1999; Wittrock, 1991), but also on a new integrated theory "involving cognitive processing in multiple channels" (Mayer & Moreno, 2002, p. 111), which was based on some research about cognitive processes for different learning channels (Schnotz, Boeckheler, & Grzondziel, 1999; Schnotz, Picard, & Henninger, 1994). Moreover, a subsequent study conducted by Mayer (2011) showed the connections between multimedia learning and educational games and investigated how to develop an effective educational game. In addition, to further support the contextual vocabulary learning theory and multimedia learning theory, Sandberg et al., (2014) developed a mobile game for vocabulary learning. This study applied multimedia objects to create common scenes in daily life, which integrated contextual and multimedia learning. The researchers indeed found that the students had a better

learning performance when using mobile applications with game characteristics, which supported the significance of both theories.

In sum, a range of theories have been used in the current literature to either guide the design of digital games or explain the vocabulary learning results. Some are adapted from the theories that are originally proposed for language learning, and some from those for general education, most of which are not specifically proposed for digital game-based vocabulary learning.

2.5.2 Research Trends of Digital Game-Based Vocabulary Learning

As vocabulary is a crucial element of language, and vocabulary knowledge is regarded as the foundation of language acquisition, learners attribute great importance to vocabulary learning, and have strong interest in effective vocabulary learning methods (Hu & Nassaji, 2016). Various factors that play important roles in successful vocabulary learning have been identified, including but not limited to, attention toward, attempts to retrieve target information, generative use of words, and the amount of time spent engaging with the words (Schmitt, 2008; Nation & Webb, 2011). Many of these factors are induced when learners conduct explicit learning that involves conscious elaboration and deliberate mental processing of target information; however, this approach has the drawbacks of being laborious and having a high demand of focused concentration (Schmitt, 2000; Nation, 2001). In addition, educators have emphasized the importance of situating learners in the contexts of using vocabulary (Sun & Dong, 2004). To complement explicit learning with incidental learning in an enjoyable and contextualized learning mode, game-based vocabulary learning emerged (Moreno-Ger, Burgos, Martinez-Ortiz, Sierra, & Fernandez-Manjon, 2008).

With the quick and massive expansion of digital educational games in various contexts, digital game-based vocabulary learning has become a recognized field, and many studies have been conducted to investigate the effectiveness of digital games in promoting vocabulary learning and to explore factors that tend to facilitate vocabulary learning in game environments (Zou et al., 2019). Zou et al. found 10 types of digital games for vocabulary learning and generally positive effects of them on short-term and long-term vocabulary learning. There have also been a number of review studies on technology-enhanced and digital game-based language learning, however, none of which focuses specifically on digital game-based vocabulary learning in interactive learning environments. Also focusing on digital game-based vocabulary learning, Chiu, Kao, and Reynolds (2012) conducted a meta-analysis of 14 relevant studies and reported the significantly effectiveness of digital games for vocabulary knowledge development. Kucirkova (2017) reviewed interactive digital books only, and Peterson (2010) concentrated on the use of massively multiplayer online role-playing games in

language learning, so other games were not analysed. Though Hung, Yang, Hwang, Chu, and Wang (2018) conducted a scoping overview of various digital games in language education, only a small portion of their analysis was relevant to vocabulary learning. Both Tsai and Tsai's (2018) and Chen et al.'s (2018) meta-analyses focused on the effectiveness of digital games and had little mention of other aspects such as game features, players, and key factors that lead to successful retention. Moreover, Tsai and Tsai (2018) did not include games for first language vocabulary learning, and Chen et al. (2018) reviewed 10 studies only. Therefore, a more focused and thorough review of all existent literature on digital game-based vocabulary learning with in-depth discussions of the digital games, the learners, the learning processes, and the learning outcomes in interactive learning environments is necessary. In response to this call, I conducted a systematic review of digital game-based vocabulary learning and analyzed them from a range of aspects such as the features and devices of the digital games, the players' backgrounds, and learning outcomes.

Data Selection. To understand the research trends and topics in digital game-based vocabulary learning, the following three-step process were employed. The Web of Science Core Collection was used for the primary search, and the Boolean expression was (TS=("vocabulary learning" or "vocabulary learning" or "word acquisition" or "vocabulary acquisition" or "learn word" or "learn vocabulary") AND TS=(game or gaming or gamification) AND TS=(digital or mobile or ubiquitous or computer)). I selected these keywords to maximize the possibility of including all relevant articles on digital game-based vocabulary learning. The literature type was set as "article," the timespan was set as all years, and only one citation index, the Social Science Citation Index (hereafter, SSCI), was ticked for quality assurance. Many review studies, for example, Hung et al. (2018) and Zou et al. (2019), analyzed SSCI publications only to ensure that all reviewed articles were of high quality. At the prior stage of our data selection, I included book chapters, conference papers, and non-SSCI journal articles but found them of unsatisfactory quality due to the absence of important details such as descriptions and explanations of the digital games, learners, experiment setting, and research results, etc. Thus, I finalized the review target as SSCI journal articles, as reliable review results are conditional upon reliable data sources. The search result showed 48 articles. These articles were further checked according to the inclusion criterion that the articles ought to be empirical research on digital game-based vocabulary learning among normal language learners, and the exclusion criteria that review articles, meta-analysis studies, and research on learners with special needs ought to be excluded. Accordingly, 19 articles were excluded from the review list, including (1) two review articles, (2) two meta-analysis studies, (3) four studies on learners with dyslexia, autism, or special needs, (4) five

studies on learning with digital books, music, or media, (5) two research on game design, and (6) four non-SSCI articles. 29 articles were included for analyzing the research trends and topics.

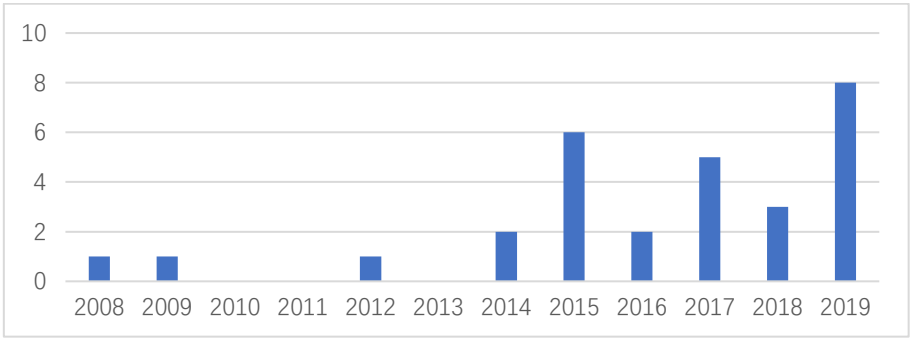


Figure 1: Publication Years

General Publication Situation. Concerning the general publication situation, the number of publications of empirical research on digital game-based vocabulary learning among normal language learners, as presented in Figure 1, has been increasing in the past decade. A total of 16 out of 29 (over 55%) studies was published in the past three years. These studies were published in 17 different journals, with more than one published in the *Journal of Educational Technology & Society*, *Computers & Education*, *Journal of Educational Computing Research*, *Computer Assisted Language Learning*, *British Journal of Educational Technology*, *Computers in Human Behavior*, and *ReCALL* (see Figure 2).

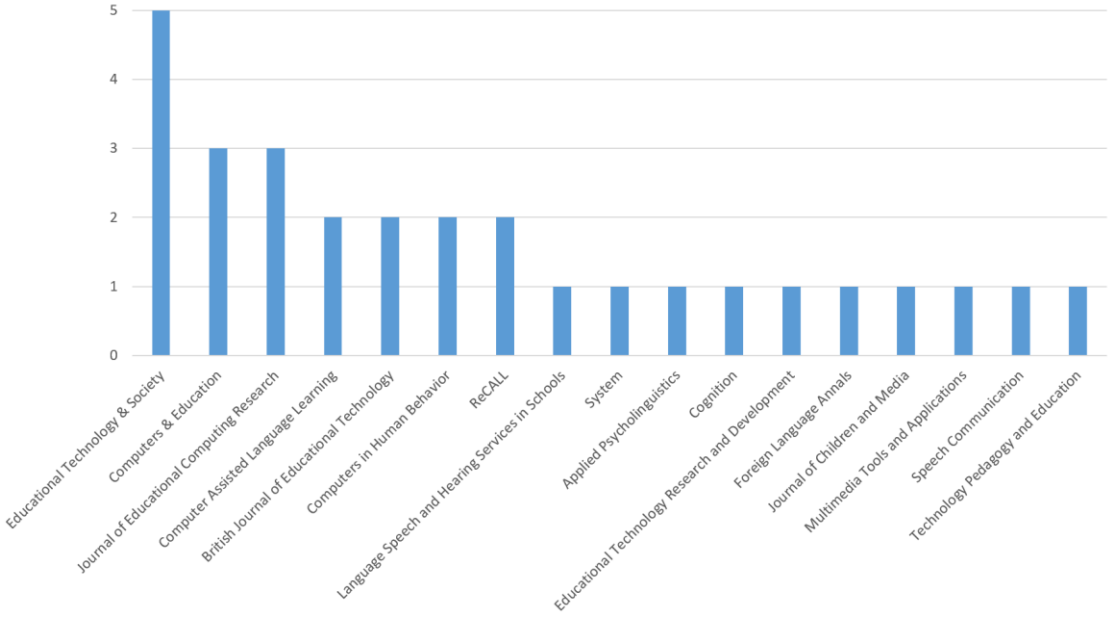


Figure 2: Journal titles and numbers of publications in the journals

Research Methods. Overall, a total of 25 studies applied experimental or quasi-experimental treatment to investigate digital game-based vocabulary learning. Pre- and post-tests were common means of evaluation of the effectiveness of digital games in promoting vocabulary learning. Approximately 90% of studies (n = 26) measured the participants' short-term vocabulary learning outcomes through immediate post-tests (e.g., Franciosi, 2017; McGraw et al., 2009; Mohsen, 2016; etc.). Six also conducted delayed post-tests to evaluate the participants' long-term retention of the target vocabulary knowledge (e.g., Calvo-Ferrer, 2017; Huang & Huang, 2015; McGregor, Marshall, Julian, & Oleson, 2019; Pan, 2017; Wei et al., 2018; Young & Wang, 2014). The number of studies on long-term effects of digital game-based vocabulary learning was smaller than that of studies on short-term effects because long-term research is more demanding and involves more distractors (Nation, 2001). It has higher demands of participant commitment and needs additional resources such as funding and human resources to maintain the implementation of long-term projects. Moreover, the games ought to be fun so that students are willing to play them for a long period of time. Research on the short-term effects of digital game-based vocabulary learning is limited as it is common for students to lose the vocabulary knowledge that they have learned before, and it is possible that some games promote very effective immediate vocabulary learning, while students forget the words they have learned quickly in a short period of time. It is therefore necessary for students to keep playing the game so that their short-term memory of the words can be transformed into long-term memory (Masoura & Gathercole, 2005).

It is noteworthy that some researchers also evaluated the participants' productive vocabulary knowledge in the post-tests, in addition to the receptive knowledge (e.g., Dore et al., 2019; Hao, Lee, Chen, & Sim, 2019; Sandberg et al., 2014; Smeets & Bus, 2015; Sundqvist & Wikstrom, 2015). Nevertheless, the learning outcomes concerning productive knowledge development were not as positive as those concerning receptive knowledge, which is likely because much more practices are necessary for the development of productive knowledge (Nation, 2001), yet most participants in studies on DGBVL spent limited time on game.

Questionnaire surveys were also frequently applied in research on digital game-based vocabulary learning to investigate different learner-related aspects. The foci of the questionnaires used in game-based vocabulary learning research and the representative studies are summarized in Table 7.

Research issue	Representative studies
Learner motivation	Calvo-Ferrer, 2017; Huang & Huang, 2015; Li, Meng, Tian, Zhang,

	& Xiao, 2019; Young & Wang, 2014
Learner attitude	Chen et al., 2019; Hao et al., 2019; Ranalli, 2008; Mueller et al., 2018; Young & Wang, 2014
Learning anxiety	Hong, Tai, & Ye, 2019; Wei et al., 2018
Learners' flow experience	Hong et al., 2019; Hung et al., 2015; Li et al., 2019
Learners' self-efficacy	Hong et al., 2019
Learners' self-esteem	Hong et al., 2019
Learners' cognitive load	Hwang & Wang, 2016

Table 7: Foci of the questionnaires used in game-based vocabulary learning research

Interviews, as another important research method, were used in studies of Chen, Liu, & Huang (2019), Hao et al. (2019), Hung et al. (2015), and Hwang & Wang (2016). Additionally, several studies involve observation as a means of data collection (e.g., Chen et al., 2019; Hung et al., 2015; Sundqvist & Wikstrom, 2015).

Game Types. According to the game taxonomies of Dondi and Moretti (2007) and Connolly et al. (2012), as well as our knowledge of GBVL, I coded the game types with nine categories.

(1) Simulation games are the games that imitate a real life or imaginary situation where players adopt different strategies to solve problems in a game environment (Reiners & Wood, 2015; Khenissi et al., 2016).

(2) In role-playing games, players assume the roles of characters and take responsibility for acting out these roles to complete certain missions in an immersive fictional setting (Crawford, 2003; Cornillie, Thorne, & Desmet, 2012).

(3) Card games involve using playing cards as the primary device for players to take turns to play (Tsai, Tsai, & Lin, 2015; McGraw et al., 2009).

(4) Puzzle games usually require players to solve a puzzle (e.g., identify a logic problem or navigate complex locations such as mazes) for advancement, and they typically involve a series of sequentially presented stages which are related in that players need to benefit from each achieved step to solve the following one (Wolf, 2001; Khenissi et al., 2016).

(5) Exergames require players' physical movements for task completion, so they rely on technology that tracks body movement or reaction (Godwin-Jones, 2016). AR and VR technologies are usually involved, the former of which offers a semi-transparent vision blended with reality while the latter blocks out users' actual environments by replacing a self-contained one (Godwin-Jones,

2016).

(6) Tutorial games are simple games that are particularly designed for educational purposes, and they are normally at the beginning point of a game level coordinate (Criswell, 2009).

(7) Gamification refers to a process of transforming non-game learning environments into more game-like environments and applying game design and mechanics to strengthen non-game activities by adding participation, competition, collaboration, and engagement; it may not necessarily be a newly developed game (Deterding et al., 2011; Werbach, 2014).

(8) Games of hybrid genre refer to any games that involve more than one of the above game types or features.

A summary of the game types is presented in Table 8. To ensure that I cover all possible game types, I also include (9) Others, which are defined as games that are not in the range of the above types.

Simulation games	Games that imitate a real life or imaginary situation where players adopt different strategies to solve problems.
Role-playing games	Games where players assume the roles of characters and take responsibility for acting out these roles to complete certain missions in an immersive fictional setting.
Card games	Games with cards as the primary device for players to take turns to play.
Puzzle games	Games where players solve puzzles.
Exergames	Games that require players' physical movements for task completion.
Tutorial games	Games that are particularly designed for educational purposes.
Gamification	A process of transforming non-game learning environments into more game-like environments and applying game design and mechanics to strengthen non-game activities.

Table 8: Game types

Concerning the popularity of vocabulary learning games that were investigated in the literature, puzzle games, tutorial games, gamification and hybrid genre were more popular than simulation games, role-playing games and card games. And only a very small number of Exergames were investigated. Moreover, compared to the early years (i.e., first half of the review period, which is from 2008 to 2013), three new types of games emerged in the recent years (i.e., second half of the review period, which is from 2014 to 2019), namely simulation games, Exergames and gamification, showing that research on digital game-based vocabulary learning had extended to a wider range of game types. Additionally,

the growth rates of tutorial games, role-playing games, card games and hybrid genre from the first to the second period were respectively 250%, 200%, 150% and 100%. Such tremendous increases demonstrated significant developments in the field in terms of the number of games and reflected that vocabulary acquisition was transforming from tedious repetitive practice to enjoyable digital game-based vocabulary learning.

Game Features. Based on Dondi and Moretti's (2007) analysis and Fu, Su, and Yu's (2009) list of game factors, together with our knowledge of the relevant literature, the major features of vocabulary learning games are listed in Table 9. In addition to the general frameworks of Dondi and Moretti (2007) and Fu et al. (2009), the specific features for some categories were adapted from representative research in the corresponding areas. The features concerning personalization had the foundation in Peterson (2010), Butler (2017), Kucirkova (2017) and Hsu (2017); those of multimodality in Kucirkova (2017) and Butler (2017); those of goal clarity in Chou, Hung, and Hung (2014); those of feedback and challenge in Butler (2017); and those of social interaction in Ryu (2013) and Peterson (2010).

Feature	Content
Personalization	Does the game provide a track record for the player to view their learning process? Does the game collect players' personal data (e.g., gender, age, hobbies, etc.) and use them to customize the learning experience? Can players select the types of activities that they want to play? Can players select the types of target words that they want to learn? Can players select the levels of difficulty of the game? Do players have their own avatar?
Multimodality	Does the game involve images of the target words? Does the game involve audio of the target words? Does the game involve video of the target words? Does the game involve hyperlinks to other explanatory content? Does the game involve AR or VR features? Does the game provide clues that aim to help players find out the answers to questions or solutions to problems?
Goal clarity	Are students given in-person briefing or guidance on how to play the game?
Feedback	Does the game provide feedback on players' performance in

	the game?
Social interaction	Can players talk to other players in the game? Can players send messages to other players in the game? Can players compete with other players in the game? Can players collaborate with other players to complete a task together in the game?
Challenge	Are there speed or time limitations? Do players need to overcome different types or levels of challenges in the game?
Immersion	Do players feel that they are involved in the game?

Table 9: Game features

Among various features, personalization and multimodality were the two most popular, a possible reason of which is that these two features can be realized through wider ranges of approaches than features such as goal clarity, feedback, and challenges (Sundqvist & Wikstrom, 2015; Butler, 2015, 2017; Chen & Lee, 2018). Personalization, for example, can be achieved by providing students with track records so that they can view their learning processes, by letting students decide the levels of difficulty of the games, choosing the types of activities that they want to play and selecting their own avatars, as well as by customizing students' learning experiences according to their preferences and language skills (Peterson, 2010; Butler, 2017; Kucirkova, 2017; Hsu, 2017). There are also various formats and modes of multimedia, including but not limited to images, audio, video, hyperlinks and AR or VR features. Another possible reason for the popularity of personalization and multimodality is the wide acknowledgement of their effectiveness in promoting successful learning. Researchers like Türk and Ercetin (2014), Mohsen (2016) and Boers, Warren, Grimshaw, and Siyanova-Chanturia (2017) all argued that multimodality was conducive to vocabulary learning. Studies such as Huang, Huang, Huang, and Lin (2012) and Huang, Liang, Su, and Chen (2012) all showed empirical support for the contribution of personalization to effective vocabulary learning.

Specifically, concerning the detailed use of multimodal elements in games, images and audio were the most frequently used, followed by clues and videos. This is perhaps because it was easier to integrate images and audio in games, while requiring more time and effort to develop clues and videos. Hyperlinks and AR and VR features were newly emerged elements of the games in the second period, implying important new trends beyond traditional multimodality for vocabulary learning games.

Among various features, personalization is the one with the biggest increase, showing that game designers were aware of its great roles in increasing learner engagement and motivation. Concerning the detailed application of personalized learning features in games, activity selection increased from zero in the early years to 13 in the recent years, indicating a greater diversity of game features with the advancement of educational technology in recent years. The players could not only decide what target words to learn and what difficulty levels to select but also the types of activities that they wanted to play. Moreover, only one game in the early years provided learners with personal avatars, while 14 in the recent years empowered players to select and design avatars. The growth rate was huge. Overall, the increasingly more varied options and sophisticated personalization elements are likely associated with the great development of educational technologies and continuous efforts of educators and researchers.

Following personalization and multimodality, goal clarity, feedback, and social interaction, as three essential elements for educational games, were also commonly integrated in vocabulary learning games. These features tended to provide learners with enjoyable, effective, and interactive learning environments (Tsai & Tsai, 2018; Chen et al., 2018; Kucirkova, 2017). Specifically, concerning social interaction features, more attempts which aim to provide more interactive learning environments and empower players with more interaction opportunities appear necessary for future game development, given that the importance of discussions, competition and collaboration for successful vocabulary learning is widely acknowledged (e.g., de la Fuente, 2002; Barcroft, 2004; Ellis, 2006). The numbers of challenge and immersion features, however, were smaller, indicating less mature developments in these aspects. This probably resulted from the fact that it was time-consuming to embed elements that induce challenge and immersion in the games. These two features however play important roles in leading to successful and enjoyable vocabulary learning (Sweetser & Wyeth, 2005; Ke & Abras, 2013; Butler, 2015, 2017); future games are therefore advised to integrate more of them.

Game Devices. Concerning the devices for digital game-based vocabulary learning, over half of the studies adopted computers, possible reasons of which include that computers are easy-to-access, and it is less time-consuming to develop games that are compatible with computers. Moreover, the proportion of the studies that did not adopt any digital devices was very small, showing a preference for digital over traditional games. Additionally, no research involved the use of tablets or smartphones in the early years, while over one fourth of the studies in the recent years used tablets or smartphones, indicating the start of an era for ubiquitous game-based vocabulary learning. The fast advancement

and popularity of mobile technologies also led to the development of vocabulary learning games towards the direction of mobile learning (Tsai & Tsai, 2018; Chen, Tseng, & Hsiao, 2018). Such a development trend is promising as mobile assisted language learning is advantageous in promoting effective learning, increasing motivation, satisfaction, confidence, and authenticity (Kukulska-Hulme & Viberg, 2018), which play important roles in vocabulary learning. One study has been conducted on the use of wearable devices for vocabulary learning in the second period, which is perhaps because games that are compatible with wearable devices need to be built from scratch and demand up-to-date IT skills and knowledge from educational researchers (Bower & Sturman, 2015). Other possible reasons for the small number of this type of games include the lack of mature software for the development of the games and the knowledge gap between IT engineers and language teachers and researchers (Shadiev, Hwang, & Liu, 2018). A summary of the game device and their affordances and limitations is presented in Table 10.

Game devices	Affordances	Limitations
Computers	Easy-to-access. Less time-consuming to develop games that are compatible with computers.	Not as innovative as mobile and wearable devices.
Tablets or smartphones	Support mobile learning.	Time-consuming to develop games that are compatible with mobile devices.
Wearable devices	Innovative. Can collect rich user data.	Lack of mature software for the development of the games. Knowledge gap between IT engineers and language teachers and researchers.

Table 10: Game devices and their affordances and limitations

Learning Elements Integrated in Games. Among various factors that are conducive to effective vocabulary learning, I focused on retrieval and generation from the learning process dimension, because factors such as motivation, engagement and noticing were closely associated with game features which have been discussed from the game dimension. The general status of the involvement of retrieval and retention in digital game-based vocabulary learning is that retrieval was much more widely used. Given that the effectiveness of generation in promoting vocabulary learning has been

acknowledged among researchers and educators for years (e.g., Laufer & Hulstijn, 2001; Kim, 2008; Zou, 2017). Moreover, retrieval was not commonly used in the games that were developed in the early years, and it was until the recent years that the integration of this element was hugely increased, indicating a developing awareness of the significance of retrieval among game developers in the second period. However, the type of retrieval being adopted was mainly simple retrieval, yet multiple and spaced retrievals, which were of higher levels, were employed with a rather low frequency. In sum, vocabulary learning games had been comparatively mature in terms of increasing learners' motivation, engagement and noticing of the target vocabulary, yet aspects such as retrieval and generation were less widely employed.

Retrieval, different from re-studying which simply involves repetition, is the act of accessing and utilizing the knowledge stored in memory (Roediger, 2000). Multiple retrievals play important roles in consolidating memory traces (Karpicke, 2017; Karpicke & Roediger, 2008). Spaced retrievals promote better learning than successive retrievals as distributed training conditions lead to larger amounts of learning than massed conditions (Baddeley, 1997; Rohrer & Pashler, 2007). Generation, or generative use, refers to the encountering or using of words in new contexts that differ from the previous encounters (Nation, 2001). Receptive generation prompts reconceptualization of a learner's knowledge of a word while meeting it in new contexts through listening or reading, while productive generation entails creating an original context for a word through speaking or writing (Nation, 2001). Thus, the latter involves a greater depth of processing and promotes more effective learning (Zou, 2017).

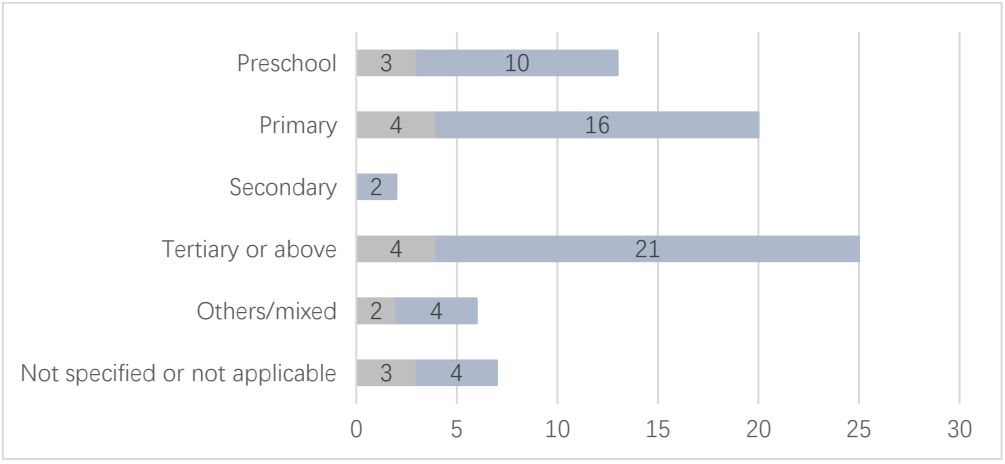


Figure 3: The education levels of the participants in the early years (gray) and the recent years (blue)

Study Participants. I read the methodology sections of the studies carefully, focused on the background information of the participants, and found that almost 35% of the studies were conducted

among tertiary students or above, over 25% among primary students, and almost 20% among preschool students as shown in Figure 3. This is perhaps because researchers had easier access to tertiary students, and primary and preschool teachers and students tended to have a more open mind about games. Moreover, no study investigated games for secondary students in the first period, while two explorations were conducted among this new learner group in the second period, demonstrating certain developments in the field, as well as calling for more investigations. Additionally, the growth rate of studies among primary students was significant in recent years, showing an increasing focus on this learner group. Many researchers argued that games were very helpful for young learners in such aspects as increasing their learning motivation and effectiveness, releasing anxiety, and promoting active and interactive learning (e.g., Kuppens, 2010; Sandberg et al., 2014; Butler, 2015; Hung et al., 2015; Pan, 2017; Chen & Lee, 2018, etc.). Similar benefits of digital game-based vocabulary learning have also been found among students at tertiary or above levels (e.g., McGraw et al., 2009; DeHaan et al., 2010; Huang & Huang, 2015; Verga & Kotz, 2017; Calvo-Ferrer, 2017; Wei et al., 2018, etc.).

The participants' first languages were mainly Chinese, Japanese and English, and their target languages were mainly English. Specifically, based on the information as reported in the methodology of the articles, the first language of almost 30% of the participants was Chinese, and the target language of almost 80% of the participants was English. This is probably because of three reasons. Firstly, English as a lingua franca was the target language of a large proportion of learners (Jenkins, 2007); and secondly, Chinese learners are particularly eager to learn English and regard vocabulary as crucial to successful language learning (Zou, 2016; Zou, 2017). Moreover, over 30% of the reviewed articles were published by Taiwan researchers and 10% by Japanese researchers, indicating that scholars from these two areas were the leading players in the field.

Learning Outcomes. Very positive results were reported by research on digital game-based vocabulary learning, including improved language-related knowledge (e.g., vocabulary knowledge, reading proficiency levels, etc.), language learning-related knowledge (e.g., vocabulary learning strategies and methods, etc.), and affective state (e.g., motivation, confidence, etc.). Among all studies, only one study (i.e., Young & Wang, 2014) reported that digital game-based vocabulary learning was not more effective than other approaches in promoting long-term retention of vocabulary meaning. Most studies observed greater effectiveness of digital game-based vocabulary learning than others in leading to successful vocabulary learning (e.g., Sandberg et al., 2014; Huang & Huang, 2015; Hung et al., 2015; Hwang & Wang, 2016; Franciosi, 2017; Wei et al., 2018, etc.).

From the perspective of target vocabulary knowledge aspects, meaning was the most frequently investigated (slightly below 35%), followed by form: pronunciation (over 20%) and spelling (slightly below 20%). This is likely because it is easy to integrate word meanings and forms in games, and there are diverse ways to enhance knowledge of form-meaning links (Nation, 2001). Vocabulary use, however, demands more sophisticated game design, and thus was comparatively less examined (slightly below 15%). But its growth rate from the early years to the recent years (over 450%) was greater than that of the average (approximately 300%), indicating that studies on digital game-based vocabulary learning were attending to this aspect. Possible reasons of this include researchers' increased realization of the importance of productive vocabulary knowledge and the development of sophisticated commercial vocabulary learning games (Wei et al., 2018; Yu, 2018).

To sum up, the above results show the following research trends and future development directions:

(1) Research on digital game-based vocabulary learning mainly adopted quantitative methods but attempts to investigate this field via qualitative approaches were growing. Most studies were short term, so research on the long-term effects of digital game-based vocabulary learning is necessary.

(2) Most games were for tertiary or above and primary or below students; there is thus a call for games for secondary school students. The games were mainly for English language learning, and the majority were custom-built by researchers. Computers were the most common device for game playing, but the number of games that are compatible with tablets or smartphones was growing fast.

(3) Simulation, role-playing and puzzle games were the three most common types of games, the growth rates of which were also the biggest. The games were developing in the direction of refinement and sophistication, providing players with a greater number and a wider range of game features. Moreover, the games were comparatively mature in terms of increasing learners' motivation, engagement and noticing of the target vocabulary, yet aspects such as retrieval and generation were less widely employed, so future game design and research are advised to focus more on these two aspects.

(4) Most games focused on the development of language-related knowledge, and some on affective states. A small number of studies investigated language learning-related knowledge, indicating a need for research in this direction.

(5) The target knowledge aspects of vocabulary learning games were mostly meaning and form (pronunciation and spelling); more research on the learning of word usage is necessary.

Moreover, vocabulary knowledge involves various aspects such as receptive and productive meaning, form and use, and current vocabulary learning applications concentrate mainly on receptive knowledge of word meaning and form. As deep levels of processing and elaboration on vocabulary are conducive to the learning of word usage, it is suggested that future games integrate more elements that induce deep processing and elaboration (Zou, 2016; Zou 2017). Speaking and writing are typical learning activities that induce deep processing and effective learning of word usage (Kim, 2008), and it is recommended that vocabulary learning games provide more opportunities for students' speaking or writing practice in the formats of collaboration or competition within games through audio or video recording or text inputting. These results indicated that the current research on digital game-based vocabulary learning is limited as they mostly investigated games for intentional vocabulary learning, rather than incidental vocabulary learning. Considering the importance of contextualized and situated vocabulary learning, it is suggested that developers of vocabulary learning games focus more on designing meaningful and facilitative contexts for learning (Kinginger & Wu, 2018). Contexts can also increase students' learning motivation, which is essential for effective learning (Lin, Hwang, Fu, & Chen, 2018), so it is of significance to develop games that provide contextualized incidental vocabulary learning. In response to this call, I attempted to enhance students' vocabulary learning while playing a digital role-playing game in this study. I managed to situate the learning of target vocabulary in a digital game environment and designed the game settings in a way that aligned with the learning context. So, when students are asked to complete the digital game-based learning tasks, they are immersed in the gameplay and incidentally learn the target vocabulary in contexts as presented by the digital role-playing game, and their main foci is not decontextualized intentional vocabulary learning.

To achieve these objectives, it is necessary that different parties, including game designers, technicians, researchers, language teachers and learners, etc., collaborate on the game development. The use of vocabulary learning theories (e.g., the involvement load hypothesis, the technique feature analysis, etc.) as the theoretical frameworks of the games is particularly important, as e-learning systems based on comprehensive linguistic frameworks tend to promote better vocabulary learning than those without strong linguistic support (Hu & Nassaji, 2016; Zou & Xie, 2018). Additionally, the integration of innovative educational technologies plays a paramount role in contributing to the

development of successful vocabulary learning games, so it is advised that more use of innovative technologies and involvement of technicians is indispensable for the advancement of vocabulary learning games. With interesting and effective vocabulary learning games, students tend to be more willing to spend more time playing the games for a longer period; it is thus suggested that more research ought to be conducted on the long-term effects of digital game-based vocabulary learning. Therefore, I investigated students' retention of the target vocabulary knowledge in this study by conducting a delayed post-test one week after the learning treatment. Many studies on task-based vocabulary learning in conventional learning environments (e.g., Folse, 2006; Laufer & Hulstijn, 2001; Keating, 2008; Zou, 2016, 2017) also applied this method of conducting delayed post-tests to measure the long-term effects of vocabulary learning tasks on students' vocabulary learning performance.

2.6 Summary

In this chapter, I firstly reviewed in Section 2.2 what vocabulary knowledge and vocabulary learning are, what aspects of knowledge are included, and what the main approaches to vocabulary learning are. I then summarized in Section 2.3 the most popular theories, hypothesis, and arguments about vocabulary learning, including (1) psycholinguistic factors such as motivation, noticing, frequency of exposure, and generative use, (2) the levels of processing theory, (3) the involvement load hypothesis, and (4) the checklist of task feature analysis. In Section 2.4, I reviewed the development of computer assisted language learning. In Section 2.5, I reviewed theories related to game-based vocabulary learning and the research trends of digital game-based vocabulary learning.

Based on the literature review, I found that lots of theories and hypothesis have been proposed to explain and discuss why some vocabulary learning tasks are more effective than others, for example, the noticing hypothesis, the levels of processing theory, ILH, and TFA. However, most of them are for vocabulary learning in conventional learning environments, while few are specifically proposed for digital game-based vocabulary learning. Thus, in this research, I attempted to investigate whether the theories and hypothesis that were commonly applied to evaluate conventional vocabulary learning tasks could be generalized to examine vocabulary learning in digital game environments. As the ILH is easy to apply and is the most widely adopted theory in the current literature on task-based vocabulary learning, I selected it as the theoretical framework of the study.

I also noticed that although many previous studies on digital game-based vocabulary learning reported positive results concerning the effectiveness of digital games in promoting vocabulary

learning, most of them did not explain why digital games were effective from the theoretical perspective or discuss the theoretical implications of these studies in the field of digital game-based vocabulary learning. So, in this study, I attempted to investigate the effectiveness of digital games in promoting vocabulary learning from the perspective of task-induced involvement load.

Moreover, to fill the gap that many previous studies on digital game-based vocabulary learning did not investigate the long-term effects of digital games on students' vocabulary learning, I conducted a delayed post-test one week after the learning treatment to measure students' retention of the target vocabulary knowledge.

In sum, the current literature on digital game-based vocabulary learning is mainly limited in that (1) few previous studies have been designed with reference to vocabulary learning theories, (2) the theoretical implications of the previous research in the field of digital game-based vocabulary learning have not been thoroughly discussed, and (3) few studies have conducted delayed post-tests to investigate the long-term effects of digital games on vocabulary learning. Thus, I designed this study with ILH as the theoretical framework, attempted to discuss its theoretical implications in the field of digital game-based vocabulary learning, and measured students' retention of the target vocabulary knowledge through a delayed post-test.

My a priori hypotheses are that the ILH, in its present form, may not be able to predict differences between digital and non-digital environments, and learners perform better in a digital game environment than in a conventional one for the same vocabulary learning task with the same involvement load. This hypothesis is based on the previous research results which indicated that digital game environments could motivate vocabulary learning and led to effective learning (see Section 2.5 Digital game-based vocabulary learning for more details). Secondly, I hypothesized that for two different vocabulary learning tasks in the digital game environment, learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form. This hypothesis is based on the previous research results which indicated that the ILH is limited in terms of its uncertainty about the relative weight of the components, and tasks with the same involvement load but different distribution of the components are not equally effective (see Section 2.3.3 Involvement load hypothesis for more details).

CHAPTER 3: METHODOLOGY

3.1 Introduction

As discussed in Section 1.2, I aim to investigate whether the ILH, a hypothesis which was developed by Hulstjn and Laufer (2001) for estimating task-induced involvement loads in conventional incidental vocabulary learning environments, can be applied to the tasks in digital game environments. This research objective is founded on two limitations that I have identified in the current literature. Firstly, it is still challenging to generalize the findings and empirical results attained from conventional learning tasks when measured against the results reached in tasks in digital game environments. Secondly, there has been limited research on how to estimate the involvement loads for learning tasks in digital game environments. Most studies on ILH were conducted in conventional learning environments, while few investigated the application of ILH in digital game environments.

The main research question that guides my study is: Can the ILH, in its present form, predict differences between digital and non-digital environments? It consists of two sub-questions.

(1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one?

(2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form?

The second research question was proposed from the perspective of ILH, which argues that tasks with the same involvement load share similar effectiveness in promoting vocabulary learning. For example, as the two tasks: (1) reading comprehension and looking up target words in a dictionary and (2) reading comprehension and consulting a teacher about knowledge of target words induce the same involvement load, they are estimated to have similar facilitative effects on vocabulary learning. According to ILH, the involvement load of the two tasks is moderate need, search, and no evaluation. Tasks with higher involvement load tend to be more effective than those with lower involvement load, and tasks with the same involvement load tend to be similarly effective in promoting vocabulary learning. It is noteworthy that the ILH is proposed from the dimension of tasks and estimates task effectiveness based on task features, and that learner factors are not the main foci of the ILH, so the estimation of the effectiveness of a task based on its involvement load is from the task dimension, rather than the learner dimension.

In the context of this study, I examined task effectiveness in terms of promoting vocabulary learning from two perspectives, (1) the initial learning of the target vocabulary knowledge, which was evaluated by an immediate posttest which was conducted immediately after the participants completed the learning tasks; and (2) the retention of the target vocabulary knowledge, which was evaluated by a delayed posttest which was conducted one week after the participants completed the learning tasks.

In the following sections of this chapter, I introduce and elaborate on the learning tasks, participants, experimental design, learning tasks, and task effectiveness measurement.

3.2 Task Design

According to ILH, each vocabulary task induces involved loads on three components (i.e., *need, search,* and *evaluation*). Different vocabulary learning tasks can have the same involved loads. For example, the two different vocabulary learning tasks: "Reading a text and answering reading comprehension questions" and "Reading a text and inferring meanings of the underlined words", induce the same involvement loads (i.e., 1+1+1) for *need, search*, and *evaluation* (Hu & Nassaji, 2016).

Specifically, in the reading comprehension task, learners read and understand the text where the target words are highlighted and then answer several multiple-choice questions that focus on the comprehension of the sentences containing the target word. In other words, learners ought to understand the meanings of the target words to be able to answer the multiple-choice questions correctively. So, learners' need is moderate as it is imposed by the task requirement; search is induced as learners need to infer meanings of the target words to understand the text; and moderate evaluation is involved as learners encounter the target words being used in new contexts by reading when they complete the multiple-choice questions. Thus, the involvement load of this task is 1+1+1, which denotes "*Need* at level 1 (moderate), *Search* at level 1 (moderate), and *Evaluation* at level 1 (moderate)" (Hu & Nassaji, 2016).

Concerning the reading and meaning inferencing task, learners read and guess the meanings of the target words based on the contexts, so their need is moderate as the learning of the target words is imposed by the task requirement, and search is induced as meaning seeking is involved. Moderate evaluation is also involved as learners are guided to compare their inferred meanings of the target words to the provided multiple choices based on the contexts. Thus, the involvement load of this task

is also 1+1+1 (Hu & Nassaji, 2016).

As these two tasks induce the same involvement load, they tend to promote vocabulary learning at similar levels of effectiveness, according to the ILH (Laufer & Hulstijn, 2001). Many empirical studies have also provided evidence of support concerning that tasks with the same involvement load are of similar effectiveness in terms of promoting vocabulary learning in conventional exercises performed on paper (Hulstijn & Laufer, 2001; Yaqubi, Rayati, & Allemzade Gorgi, 2012; Hu & Nassaji, 2016). However, it is still unknown whether this finding can be applied to game-based environments for these two tasks. Thus, I focus on these two vocabulary-learning tasks in this study.

In summary, the rationale for selecting these two tasks include four aspects.

(1) *Same Involvement Loads.* To eliminate the potential impact from different sums or different combinations of the involvement loads from three components in the ILH, it is necessary to select two learning tasks with the same involvement loads in all three different components (i.e., *need, search*, and *evaluation*). Currently, the levels of involvement load for the two selected learning tasks in these three components are the same (i.e., moderate levels).

(2) *Balanced Involvement Loads.* These two tasks have balanced involvement loads in three components (i.e., *need, search*, and *evaluation*) in the ILH. If a learning task does not have these two characteristics, the task will have unbalanced involvement loads in these three components. For example, if the task has involvement load of either (1+0+1) or (1+2+1), the higher/lower involvement load for a specific component, i.e., *search* in this case, can create a potential bias which would lead to different learning performances. Thus, the selection of these two tasks can also eliminate the possible bias caused by unbalanced involvement loads from a certain component (Zou, 2017)

(3) *Minimal Differences*. There are minimal differences between the two selected tasks. The same reading text and format of target words can be used in the two tasks. If I select two significantly different learning tasks, such as dictionary-induced vocabulary learning and inferencing with the same involvement loads (Zou, 2016), there would be two potential problems. One is that the gamification of these two tasks becomes a challenge. First, a built-in dictionary would be needed in the digital game. And secondly, the game environment for these two tasks would be significantly different, e.g., the built-in dictionary would need additional data sources, so the game flow and design would be different. To have a consistent game environment and reduce the cost, keeping minimal differences between the two selected tasks is necessary.

(4) *Easy implementation*. The selected two tasks can be easily implemented both on paper and in the digital game. Some vocabulary learning tasks, including composition-writing and sentence-writing (Zou, 2017), can be easily implemented on paper, whereas it is difficult to adopt and integrate them into game flow.

	Task Description	Environment	Target Words	Learning Materials
Task 1	Reading a text and answering reading comprehension questions	paper	same ten target words	same reading text
Task 2	Reading a text and answering reading comprehension questions	digital game	same ten target words	same reading text
Task 3	Reading a text and inferring meanings of the underlined words	digital game	same ten target words	same reading text

Table 11: Details of the three tasks

After finalizing the types of the selected tasks, the format of the tasks was decided as follows based on the research questions and research objectives. Three tasks were investigated (see Table 11). Task 1: Reading a text and answering reading comprehension questions is paper based. Task 2: Reading a text and answering reading comprehension questions is digital game based. Task 3: Reading a text and inferring meanings of the underlined words is also digital game based. To minimize the differences between Task 2 and Task 3, the same target words and reading text were used. According to the ILH, these three tasks have the same ILH loads (i.e., 1+1+1) in three components "need," "search," and "evaluation" (Hu & Nassaji, 2016), and it is estimated that they are similarly effective in terms of promoting vocabulary learning (Gu, 2003; Webb, 2007; Milton, 2008; Laufer, & Rozovski-Roitblat, 2011). More details of the reading text and the target vocabulary are presented in Section 3.6 Learning materials.

3.3 Research Design

To answer the two research questions, which are (1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one? And (2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form? The experiments were designed following the framework illustrated in Figure 4. Participants were divided into three groups (i.e., Group A, B, and C). Each group of participants was required to complete one vocabulary learning tasks (i.e., Task 1, 2, or 3). More details of the participants are presented in Section 3.4 Participants. The details of the group division and task description are introduced as follows.

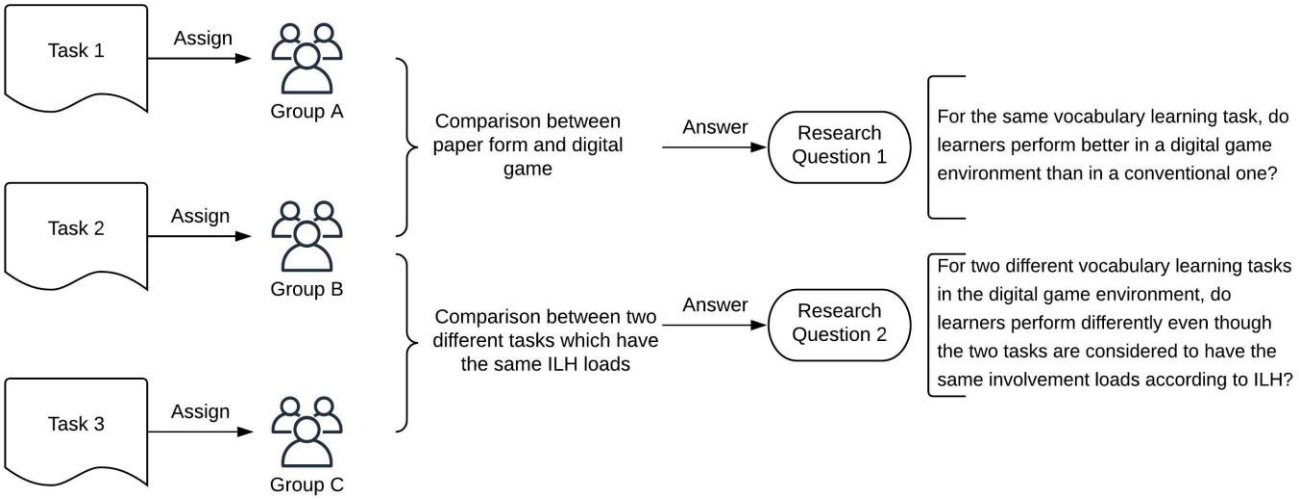


Figure 4: *The research framework of this study*

Essentially, there were two pairs of comparisons among the three groups of participants. The first one was the comparison between Group A and Group B. This comparison aimed to answer the first research question about whether for the same vocabulary learning task, learners perform better in a digital game environment than in a conventional one. Therefore, the only experimental setting difference between these two groups was the game environment. In other words, I employed the conventional learning environment (i.e., on paper) in Group A, while the digital game-based learning environment for the same task was adopted in Group B. The second one was the comparison between Group B and Group C. This comparison aimed to answer the second research question about whether for two different vocabulary learning tasks in the digital game environment, learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH. In other words, I used "Reading a text and answering reading comprehension questions" as the

learning task in Group B, while "Reading a text and inferring meanings of the underlined words" was employed in Group C as the learning task. Both tasks were under the same digital game-based environment (see Table 12). More details of the game are presented in Section 3.5 Game design.

	Task 1 (Group A)	Task 2 (Group B)	Task 3 (Group C)
Task Type	Reading a text and answering reading comprehension questions		Reading a text and inferring meanings of the underlined words
Environment	On paper	In a digital game	
Target Word	The same 10 target words		
Learning Material	The same reading text		
Involvement Load	Moderate need, search, moderate evaluation (1+1+1)	Moderate need, search, moderate evaluation (1+1+1)	Moderate need, search, moderate evaluation (1+1+1)
Research Question 1	For the same vocabulary learning task, do learners perform better in a digital game environment than in a conventional one?		[redacted]
Research Question 2	[redacted]	For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH?	

Table 12: The details of three learning tasks and their relationships to the research questions

As illustrated in Figure 4 and Table 12, the three vocabulary learning tasks were assigned to three groups of participants, Group A, Group B, and Group C, respectively. Task 1 required participants to read a text and complete the multiple-choice task on the text on paper. The reading text and target words of Task 2 were the same as those of Task 1; however, participants needed to complete it in a digital game learning environment. For Task 3, participants needed to read the same text, learn the same target words, and choose the most suitable definitions for the target words while doing the

post-reading exercises. Similar to Task 2, participants completed Task 3 in a digital game learning environment. The similarities and differences between the three tasks are presented in Table 12.

In summary, the learning environment was the main difference between Task 1 and 2, whereas the task types were the main difference between Task 2 and 3. Note that Task 2 and Task 3 are of different task types but induce the same involvement load (i.e., 1+1+1) with reference to the ILH. Research question 1 can be answered through a comparison between Task 1 and 2 by using the same task content with different learning environments. Task 2 and 3 addressed question 2 by requiring students to complete different tasks in the same learning environment. Each group completed only one specific vocabulary learning task, illustrated in the first row in Table 7. This experimental design is similar to other language learning studies on ILH (Laufer & Hulstijn, 2001; Laufer & Girsai, 2008).

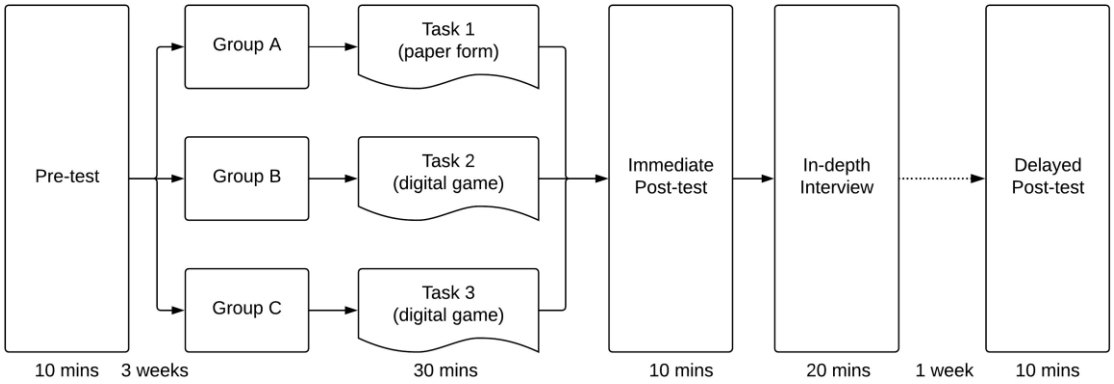


Figure 5: The designed experimental procedures

This experiment consists of five stages: the pre-test, the learning process, the immediate post-test, the in-depth interview, and the delayed post-test, as shown in Figure 5. More details of the pre-test are presented in Section 3.3.1; details of the learning process are presented in Section 3.3.2; details of the immediate post-test are presented in Section 3.3.3; details of the interview are presented in Section 3.3.4; and details of the delayed post-test are presented in Section 3.3.5. As a brief summary, the experimental procedures are elaborated as follows.

3.3.1 Pre-test

A pre-test was conducted to examine the participants' prior knowledge of the ten target words. The measurement tool of the pre-test was the same as those of the immediate posttest and delayed posttest. More details of this measurement tool are presented in Section 3.7. A total of 159 participants were asked to attend the pre-test three weeks before the experiment. The pre-test lasted ten minutes.

During the pre-test, the participants were asked to respond whether they have met these words before. This pre-test design is similar to other vocabulary acquisition studies for foreign language learners (Laufer & Hulstijn, 2001; Laufer & Girsai, 2008). The students who knew more than two target vocabulary were not invited to further participate in the project, and a total of 24 students were excluded at this stage. More details of the participants are presented in Section 3.4.

I used Folse's (2006) modified version of Paribakht and Wesche's (1997) vocabulary knowledge scale as the measurement tool of the pre-test because it could evaluate learners' different levels of vocabulary knowledge, from not knowing the meaning of a word to being able to create an original context using the word (Schmitt, 2000). Several previous studies on the ILH also used this measurement tool to evaluate learners' initial development of knowledge of target words, for example, in Folse (2006) and Zou (2016, 2017). The aspects of vocabulary knowledge evaluated by this modified vocabulary knowledge scale (i.e., the receptive knowledge of word meanings and productive knowledge of meanings and use) are the aspects which are most likely to be learnt by language learners at the initial stage of vocabulary knowledge development, so most previous studies also concentrate on the measurement of these knowledge aspects (Schmitt, 2000; Zou, 2017).

3.3.2 Learning Processes

After completing the pre-test, the selected 135 participants, who knew no more than two target vocabulary, were invited to participate in the learning session. They were randomly and equally assigned to three groups and then asked to complete the vocabulary learning tasks. Groups A, B, and C completed Tasks 1, 2, and 3, respectively. As introduced earlier in Section 3.2 Task design, Group A read a given text and completed the multiple-choice reading comprehension exercises on paper, while Group B completed the same task in digital game format. For Group C, the participants read the same text as Group A and B and inferred meanings of the underlined words while playing the digital game.

The location for the learning was a classroom equipped with more than 30 laptops at The Education University of Hong Kong. The duration of the learning process lasted about 30 minutes. Due to the differences between the three learning tasks, the learning settings for the three groups of participants were arranged as follows. For Group A, the learning task was paper based, and therefore all participants were provided with the pre-printed hard copies of the learning materials, and all copies of learning materials were returned to the research team members to prevent the exposure of the

learning materials to other participants before the experimental procedures. For Group B and C, as the two learning tasks were slightly different, two versions of digital games were developed and installed in the personal computers in the classroom. To make the learning process smooth, several student helpers had already been trained to be familiar with the game operations and provided technical support to the participants using the personal computers and guided the game playing in the classroom when necessary.

I conducted a pilot study before the project experiment to estimate the length of time the participants might need for task completion. Six participants were selected for this pilot study according to the four criteria (i.e., the 1st year postgraduate master students from a local university at Hong Kong; their English proficiency is between International English Language Testing System (IELTS) Band 5.5 to 7.0; they have minimal prior knowledge about target words; and their first language is not English). These participants were asked to complete all tests without a limited time. A timer was used to record their time for completing each test. The average time for the participants' completion of the learning tasks was 27 minutes, and the longest time they needed was 29 minutes. Thus, it seemed that 30 minutes were appropriate for the learning period.

In sum, all learners participated in the experiment in normal class hours in a typical classroom setting. I explained to them the project purposes and general background and obtained their consent to participate in the research prior to the experiment. I also obtained relevant ethical approval from University of Bristol and The Education University of Hong Kong. I was an academic staff member of The Education University of Hong Kong at the time of data collection. All students participated in the project voluntarily and completed the learning in 30 minutes. More details of the participants are presented in Section 3.4 Participants, and details of the research ethics are in Section 3.8 Research ethics.

3.3.3 Immediate Post-test

After completing the 30-minute learning process, a 10-minute immediate post-test was immediately arranged for all three groups of participants. To eliminate the potential impact caused by different test formats, I employed paper-based post-tests for all three groups rather than using a game-based assessment, which could be considered a negative or positive factor for the assessment performance (Mavridis, & Tsiatsos, 2017; Ventura, & Shute, 2013). The post-test was the same as the pre-test, as illustrated in Appendix 4.

The reason for using the same test for pre-test and immediate/delayed post-tests are as follows. First, the effectiveness of word retention should be measured by the differences in performance between the same set of tests (Hulstijn & Laufer, 2001; Newton, 1995; Ulanoff & Pucci, 1999). Second, the inconsistency between the format of tests could lead to the learners' different performance for the same set of tests as found by many studies (Bell & Hay, 1987; Currie, & Chiramanee, 2010).

The post-test duration was strictly controlled at 10 minutes according to the results of the pilot study concerning the needed time for the post-test completion. The average time for the participants' completion of the post-test was 8 minutes, and the longest time they needed was 10 minutes. Therefore, the duration of the post-tests was set as 10 minutes.

This immediate post-test aimed to measure the participants' initial learning of the target vocabulary knowledge, and it was conducted immediately after the participants completed the learning tasks. The details of the measurement tool are presented in Section 3.7 Measurement tool: the modified vocabulary knowledge scale.

3.3.4 In-depth Interview

To collect qualitative data concerning the participants' learning processes, 20 minutes of in-depth interviews were conducted after the immediate post-test. A total of 30 participants, 10 from each group, were selected by random sampling (Robinson, 2014). More details of the participants are presented in Section 3.4. The interview was conducted in a standard classroom at The Education University of Hong Kong. Interviews were conducted in the participants' first language (i.e., Chinese) using retrospective interview protocols because they are considered as the least disturbing method towards participants' cognition (Someren, Barnard, & Sandberg, 1994). The protocols allowed the researcher to ask relevant questions to gather sufficient qualitative data for further analysis (Fehr, Fischbacher & Rosenbladt, 2003). The interview was audio-recorded, and the participants were encouraged to express themselves freely in whatever languages they preferred (i.e., English, Mandarin, Cantonese, or a mixed use of them).

Sample questions concerning the learning tasks include "What do you think about the learning task?" "What did you notice during learning?" and "What did you find useful or useless for vocabulary learning concerning the learning task?". Sample questions concerning the digital game include "What

do you think about the digital game?" "Do you like this way of learning?" and "What did you find useful or useless for vocabulary learning concerning the digital game?".

The semi-structured interview was collected to triangulate the quantitative data collected from the post-tests. The participants were encouraged to express themselves freely and report whatever they thought about the learning tasks, approaches, and processes. They were especially asked to reflect on their learning experiences and what they found useful for their learning of the target vocabulary, what helped them concentrate on the vocabulary learning, and what distracted them from the learning or placed negative influences on their learning experiences. Analysis of these features could contribute to the better understanding of the experimental results concerning the effectiveness of the three tasks.

3.3.5 Delayed Post-test

The delayed post-test aimed to verify long-term word retention after completing the learning task (Schellekens, Sijtsma, Vegter, & Meijman, 2000; Nakata, 2008). Thus, I conducted delayed post-tests one week after the immediate post-test and the interview. One week has been regarded as the standard interval between immediate post-test and delayed post-test in previous research (Hulstijn & Laufer, 2001; Laufer & Girsai, 2008). To minimize any impact caused by environmental factors, a similar test environment (i.e., a classroom) and the same duration (i.e., 10 minutes) were employed for the delayed post-tests. In addition, the same test papers, as shown in Appendix 4, were used.

The interviewees were not post-tested to avoid potential influences of the interview on the participants' performance in the delayed posttest. That is, after the immediate post-test, the students participated in either the interview or the delayed post-test, rather than both. More details of the participants were presented in Section 3.4.

This delayed posttest aimed to evaluate the participants' retention of the target vocabulary knowledge, and it was conducted one week after the participants completed the learning tasks. During this one week, the participants were asked to not attempt to recall any information about the 10 target words or review the target vocabulary knowledge in any way. As the target words were out of the most frequently used 8000 words, according to Davies' (2012) Corpus of Contemporary American English (COCA), it is unlikely that the participants would encounter these words during this one-week period. Thus, it could infer that the main influence on the participants' performance in the delayed posttest was the tasks that they completed during the learning session, rather than other factors. Many empirical

studies on the ILH were also designed in this way to investigate the effects of task types on students' retention of target vocabulary knowledge, for example, Laufer and Hulstijn (2001), Folse (2006), Keating (2008), Kim (2008), and Zou (2012, 2016, 2017).

3.4 Participants

The participants recruited in this study were master's degree students from a Hong Kong university. To ensure that the recruited students' language proficiency was at similar levels, the following methods for selecting candidates was employed:

- All participants had to be students in their first year of a master's degree programme from The Education University of Hong Kong. They had learned English as a foreign language for approximately 12 years on average. This aims to minimize possible influences of the participants' backgrounds on their vocabulary learning performance in this research.
- All participants had to have obtained scores of a Band 5.5 to Band 7.0 from the International English Language Testing System (IELTS), which is a typical entry score for admission to postgraduate programmes of local universities. (If the IELTS scores were not available, the participants' HKDSE English Language Level used a reference, with Level Three of HKDSE scores being accepted). This aims to minimize possible influences of the participants' language proficiency levels on their vocabulary learning performance in this research.
- All participants were asked to take a pre-test. I did not invite the students who knew more than two target words to participate in the experiment but only selected those who had little prior knowledge of the target words. This aims to minimize possible influences of the participants' prior knowledge of the target vocabulary on their vocabulary learning performance in this research.
- The first language (L1) of all participants was either Mandarin or Cantonese but not English. This aims to minimize possible influences of the participants' first languages on their vocabulary learning performance in this research.

In sum, these criteria aimed to minimize the possible differences among the participants. Note that the reason for setting the scores between Band 5.5 and Band 7.0 from the IELTS as a compulsory condition were: (i) first-year postgraduate non-local students have obtained this score or above for the admission to a local university; (ii) some local students in Hong Kong do not have to take IELTS to gain admission into master's degree programmes when the medium of instruction of their Bachelor programme is English. To ensure that the participants without IELTS scores had similar language

proficiency, they were required to provide their HKDSE score, as the Level Three for HKDSE – English is equivalent to a Band 5.5 from the IELTS according to the mapping relationship between HKDSE English Language Levels and Mean IELTS Band Scores, which is based on a data released by the Education Bureau of Hong Kong (HKEAA, 2013). When students have studied for four years in an undergraduate programme, their average scores on the IELTS are 6.76 according to a comprehensive assessment survey of English proficiency by University Grant Council (UGC, 2014).

In addition, taking the pre-test helped to confirm that all participants not only had similar language proficiency but also little prior knowledge of the target words in the experiments. Some studies have revealed that the prior knowledge on the words of participants tends to result in a better performance in word retention (Hulstijn, Hollander, & Greidanus, 1996; Shin, 2010).

Based on the selection criteria, I invited a total of 192 students to participate in this research at the first stage through different channels, including emails, social media platforms, posters, and so on. These participants were mainly (i) 1st year master's-level students from a local university in Hong Kong; (ii) their English proficiency was between IELTS Band 5.5 to 7.0; (iii) they had minimal prior knowledge about the target words; and (iv) their first language was Mandarin or Cantonese, not English. The criteria (i), (ii), and (iv) were checked through a self-reported list before attending the pre-test. For criteria (iii), I implemented the following methods to check whether the participants met it:

- The participants were asked to complete a pre-test and try their best to answer questions about the target vocabulary knowledge. The material for the pre-test is shown in Appendix 4.
- If a participant knew more than two target words, s/he would not be invited to participate in the following parts of the project.

Stages		Number of Participants
Pilot studies (n = 45)	The pilot study on how much time students need for task completion	n = 6 (The details are presented in Section 3.3.2 Learning processes.)
	The pilot study on what digital device should be used for the digital game	n = 10 (The details are presented in Section 3.5.3 Selection of the digital tool for playing the game.)
	The pilot study on the appropriateness of the game	n = 7 (The details are presented in Section 3.5.4 Game flow.)
	The pilot study on the appropriateness of the reading text	n = 8 (The details are presented in Section 3.6.1 The reading text.)
	The pilot study on the appropriateness of the target vocabulary	n = 9 (The details are presented in Section 3.6.2 The target vocabulary.)
	The pilot study on the importance of the target vocabulary	n = 7 (The details are presented in Section 3.6.2 The target vocabulary.)
Experiments (n = 135)	Enrollment stage	n = 192
	Pre-test	n = 159
	Learning session	n = 135
	Immediate Post-test	n = 135
	Interview	n = 30 (10 participants in Group A, 10 participants in Group B, and 10 participants in Group C)
	Delayed Post-test	n = 105 (35 participants in Group A, 35 participants in Group B, and 35 participants in Group C)

Table 13: The number of participants in different experimental stages

As demonstrated in Table 13, among the 192 students who participated in this research at the

first stage, 135 of them met the selection criteria, so they were invited to participate in the experiment. Among these 135 participants, 132 were mainland Chinese, and three were Hong Kong students. These participants were assigned to three groups randomly, with 45 participants in each group. The three Hong Kong students were assigned to the three groups randomly. Group A learned the target vocabulary through completing Task 1: Reading a text and answering reading comprehension questions on paper; Group B learned the target vocabulary through completing Task 2: Reading a text and answering reading comprehension questions in a digital game environment; and Group C learned the target vocabulary through completing Task 3: Reading a text and inferring meanings of the underlined words in a digital game environment. Among the 45 participants in each group, 10 of them were interviewed after the immediate posttest, and the remaining 35 participants were tested again one week later to measure their retention of the target vocabulary. The details of the tasks are presented in Section 3.2 Task design; the details of the research design are presented in Section 3.3; and the details of the pre-test, learning processes, the immediate and delayed post-tests, and the interview are presented in Section 3.3 Research design.

A group size of 35 was selected because (i) the error of t-distribution would remain stable and small when the sample size is 30 or larger (Hogg & Tanis, 2009) and (ii) it is convenient to administer with the sample size of 30 to 40 since the normal classroom size is about 30 to 40 at Hong Kong universities.

3.5 Game Design

In this sub-section, I firstly explain the rationale for my selection of the game development tool and the selection of the digital tool for playing the game. After that, I explain the game flow explicitly with associated screenshots of the digital game.

3.5.1 Selection of the Game Type

I selected digital role-playing games as the digital game type of this study based on the following considerations. Firstly, digital role-playing games could effectively immerse students in digital game environments when they act certain roles in the game to complete game challenges (Zou et al., 2019). Secondly, it is easy to embed learning activities in digital role-playing games, as game designers can place different Non-player Characters (NPC) in the game to guide players to complete various tasks (Hwang & Wang, 2016). The game players can also have conversations with the NPCs. In this way, learning activities in digital role-playing games are normally very interactive, and the

literature indicated that students basically showed positive attitudes towards learning while playing digital role-playing games (Cornillie, Thorne, & Desmet, 2012). Thirdly, digital role-playing games are easy to develop and effective in promoting vocabulary learning (Hwang & Wang, 2016; Sylven & Sundqvist, 2012).

3.5.2 Selection of the Game Development Tool

For Task 2 and Task 3, there were many possible software tools for developing digital games. Using a survey conducted by a professional game design organization called Gamedesigning (2019), I compared a list of the most popular game development tools (Table 14). These tools are GameMaker, Unity, Stencyl, Construct 3, Cocos2D, PlayCanvas, RPGMaker, and MonoGame. To pick the most suitable game development tool on the list, I set some selection criteria. The criteria can be summarized as follows:

- *Dedicated to 2D Game Development*. The reason for selecting 2D games rather than 3D games for our project was that the cost of 3D games is much higher. For example, the cost of developing a popular 2D game called "Flappy Bird" was $US300, whereas the cost of developing a 3D game called "Candy Crush" with a similar style to "Flappy Bird" was about $US100,000 (MeliorGames, 2019). Moreover, if the game development tool is not dedicated to 2D games, the software tool is more complicated and requires more domain knowledge and experience in game development. To reduce the cost and required manpower, the game development tools dedicated to 2D digital games development were selected.
- *Easy Use of Programming Language*. As shown in Table 9, several programming languages like C#, C++, Python and Lua are supported by different game development software tools. A critical factor for selecting these software tools is the easy use of the programming language because the maintenance and revision costs of the digital games are significantly reduced if the programming language can be easily learned and used. In other words, I would not need to seek a professional C++ game developer for further revisions of the digital game if the source code of the game is easy to be maintained.
- *Widely Used in Educational Research*. As digital games are an essential tool in in educational research, it is necessary for the selected game development tool to be widely and maturely used in academic communities. Notably, the game development tool should be used in language or vocabulary learning research.

Software	Price	Platforms	Develop skills	Official Website
GameMaker	Free	- Desktop	Without programming	www.yoyogames.com/s

		- Mobile - Console - Web - UWP - Mobile - Console	(drag-and-drop system) for 2D digital games	tudio
Unity	Conditionally Free	- Desktop - Web - UWP - Mobile - Console - VR/AR	Uses C# for scripting for both 2D and 3D digital games	unity3d.com
Stencyl	Free	- Desktop - Web - Mobile	Uses Haxe programming language and drag-and-drop system	www.stencyl.com
Construct 3	Free	- Desktop - Web - Console	Uses visual programming for 2D digital games	www.scirra.com
Cocos2D	Free	- Desktop - Web - Mobile - Console	Uses various programming languages including C++, Javascript, Lua, Python, and so on for 2D digital games	www.cocos2d.org
PlayCanvas	Free	- Desktop - Web - Mobile	Uses Javascript for 3D digital games	www.playcanvas.com
RPGMaker	Free	- Desktop - Mobile - Console	Uses simple programming languages for 2D digital games	www.rpgmakerweb.com
MonoGame	Free	- Desktop - Mobile - Console	Uses XNA and C# for 2D digital games	www.monogame.net

Table 14: The most popular free game development tools (Gamedesigning, 2019)

Using the above three criteria, I eventually selected the RPGMaker as the game development tool for this study. RPGMaker is a dedicated game development software for 2D games, and there have been thousands of 2D digital games created using RPGMaker (Itch, 2019). Although RPGMaker supports a programming language for game scripting, it is more frequently used as a non-programming interface to implement the internal flow and logic of games without any programming (Whitehead, 2008). Therefore, it is relatively easy to maintain and revise games by using RPGMaker. RPGMaker has also been widely employed to develop digital games in the academic communities of educational technologies. For example, Yang and Chang (2013) employed RPGMaker in 19-week-long biology

and computer programming classes and found that the students gained better performance in terms of academic achievement and critical thinking skills than a control group of students who were using flash games for learning. Hwang and Wang (2016) proposed to compare two different guiding strategies (i.e., single-loop and double-loop learning) for English vocabulary acquisition in situated computer games developed by RPGMaker. From the behavioral patterns identified in this study, the digital games developed by RPGMaker engaged the participants in both single and double-loop learning strategies. Furthermore, several studies (e.g., Jian, Shen, Huang, Chen, & Chen, 2015; Zou et al., 2019) have demonstrated that RPGMaker is an effective game development tool for educational studies.

3.5.3 Selection of the Digital Tool for Playing the Game

Concerning the selection of the digital tool for playing the digital game, I conducted a survey among 10 students who had similar backgrounds as the participants of the experiment. These students were asked to indicate the degrees of their preferences for the devices used for playing the digital game. The survey used a five-point Likert scale (i.e., 5 - "Strongly Preferred," 4 - "Preferred," 3 - "Neutral," 2 - "Not Preferred" and 1 - "Strongly Not Preferred") to denote the degree of preferences from the most to least preferred medium for the experiment. If the students strongly preferred a digital device, they were asked to select 5; and if they strongly dis-preferred a digital device, they were asked to select 1. The participants were asked to give scores for all three types of digital devices, which were computers, tables, and mobile phones.

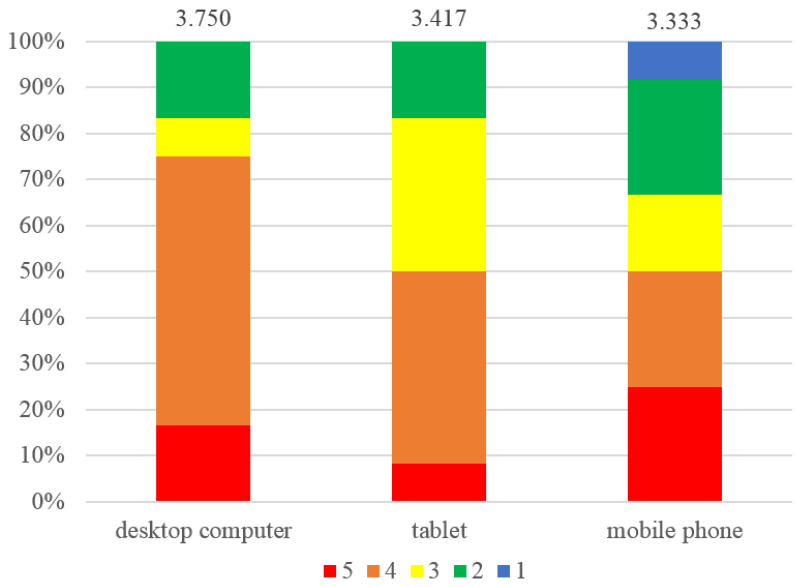


Figure 6: The results of the survey for the preferred tool for playing the digital game

The survey results showed that the students preferred playing the digital game using computers over tablets and mobile phones. There are two potential reasons for these survey results. First, the mobile phone may not be so acceptable for certain text-based learning tasks as they involve high cognitive loads. For example, students prefer not to use mobile phones and tablets for learning tasks involving the review of products and services due to the complexity and necessity of accessing multiple information sources (Calabrich, 2016). Second, the five participants in the pilot study indicated that mobile phones and tablets might not be suitable for this learning task due to the limited screen size for reading the task descriptions and learning materials, whereas key information can easily be located when displayed on computers as the screen sizes were much bigger.

In sum, based on the above considerations and survey results, I decided to use RPGMaker to develop a computer-based digital game for this research. That is, the participants who did Task 2: Reading a text and answering reading comprehension questions and Task 3: Reading a text and inferring meanings of the underlined words learned the target vocabulary through playing a digital role-playing game using computers. The details concerning the selection of the game type are presented in Section 3.5.1.

3.5.4 Game Flow

The selected game development software RPGMaker can be used to develop various storylines and ways for participants to experience digital games. Previous studies have revealed that the flow experience in game-based learning provides a better flow experience and various kinds of cognitive loads compared to non-game learning (Chang et al., 2017; Wang & Chen, 2010). Other studies (Erhel & Jamet, 2013; Erhel & Jamet, 2019) have suggested that other factors, including instructions, student motivation and feedback, can lead to different performances for students with similar prior knowledge in terms of learning outcomes, engagement in the learning process and so on. To minimize the potential impact caused by the above factors, I developed a simple and straightforward game flow with only two scenarios.

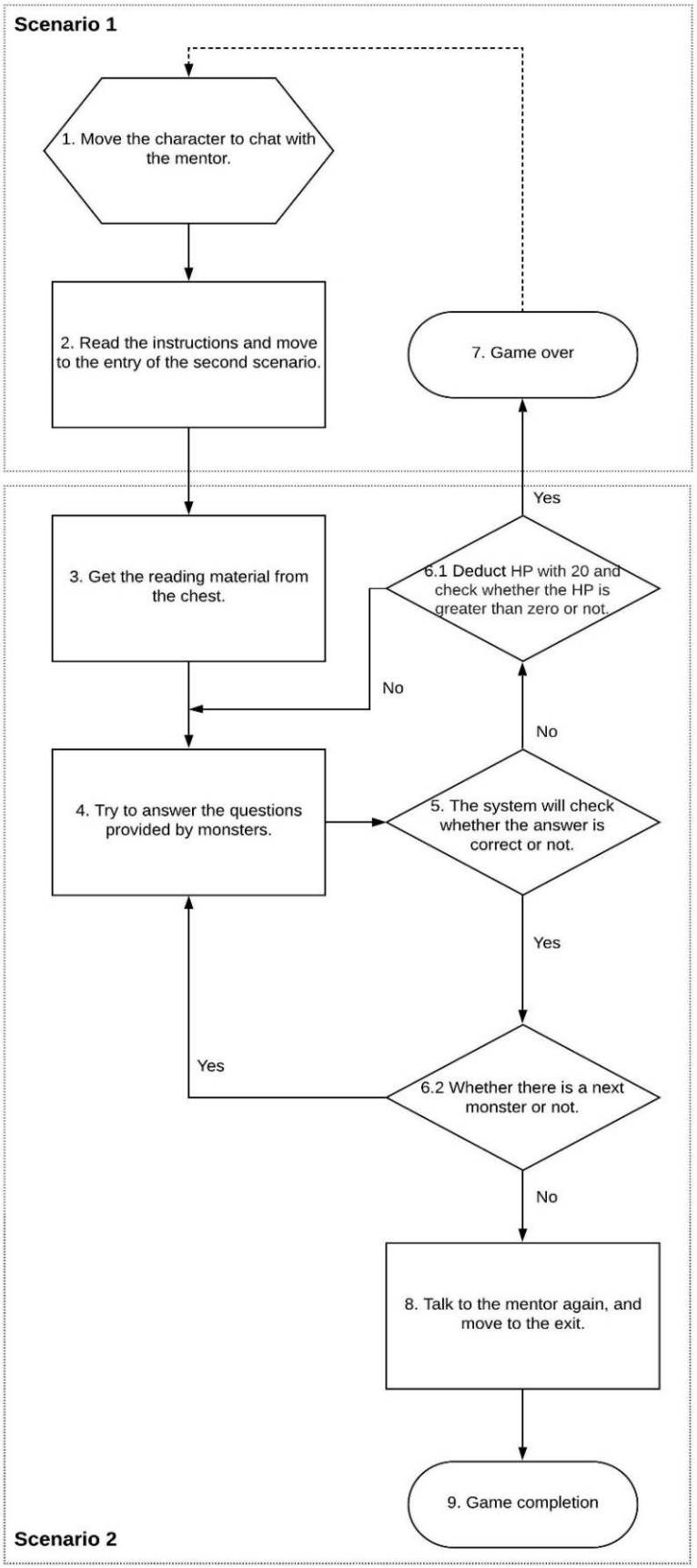


Figure 7: The detailed game flow of the learning task

Figure 8: The game interface of the first scenario

Figure 9: The game interface of the first scenario with the instructions

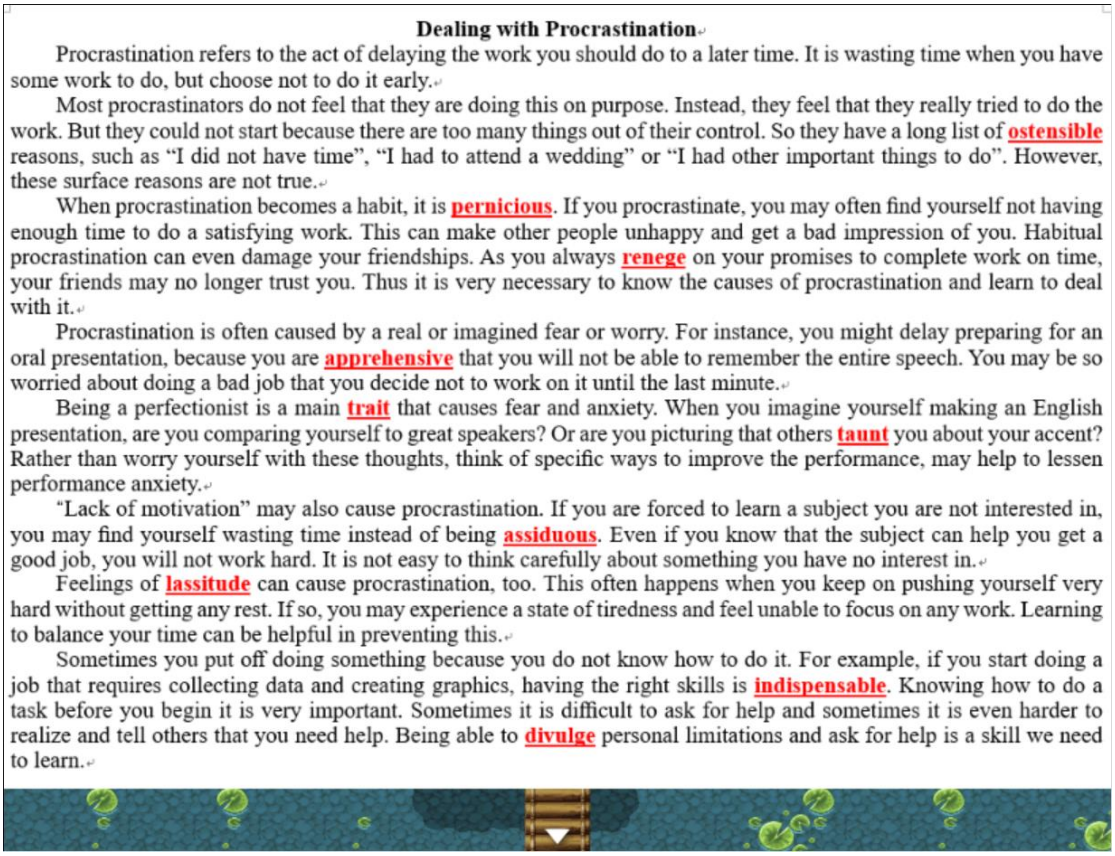


Figure 10: The game interface of the second scenario with the task card

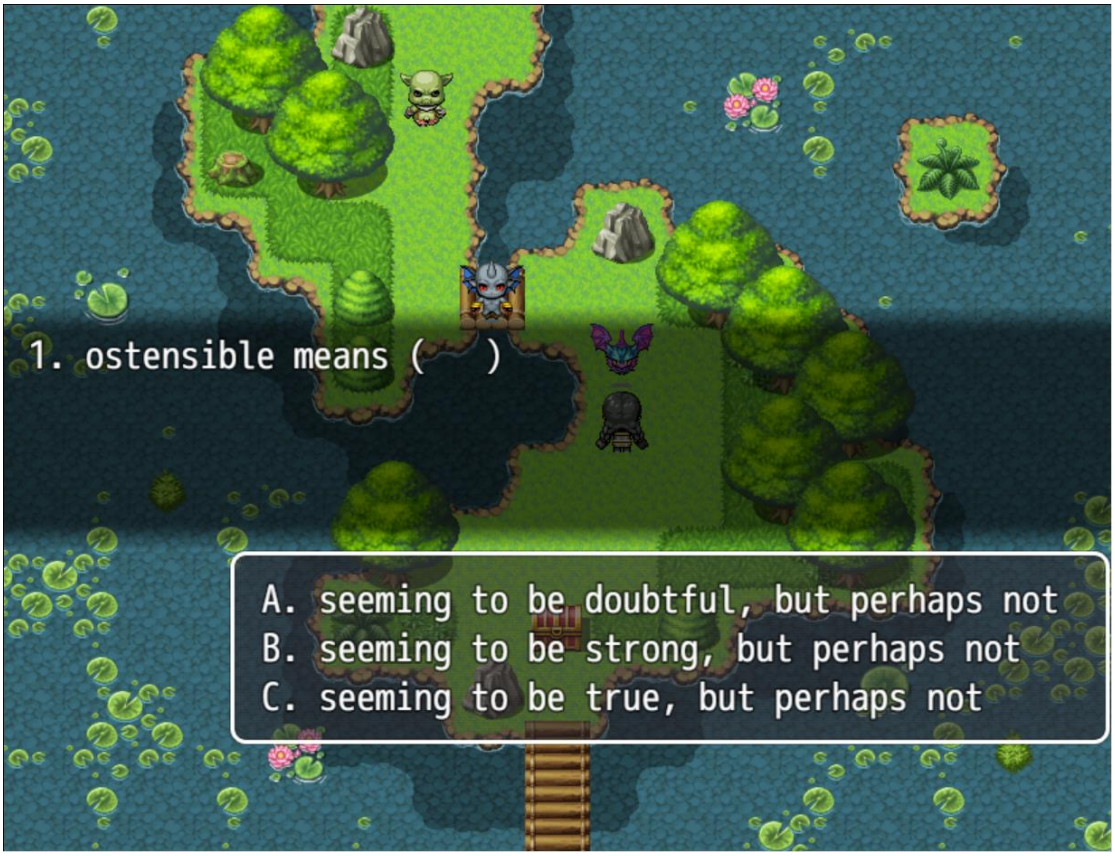


Figure 11: The game interface of the second scenario of question items

Figure 12: The game interface of the second scenario of completion of all questions

The first scenario was the entry point, which provided some instructions, e.g., "Please read the text on the task card and then answer the ten comprehension questions given by the monsters" (Appendix 2), and "Please read the text on the task card and then choose the correct definitions for ten questions given by the monsters" (Appendix 3). The game interface for the first scenario is shown in Figures 8 and 9. The game flow (Figure 7) is summarized as follows:

(1) The participant can control the game character on the hilltop to move and then chat with a Non-player Character (NPC) mentor.

(2) The participant reads the instructions given by the NPC mentor. The instructions contain the main descriptions of the tasks. Then, the participant can move the game character to the entry of the second scenario, as shown in Figure 9.

(3) After entering the second scenario, the game character, controlled by the participant, receives a task card from a treasure chest. Note that this step is set as compulsory in the game. The task card contains the reading materials of the learning task, as shown in Figure 10. Note that the participant can open or close the task card by using the menu or the keyboard at any time during the game.

(4) The game character encounters 10 monsters on the path to completing the mission. Each

monster gives a multiple-choice question as shown in Figure 11. The question items are listed in Appendixes 2 and 3. The participant selects an answer for the questions provided by 10 monsters.

(5) The system checks whether the answer is correct or not. If the answer is incorrect, the system moves to step 6.1. Otherwise, the system moves to step 6.2.

(6) Step 6.1: The health points (HP) of the game character are deducted by 20 points. Note that the initial HP of the character is 200 points. The system then checks the HP of the character.
   - If the HP is greater than zero, the participant can continue the game.
   - Otherwise, the game flow moves to Step 7.
   - Step 6.2: If the participant selects the correct answer, the monster is destroyed. The system then checks whether there is a next monster or not.
   - If some monsters remain, the participant goes back to Step 4 to repeatedly continue the procedure.
   - Otherwise, the game flow moves to Step 8.

(7) If the character's HP is less than or equal to zero, the game is over. The participant is then forced to go to the game's starting point and play it again.

(8) When reaching this step, the participant can control the game character to chat with the NPC mentor again. The mentor congratulates the participant on the success of completing the whole learning task, as shown in Figure 12.

(9) If the participants reached this step, they have successfully completed the vocabulary learning task.

From the detailed steps of the above game flow, note that there is a mechanism, to force participants to read the learning materials and try to understand the target words which leads to the conclusion of the game if the participant answers the questions correctly. In addition, helpers in the classroom can assist and teach the participants how to play the game to ensure that all participants strictly follow the flow to complete the learning task. Participants can easily play the game by using a keyboard or mouse. The four keyboard arrows control the movement of the character in the game, while the space bar or the “enter” key triggers events, including answering the question items, completing the instructions, and so on. Alternatively, they can use the mouse to control movement and trigger events of the character in the game. Before playing the game, participants are given oral instructions on how to play. During the main study, there were two helpers who were familiar with the game and computer settings in the classroom to address any technical and game-playing issues for all

participants while they played.

To ensure that the developed role-playing game was easy-to-play and to minimize the impact of various impediments in the game-flow design, I invited seven participants to join another pilot study. This pilot study aimed to investigate (i) whether the game was easy to play, (ii) whether the game flow was appropriate or not, and (iii) whether the process of completing the learning task in the digital game was distracted by the game design or not. The participants in this pilot study completed the digital game learning tasks first. Four participants completed the digital game for Task 2: Reading a text and answering reading comprehension questions, while three participants completed the digital game for Task 3: Reading a text and inferring meanings of the underlined words. The next step was to ask them to indicate their opinions on three statements: (i) "You agree that the game is easy-to-play," (ii) "You agree that the game flow is appropriate," (iii) "You agree that your learning is not distracted by the game flow," through three corresponding five-point Likert scales (i.e., 1. strongly disagree, 2. disagree, 3. neutral, 4. agree, 5. strongly agree).

The mean score asking whether the participants agreed with the first statement "You agree that the game is easy-to-play" was 4.29; the mean score asking whether the participants agreed with the second statement "You agree that the game flow is appropriate" was 4.14; and the mean score asking whether the participants agreed with the third statement "You agree that your learning is not distracted by the game flow" was 4.00. These results indicate that the simple design of game flow and scenarios successfully led to easy-to-play, appropriateness, and no distraction for game-based learning. No participants gave scores of "1 strongly disagree," or "2 Disagree." In other words, all participants had somewhat agreed with the three statements. This result is another indicator to verify the appropriateness of the game design.

3.6 Learning Materials

3.6.1 The Reading Text

The three learning tasks of this study were reading-based and shared the same reading text. The selection of the reading text played an important role in the design of the learning tasks. After a pilot study among five students, I decided to follow the reading text employed in Zou (2017) as it has the following two features: (i) most students were familiar with the topic of the reading text; and (ii) its level of difficulty was suitable for intermediate-level foreign language learners. As explained in the study conducted by Zou (2012), topic familiarity was a prime factor for developing or selecting the

reading text as previous studies (Ellis, 2001; Pulido, 2003; Xie et al., 2016) have revealed that the learning performance can be improved in terms of word retention or reading comprehension if the learner is more familiar with the topic of the text. The other important factor was the level of difficulty of the reading text. To ensure the difficulty of the reading text was at an appropriate level, Zou (2012) used three criteria and the interview in her pilot study. Specifically, the three criteria are "(i) the density of words unfamiliar to the subjects should be approximately 2% of the text; (ii) the text should be about 500 words; and (iii) the contexts of the target words should clearly indicate their meanings" (Zou, 2012, p. 108). These criteria have been supported by several previous studies (Liu and Nation, 1985; Hirsh and Nation, 1992; Laufer, 1997; Nation, 2001; Robinson, 2003).

Therefore, I used the same reading text used by Zou (2012), the topic of which was "coping with the procrastination," as shown in Appendix 2. This reading text was extracted from the book "Wordsmith: A Guide to College Writing" (Arlov, 2000). This text has been employed in previous studies (Kim, 2011; Zou, 2012; Zou, 2017) by revising the length and complexity depending on the language proficiency of the participants. I used Zou's (2012) version of the text. This version has been reduced from 640 words to 507 words by taking the density of unfamiliar words and the suitability of length into account. Specifically, the text is composed of eight paragraphs, and the 10 target words are evenly distributed in the eight paragraphs. Each paragraph includes no more than two target words, and no sentence includes more than two target words. Except the 10 target words, all other words are among the most frequently used 5000 words, so it is unlikely that any participants of this study may not know any words except the target ones.

The participants of the present study were similar to Zou's study (2012), i.e., foreign language learners with intermediate English language proficiency. However, there were still some minor differences between the participants in these two studies. The main participants of Zou's (2012) study were senior-year undergraduate students at Tsinghua University, which is ranked as the top university in Mainland China. In our study, the students were first-year master’s students in The Education University of Hong Kong.

To assess the degree of topic familiarity and the level of difficulty of the reading text, I conducted the pilot study by asking the eight participants to indicate their perceptions and feelings about the reading text. Similar to the previous pilot surveys, the survey used a five-point Likert scale (i.e., 5 - "Very familiar," 4 - "Familiar," 3 - "Neutral," 2 - "Unfamiliar" and 1 - "Very unfamiliar") to

denote the degree of topic familiarity. Similarly, a five-point Likert scale (i.e., 5 - "Very easy," 4 - "Easy," 3 - "Neutral," 2 - "Difficult" and 1 - "Very difficult") was used to denote the level of difficulty of the reading text. The participants in the pilot study were also invited to anonymously indicate their perceptions and feelings about these two criteria.

	Mean	*SD*
Topic Familiarity	3.167	0.937
Levels of Difficulty	3.250	0.965

Table 15: Survey results of the topic familiarity and level of difficulty

As illustrated in Table 15, the survey results of both topic familiarity and the level of difficulty showed that the reading text from Zou (2012) could be used for this study. The mean score regarding topic familiarity was 3.167 with a standard deviation of 0.937 and the mean score of the topic difficulty was 3.250 with a standard deviation of 0.965. These results indicate that almost 90% participants had neutral feeling about the two criteria. Therefore, I can draw a preliminary conclusion that the employment of the reading text from Zou (2012) was reasonable and acceptable for the participants in our study in terms of topic familiarity and the level of difficulty.

3.6.2 The Target Vocabulary

A total of 10 target words in the reading text were chosen based on two reasons.

- Comparable to previous studies- Many previous studies (Hulstijn and Laufer, 2001; Laufer, 2003; Folse, 2006; Laufer and Girsai, 2008; Kim, 2011; Zou, 2012; Zou, 2017) have also decided upon, 10, for the number of words to use in a test. This choice of 10 target words helps to remain consistent with other experimental results for the sake of comparison.
- Density of unfamiliar words of text- As mentioned, the first two criteria for selecting the reading text were "(i) the density of words unfamiliar to the subjects should be approximately 2% of the text; and (ii) the text should be about 500 words" (Zou, 2012, p. 108). Therefore, the number of target (i.e., unfamiliar) words should be 2% of 500 words, which is ten.

After confirming the number of target words, the next step was to select the ten target words. In this study, I follow the criteria proposed by Zou (2012) and Christ, Wang, and Chiu (2017). Detailed criteria about how to select target words were specified as follows.

- Unfamiliar to participants- The participants in this study had to be unfamiliar with the selected target words. If participants are familiar with some target words, it would be challenging for us to identify whether the target words were learned through the designed learning tasks or whether the words were already known by the participants. Although there was a minor difference between the groups of participants of Zou's study (2012) and the current study, the same target word selection strategy, which used "words outside the most frequently used 8000 words listed in the Test for English Majors Band 8 (TEM 8), was the most likely way to select the unknown target words, as the vocabulary list of TEM 8 has already included all words listed in the College English Test Band 4 (CET-4) and College English Test Band 6 (CET-6) in Mainland China" (Zou, 2012, p.128), was employed in the current study. This was because CET-6 has frequently been used as proof of the English proficiency for the admission of postgraduate programmes at local Hong Kong universities for several years (EdUHK, 2019).
- Various parts of speech- The selected target words had to represent various part of speech as the potential impact from grammatical categories has been identified in previous studies (Ludwig, 1984; Laufer, 1990; Nation, 2001; Folse, 2006; Christ et al., 2017; Zou, 2012). Specifically, Ludwig (1984) showed that the learning difficulty seems to be different if the target words were verbs, nouns, adverbs, or adjectives. Laufer (1990) found that adverbs were most difficult, adjectives and verbs were second most difficult, and nouns were the easiest ones to be learned. Nation (2001) found that the meaning of nouns and verbs tended to be inferred more easily than other parts of speech as they have more relationships with the context. Therefore, this study chose target words of various parts of speech (Zou, 2012; Christ et al., 2017).
- Sole meaning- Target words had to contain only a single consistent meaning when embedded in the reading contexts (Zou, 2012). Zou (2012) explained the reasons for this as follows: (i) the target words might induce additional ILH involvement loads if the target words had multiple meanings in the same context; (ii) target words are more difficult to learn if they are polysemous, as supported by a research study conducted by Saeman (1970); and (iii) the consistent and single meaning of target words can ensure the reliability of vocabulary tests in this study (e.g., it is difficult to determine the degree of understanding of a target word containing multiple meanings if the participants only understand one or two of these meanings). Therefore, words with a single consistent meaning in the reading context were selected.

- Conceptually familiar- As suggested by Zou (2012) and Christ et al. (2017), target words should convey concepts which are familiar to the participants as participants do not have to use extra mental effort to construct new concepts during the learning process. According to the findings of previous studies (Daneman, & Green, 1986; Nagy, Anderson, & Herma, 1987; Nation, 2001; Rupley, & Nichols, 2005; Shefelbine, 1990), if the concepts conveyed by target words are familiar to the subjects, learning and inferring their meanings is easier than target words which convey unfamiliar concepts. Hence, the target words with the familiar concepts were selected (Zou, 2012; Christ et al., 2017).
- Important for text comprehension- The target words should be important for text comprehension (Wieland, 2008), and the participants should feel the need of learning the targe words while reading the text (Wieland, 2008). Thus, I designed the task in a way that the target words were closely related to the completion of the reading task and the understanding of the core idea of the reading text.

Based on the above criteria, I decided to employ the ten target words from Zou's study (2012) as shown in Appendix 4. Not only did the participants in Zou's study and ours have similar English language proficiency (i.e., IELTS and CET Band 6 which are considered as having comparable English proficiency levels for admission into postgraduate programmes in local universities [EdUHK, 2019]), but our pilot studies also confirmed the appropriateness of these target words. There were four pilot studies assessing the target words. The first study was to determine whether the target words were unfamiliar to the participants or not, and the second one was to examine whether the concepts conveyed were familiar to the participants or not (Zou, 2012). The third one examined whether the number of target words was appropriate or not, and the final one examined whether the target words were necessary for text comprehension.

The first pilot study, which aimed to test whether the target words would likely be unfamiliar to the project participants, was conducted on nine postgraduates who shared similar backgrounds with the intended participants of the main study. Similar to Zou's methods (2012) on the pilot study, I employed the modified vocabulary levels test (VLT) proposed by Nation (1990), which "test[s] lexical knowledge via a matching task in which words in one column have to be matched to definitions in an adjacent column" (Zou, 2012, p. 115). A sample of the modified VLT, which categorized the target words according to their parts-of-speech, is shown in Table 16. To avoid participants randomly guessing the correct answers, two distracters were added to the word list in the left column. The result

of this pilot study revealed that none of the target words were correctly matched by the participants showing that the selected ten target words were likely unfamiliar to the intended participants of the main study.

1. stymie	
2. renege	( ) to upset someone by laughing at him
3. eulogy	( ) to break a promise, an agreement, etc
4. taunt	( ) very careful to ensure that something is done properly
5. assiduous	

Table 16: *A sample of modified vocabulary levels test (Zou, 2012, p. 116)*

The second pilot study, which aimed to assess whether the concepts conveyed were familiar to the participants or not, was conducted on the same nine students. For this pilot study, the dictionary entry of each target word was given to the participants. To assist in understanding the concepts conveyed by the target words, the dictionary entry contained the word form, pronunciation, semantic meaning, part-of-speech and a sample sentence. Appendix 5 shows the entries of all 10 target words. The participants were provided with the entries and then asked to indicate their degree of familiarity with the concepts conveyed by the word on a five-point Likert scale: "very familiar, familiar, neutral, unfamiliar, very unfamiliar." The definitions of the target words were extracted from the Longman Dictionary of Contemporary English 5. Table 17 shows the familiarity scale. The result revealed that none of the concepts conveyed by the selected target words were unfamiliar to the nine participants. The familiarity score averaged 4.27 (5 denotes "very familiar, and 1 denotes "very unfamiliar" for the ten target words). For each word, the average score ranged from 3.89 to 4.79. Therefore, I concluded that the concepts conveyed by the selected ten target words were familiar to the intended participants of the main study.

The third pilot study, again conducted on the nine postgraduate students, aimed to examine whether the number of target words was appropriate or not. Similar to the second pilot study, the participants were asked to indicate their feelings and perceptions about the degree of appropriateness in learning the ten target words. The participants indicated their opinions on the statements such as: "You agree that the number of target words is appropriate," using a five-point Likert scale (i.e., 1. strongly disagree, 2. disagree, 3. neutral, 4. agree, 5. strongly agree). The resulting average score was 4.22. None of the participants marked "strongly disagree" or "disagree." In other words, almost all

participants acknowledged that the size of target words was appropriate for this learning task.

Instructions: Please indicate your degree of familiarity with the concepts conveyed by the target words. You need to select the answers from a five-point Likert scale.
1. taunt /tɔːnt $ tɒːnt/ verb [transitive] to try to make someone angry or upset by saying unkind things to them → tease taunt somebody about something The other children taunted him about his weight. taunt somebody with something They taunted him with the nickname 'Fatso.' And he'll believe you, will he?' Maria taunted. -tauntingly adverb taunt • Of course he wasn't, an inner voice taunted. • They were accosted by three white youths who taunted and then attacked them. • The older boys taunted Chris and called him a girl. • Or maybe, as she'd taunted earlier, his actions were governed by boredom. • She was held in jail overnight, and she alleges in her lawsuit that guards taunted her with ethnic slurs. • Now the telephone had acquired a personality, sat on the shelf so smug, taunting her with its silence. • He couldn't forget how they had taunted him about his appearance. • She went on taunting him until he lost his temper. • They taunt me and beat me. • When I didn't want to fight he would taunt me repeatedly. "Coward, " he would say, "coward, coward, coward...." • You can blast your buddies and taunt them verbally at the same time.
Please indicate your degree of familiarity with the concepts conveyed by taunt here: Very unfamiliar 2. Unfamiliar 3. Neutral 4. familiar 5. Very familiar

Table 17: *A sample questionnaire measuring word familiarity (Zou, 2012)*

The final pilot study concerning the appropriateness of the selected target words, which

examined whether the target words were important or not, was conducted on seven different postgraduate students with similar educational backgrounds. Specifically, I examined their participants' perceptions about the importance of the target words by asking them to indicate their feelings and perceptions about the degree of importance of the 10 target words for understanding the text. They completed the reading task and then gave their opinions on the statement, "You agree that the ten target words are important to understand the reading text" via a five-point Likert scale (i.e., 1. strongly disagree, 2. disagree, 3. neutral, 4. agree, 5. strongly agree). The average score was 4.00 indicating that these ten target words were important for the participants to understand the reading text.

3.7 Measurement Tool: The Modified Vocabulary Knowledge Scale

I used the modified vocabulary knowledge scales (MKVS) proposed by Folse (2006) in the pre-test, immediate and delayed post-tests to measure the participants' prior knowledge of the target vocabulary, their initial learning of the target vocabulary, and retention of the vocabulary knowledge one week later, as explained in Section 3.3 Research Design. This subsection introduces the measurement of the three learning tasks in detail.

As reviewed in Section 2.2.1, vocabulary knowledge can be divided into a few aspects. For example, Webb (2005) divided vocabulary knowledge into five aspects – orthography, syntax, association, grammatical functions, and meaning and forms, which can be measured by receptive and productive tests. Nation (2003, p. 31) presented "a list of the vocabulary knowledge types that native-speakers typically possess and the list including (1) a word's spoken form, (2) a word's written form, (3) a words' part-of-speech, derivative forms, and grammatical patterns, (4) a word's collocations, (5) how frequently a word is used in a language, (6) how frequently a word is used in a language, (7) the many stylistic constraints which determine if a word is appropriate in a particular context, (8) a word's conceptual meaning(s), and (9) a word's semantic network of associations".

Furthermore, the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) assumes that knowledge about word meanings and forms is the most important element for developing reading skills. "One very common characteristic of learner vocabulary was that vocabulary knowledge was restricted to certain aspects of a word rather than encompassing the full knowledge normally available to native speakers" (Zou, 2012, p. 136). Therefore, it was more likely that the participants would acquire only a partial rather than the full meaning of the selected target words within a few instances of learning the new words (Zou, 2012). Motivated by this observation, the selected tool for task

measurement should evaluate the acquisition of only a partial knowledge of the target words. Accordingly, the Modified Vocabulary Knowledge Scale (MVKS) was selected as the task measurement tool. MVKS is a modified version used by Folse (2006) and Zou (2012). The Vocabulary Knowledge Scale (VKS), the original version of MVKS, proposed by Paribakht and Wesche (1997), arrived in the form of a self-report, five-point Likert scale for measuring the degree of self-perceived and demonstrated knowledge of target words by participants (Kim, 2011). The five word-knowledge descriptors of VKS are listed in Table 18.

Scales	Descriptions
5	The participant is familiar with the word.
4	The participant does not know the meaning but is familiar with the word.
3	The participant can provide a correct translation/synonym for the word.
2	The participant can use the word with semantic appropriateness in a sentence.
1	The participant can use the word with semantic and grammatical appropriateness in a sentence.

Table 18: The five scale of VKS of vocabulary knowledge (Paribakht and Wesche, 1997; Zou 2012)

- I do not know what this word means. - I know this word. It means_____. (provide an English synonym or a translation in your native language) - I can use this word in a good example sentence. Write your sentence here: _____ (If you do #3, you must do #2 also.) *Note*: Adapted from Paribakht & Wesche, 1997

Table 19: Folse's (2006) Modified Vocabulary Knowledge Scale (MVKS)

However, previous studies (Wesche and Paribakht, 1996; Read, 2000; Kim, 2011), have claimed that VKS was developed for tracking the development of vocabulary knowledge rather than generally measuring vocabulary knowledge. Therefore, the following studies (Folse, 2006; Zou, 2012) proposed using modified versions of the VKS in their studies.

Folse's (2006) MVKS is shown in Table 19, which is a modified version of VKS proposed by

Paribakht and Wesche (1997). MVKS employs a three-point scale rather than a five-point scale in the original version as the three-point scale can easily be used in measuring the vocabulary knowledge of participants through a test and self-report form. In addition, in a previous study (Zou, Xie, Rao, Wong, Wang & Wu, 2017), the researchers found that a five-point scale of vocabulary knowledge has slightly better accuracy (i.e., 1.5%) than a three-point scale; however, the cost in terms of participant effort and evaluation was relatively more expensive. Specifically, two points are given if the participants can demonstrate not only the correct semantic meaning of a target word but also the correct usage of the word in a sentence. One point is given if participants can understand the correct meaning of the target word but cannot use it correctly in a sentence. No points are given if participants neither understand the meaning of the target word nor use it correctly in a sentence.

Modified Vocabulary Knowledge Scale

Instructions:

1. If you cannot remember the meaning of the word, please tick (√) in Column [A].
2. If you can remember the meaning of the word, please write it in Column [B]. You can use either English or Chinese.
3. If you can write a sentence by using this word, please write one in Column [C]. Your sentence should include at least seven words.

	I cannot remember the meaning of the word.	I remember the meaning. The meaning is	I can use the word in a sentence. For example,
	[0]	[1]	[2]
Word 1			
Word 2			
...			
Word 10			

Note: Adapted from Paribakht & Wesche, 1997

Table 20: The Modified Vocabulary Knowledge Scale (Folse, 2006) further adapted by Zou (2012)

Zou (2012) employed Folse's (2006) MVKS in her research as shown in Table 20. The

differences between Zou's version (2012) and Folse's (2006) were minimal: (i) the detailed instructions were moved to the top part; (ii) a concise table form was used for writing the answer; (iii) the semantic meaning of the target words was not only expressed by a synonym or a translation but also other forms; and (iv) the example sentences for the target words had to contain at least seven words. In this study, I followed Zou's (2012) version for measuring the vocabulary knowledge scale.

In sum, this MVKS was used as (1) the measurement tool for the pre-test to evaluate the participants' prior knowledge of the target vocabulary, (2) the measurement tool for the immediate post-test to evaluate the participants' initial learning of the target vocabulary knowledge, and (3) the measurement tool for the delayed post-test to evaluate the participants' retention of the target vocabulary knowledge. The maximum possible scores for the pre-test, immediate and delayed post-tests are all 20, as there are 10 target words. The minimum possible scores for the three tests are all zero.

3.8 Research Ethics

I conducted the study strictly following the research ethics of University of Bristol and The Education University of Hong Kong. In line with the relevant policies on research integrity, I applied to the Human Research Ethics Committee of both University of Bristol and The Education University of Hong Kong for ethical clearance. I obtained the ethical approval prior to any data collection or analysis as shown in Appendix 6 to 12.

Specifically, as explained in Section 3.4 Participants, I invited 192 students to participate in the research through different channels, including emails, social media platforms, posters, and so on. Based on the selection criteria, 159 students participated in the pre-test. While 24 of them, who knew more than two target vocabulary, were excluded from the research at this stage based on the selection criteria. I selected the remaining 135 students for the research and explained to them that no potential risks were involved in this study. I also gave them a brief introduction of the project, telling them that the whole experiment process, including the pre-test, learning process, and posttests, lasted for around one hour and were conducted in regular classrooms or computer labs. All students participated in the project voluntarily, and they were ensured that they had the right to question any part of the procedure and could withdraw at any time without penalty of any kind. No one withdrew from the study.

As explained in Section 3.3 Research design, a total of 135 students participated in the project.

Their prior knowledge of the target vocabulary was measured by a pre-test before the experiment. They were then randomly and equally assigned to three groups (45 students in Group A did Task 1, 45 students in Group B did Task 2, and 45 students in Group C did Task 3). All students participated in the immediate post-test, which evaluated their initial learning of the target vocabulary knowledge, after the learning session. Subsequently, 10 students from each group were interviewed to collect data concerning their learning processes. The remaining 35 students from each group were post-tested again one week later to measure their retention of the target vocabulary knowledge. More details are presented in Section 3.3.1 Pre-test, Section 3.3.2 Learning processes, Section 3.3.3 Immediate post-test, Section 3.3.4 In-depth interview, and Section 3.3.5 Delayed post-test, respectively.

3.9 Data Analysis

This study involved three vocabulary learning tasks: Task 1: "Reading a text and answering reading comprehension questions on paper", Task 2: "Reading a text and answering reading comprehension questions in a digital game," and Task 3: "Reading a text and inferring meanings of the underlined words in a digital game." T1, T2, and T3 are used as abbreviations for Task 1, Task 2, and Task 3, respectively, in this section. The involvement loads of the three learning tasks in three components, including *Need*, *Search*, and *Evaluation*, (Section 3.2) were not only the same but also equally distributed. Specifically, the involvement loads of the three components in ILH were at moderate levels (1+1+1) for the three learning tasks (Hu & Nassaji, 2016). As the involvement loads of these three learning tasks were the same and equally distributed in terms of *need*, *search*, and *evaluation*, the differences between the three tasks were essential and critical.

Research questions	Learning tasks	ILH		
		Need	Search	Evaluation
1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one?	Task 1: Reading a text and answering reading comprehension questions (on paper)	1	1	1
	Task 2: Reading a text and answering reading comprehension questions (in a digital game)	1	1	1

2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form?	Task 2: Reading a text and answering reading comprehension questions (in a digital game)	1	1	1
	Task 3: Reading a text and inferring meanings of the underlined words (in a digital game)	1	1	1

Table 21: The purposes of two comparisons among the three groups

As shown in Table 21, the only difference between Task 1 and Task 2 is the learning environment. The conventional learning environment (on paper) was employed in Task 1, whereas the digital game environment was employed in Task 2. Thus, I compared these two tasks to answer the first research question. I collected data concerning the effectiveness of these two tasks in terms of promoting students' initial learning of the target vocabulary through an immediate posttest, which was conducted immediately after the participants completed Task 1 and Task 2. I also collected data concerning the effectiveness of these two tasks in terms of promoting students' retention of the target vocabulary through a delayed posttest, which was conducted one week after the participants completed the tasks. Through comparing the participants' scores in the immediate post-test, I evaluated whether for the same learning task, learners performed better in initial vocabulary learning in a digital game environment than in a conventional one. Through comparing the participants' scores in the delayed post-test, I evaluated whether for the same learning task, learners performed better in vocabulary retention in a digital game environment than in a conventional one.

Objective 1.1: To investigate whether for the same learning task, learners performed better in initial vocabulary learning in a digital game environment than in a conventional one	Comparison 1.1: To compare the immediate post-test scores of the participants who completed Task 1 and Task 2

Objective 1.2: To investigate whether for the same learning task, learners performed better in vocabulary retention in a digital game environment than in a conventional one	Comparison 1.2: To compare the delayed post-test scores of the participants who completed Task 1 and Task 2
Objective 2.1: To investigate whether for two different learning tasks, which are similarly effective in terms of promoting vocabulary learning in a conventional learning environment, learners performed differently in a digital game environment from the perspective of initial vocabulary learning	Comparison 2.1: To compare the immediate post-test scores of the participants who completed Task 2 and Task 3
Objective 2.2: To investigate whether for two different learning tasks, which are similarly effective in terms of promoting vocabulary learning in a conventional learning environment, learners performed differently in a digital game environment from the perspective of vocabulary retention	Comparison 2.2: To compare the delayed post-test scores of the participants who completed Task 2 and Task 3

Table 22: The objectives and comparisons

The only difference between T2 and T3, was the learning environment. Therefore, I compared these two tasks to answer the second research question. I collected data concerning the effectiveness of these two tasks in terms of promoting students' initial learning of the target vocabulary through an immediate posttest, which was conducted immediately after the participants completed Task 2 and Task 3. I also collected data concerning the effectiveness of these two tasks in terms of promoting students' retention of the target vocabulary through a delayed posttest, which was conducted one week after the participants completed the tasks. Similarly, through comparing the participants' scores in the immediate post-test, I evaluated whether for two different learning tasks, which are similarly effective in terms of promoting vocabulary learning in a conventional learning environment, learners performed differently in a digital game environment from the perspective of initial vocabulary learning. Through

comparing the participants' scores in the delayed post-test, I evaluated whether for two different learning tasks, which are similarly effective in terms of promoting vocabulary learning in a conventional learning environment, learners performed differently in a digital game environment from the perspective of vocabulary retention. The objectives and comparisons are summarized in Table 22.

The interview data were analyzed to understand the participants' learning experiences. Specifically, I transcribed the audio recordings of the interviews literally first, and then read the transcripts several times to obtain a general understanding of the whole picture of the participants' learning processes. Many other studies applied similar methods as well, for example, Zou (2012, 2016, 2017). After obtaining a comprehensive understanding, I started to concentrate specifically on the episodes when the participants reported the features of the learning tasks and learning environments conducive to their learning of the target vocabulary knowledge. I also paid attention to the differences among the three tasks and wrote down notes concerning how these differences might have possible influences on the initial learning and retention of the target vocabulary.

Specifically, I read through each transcript carefully and paid attention to the parts when learners reported what they felt useful for their learning of target words. I also highlighted the important words, phrases, and sentences that indicated the students' perceptions of the learning experience, as well as what they did and noticed. After highlighting these, I found that these texts were normally what the students noticed or felt during the learning process, including what features motivated or demotivated them, what engaged them, what directed their attention to the target vocabulary knowledge, what distracted them, what they liked or disliked, and what they felt useful or useless. Based on these categories, I grouped and labeled the data accordingly. After the categorizing and labeling, I further organized and analyzed the data to identify the features of different learning tasks.

3.10 Summary

In this chapter, I have explained the research methodology in detail. In the introduction, I explained that the purpose of the current study is to investigate whether the ILH developed by Hulstijn and Laufer (2001), a measurement theory for estimating involvement loads for tasks in incidental vocabulary acquisition, can be applied to the tasks in DGBVL as the experimental results of existing studies cannot be generalized for DGBVL, and because there have been few studies on this topic. Motivated by this research objective and theoretical analysis, I asked two research questions which

were: 1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one? 2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form? For the interpretation of the research results, I also aimed to investigate whether the ILH or competing theories can account for the results.

The participants and criteria for selecting participants were then introduced and discussed. The participants were from The Education University of Hong Kong and had similar educational backgrounds (i.e., first year Master of Education students), language proficiency levels (i.e., intermediate English learners), and prior knowledge of the target vocabulary (i.e., knew no more than two target words). Their first languages were either Mandarin or Cantonese. The details are presented in Section 3.4 Participants. I explained the experimental design in detail in Section 3.3 Research design. The research framework contained two comparisons among three group, which were (i) "responding either on paper or in a digital game" (i.e., Group A vs. Group B) and (ii) "responding on two different tasks which have the same ILH loads" (i.e., Group B vs. Group C). The first comparison was to answer the first research question and the second comparison was to answer the second research question. The two comparisons among the three groups formed the research framework of this study. All participants were randomly assigned to complete one of the three tasks and were given the same amount of time to complete the tasks.

The details concerning the tasks design and research design are explained in Section 3.2: Task Design and Section 3.3: Research Design, which further includes explanations of the pre-test, the learning process, the procedure for conducting the pre-test, immediate post-test, and delayed post-test, and the interview. Subsequently, in Section 3.4: Participants, I explained the criteria of selecting the participants for the three groups. In Section 3.5, I explained the game design; and in Section 3.6, I explained the procedures and rationale for selecting the learning tasks, including the selection of reading topic, reading text, and target words. The learning task measurement was then discussed in Section 3.7: Measurement Tool. The proposed task measurement was based on Folse's (2006) Modified Vocabulary Knowledge Scale (MVKS), which was further revised by Zou (2012). Finally, the differences between Folse's MVKS (2006) and Zou's task measurement (2012) were explained in this section.

In Section 3.8 Research ethics, I explained the ethical issues and relevant procedures. In Section 3.9, I provided details concerning how different data were analyzed for the answering of the two research questions and explained why these comparisons and analysis could lead to the successful question answering.

CHAPTER 4: EXPERIMENTAL RESULTS

4.1 Introduction

In this chapter, I report the results from the experiment. Specifically, I report the results of the pre-test, immediate post-test, delayed post-test for the three groups in the following six sections. I employed the MVKS as the task measurement for scoring all tasks in our experiment. The scores on the tasks reflect the effectiveness of promoting vocabulary learning on the tasks (Zou, 2012).

The remaining parts of this chapter are organized as follows. In Section 4.2, the results are first presented to provide a global view. The detailed results of the test of normality and test of homogeneity of variance are then introduced and reported in Section 4.3 and Section 4.4, respectively. In Section 4.5, I discuss the effectiveness of the three vocabulary learning tasks. I then report the results of the first comparison between Task 1 (Reading a text and answering reading comprehension questions on paper) and Task 2 (Reading a text and answering reading comprehension questions in a digital game) to examine whether learners had a better performance in the two learning environments (on paper vs. in a digital game) for the same learning task in Section 4.6. In Section 4.7, I report the results of the second comparison between Task 2 and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) to examine whether students performed differently in a digital game learning environment for two different learning tasks with the same involvement loads.

4.2 Overall Experimental Results

The three groups of participants' scores in the pre-test are summarized in Table 18. As explained in Section 3.3 Research Design, the pre-test evaluated the participants' prior knowledge of the target vocabulary, and Folse's (2006) modified vocabulary knowledge scale (MVKS) was used as the measurement tool. I selected only the students who knew no more than two target vocabulary to participate in the research, as explained in Section 3.3.1 Pre-test and Section 3.4 Participants. So, the maximum score of the students in the pre-test was 4, and the minimum score was 0 as some students knew none of the target vocabulary before participating in the research. I further run several tests to examine the participants' pre-test scores. The results showed that the mean scores and standard deviations of the three groups of participants' pre-test scores were similar. Thus, it can be concluded that the three groups of participants shared similar prior knowledge of the target vocabulary, and the differences of their immediate post-test scores, which indicated their initial learning of the target vocabulary knowledge, and the differences of their delayed post-test scores, which indicated their

retention of target vocabulary knowledge, resulted mainly from the different tasks that they did during the treatment.

The pre-test, immediate and delayed post-test scores of the three groups of students who did Task 1, Task 2, and Task 3 were independent. All participants in each group were different. No student was in more than one group.

As presented in Table 23, the mean immediate post-test score of the participants who did Task 1 was 5.11 (with the standard deviation of 2.62), which was lower than those of the participants who did Task 2 (M = 7.22; SD = 3.01) and Task 3 (M = 10.62; SD = 2.52). The mean score of the participants who did Task 3 was the highest among the three. Concerning the three groups of participants' highest immediate post-test scores, that of the participants who did Task 1 was 12, that of those who did Task 2 was 13, and that of those who did Task 3 was 16. The three groups of participants' lowest immediate post-test scores were 0, 2, and 6, respectively. The overall immediate post-test mean was 7.66 (with a standard deviation of 3.51), showing that all three tasks promoted generally satisfactory initial learning of the target vocabulary knowledge.

	N	Pre-test scores				Immediate post-test scores				Delayed post-test scores			
		M	*SD*	*Max*	*Min*	*M*	*SD*	*Max*	*Min*	*M*	*SD*	*Max*	*Min*
Task 1	35	1.43	1.48	4	0	5.22	2.61	12	0	4.17	2.43	10	0
Task 2	35	1.40	1.35	4	0	7.22	3.01	13	2	5.46	2.55	10	0
Task 3	35	1.31	1.30	4	0	10.62	2.52	16	6	7.46	2.31	12	3
Total	105	1.38	1.36			7.66	3.51			5.69	2.76		

Table 23: Descriptive statistics of the participants' pre-test, immediate and delayed post-test scores

The three groups of the participants' mean scores in the delayed post-test displayed similar trends as those of the immediate post-test. The mean delayed post-test score of the participants who did Task 1 was 4.17 (with the standard deviation of 2.43), which was lower than those of the participants who did Task 2 (M = 5.46; SD = 2.55) and Task 3 (M = 7.46; SD = 2.31). The three groups of participants' highest delayed post-test scores were 10, 10, and 12, respectively. Their lowest delayed post-test scores were 0, 0, and 3, respectively. The overall delayed post-test mean was 5.69 (with a standard deviation of 2.76), showing that all three tasks promoted generally satisfactory retention of the target vocabulary knowledge, considering that the participants' prior knowledge of the target

vocabulary was very little.

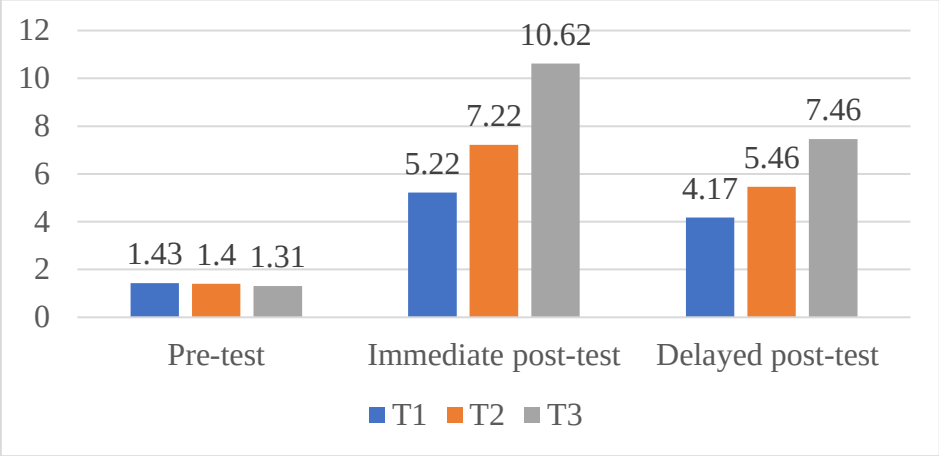


Figure 13: The overall MVKS scores of the three groups of participants in the three tests

The MVKS scores of the three groups of participants who did the three tasks in the pre-test, immediate post-test, and delayed post-test were consolidated and plotted in Figure 13. From the trends and results from the figure, it can be observed that:

(1) The mean scores of the immediate post-test of the participants who did Task 1 (Reading a text and answering reading comprehension questions on paper) was lower than those of the participants who did Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game), indicating that Task 1 was less effective than Task 2 and Task 3 in terms of promoting initial learning of the target vocabulary knowledge.

(2) The mean scores of the delayed post-test of the participants who did Task 1 was lower than those of the participants who did Task 2 and Task 3, indicating that Task 1 was less effective than Task 2 and Task 3 in terms of promoting retention of the target vocabulary knowledge.

(3) The mean scores of the immediate post-test of the participants who did Task 2 was lower than those of the participants who did Task 3, indicating that Task 2 was less effective than Task 3 in terms of promoting initial learning of the target vocabulary knowledge.

(4) The mean scores of the delayed post-test of the participants who did Task 2 was lower than those of the participants who did Task 3, indicating that Task 2 was less effective than Task 3 in terms of promoting retention of the target vocabulary knowledge.

(5) For all three learning tasks, the mean scores of the participants in the immediate post-test were higher than their mean scores in the delayed post-test.

4.3 Test of Normality

To examine whether the three groups of participants' MVKS scores in the pre-test and the immediate and delayed post-tests were normally distributed, I employed two typical statistical tests, Kolmogorov-Smirnov and Shapiro-Wilk, to examine the data.

For the three groups of participants' pre-test scores, the values of Kolmogorov-Simirnoy tests of the three sets of data were all smaller than 0.01. Similarly, the statistic values of the Shapiro-Wilk tests of the three sets of data were smaller than 0.01. These results showed that the three groups of participants' pre-test scores were not normally distributed, which is likely because I selected only the students who knew no more than two target vocabulary to participate in the study. As explained earlier in Section 3.4 Participants, this selection criterion aims to minimize the possible influences of students' prior knowledge on their learning of the target vocabulary.

For the three groups of participants' immediate post-test scores, the values of Kolmogorov-Simirnoy tests of the three sets of data were 0.11, 0.12, and 0.12, respectively. The degrees of freedom (i.e., df values) for the three tasks were all 35. More importantly, the sig. values of the Kolmogorov-Simirnoy tests of the three groups of participants' MVKS scores were 0.20, 0.20, and 0.16, respectively. A sig. value greater than 0.05, indicates that the MVKS scores followed a normal distribution. The Shapiro-Wilk tests showed whether the data follow the normal distribution. The statistic values of the Shapiro-Wilk tests of the three groups of participants' MVKS scores were 0.974, 0.963, and 0.967, respectively. The df values for the three groups were all 35. The sig. values of the Shapiro-Wilk tests of the three groups of participants' MVKS scores in the immediate post-tests were 0.55, 0.27, and 0.36, respectively, indicating normal distribution of the data.

For the three groups of participants' delayed post-test scores, the values of Kolmogorov-Simirnoy tests of the three sets of data were 0.10, 0.11, and 0.10, respectively. The df values for the three tasks were all 35. The sig. values of the Kolmogorov-Simirnoy tests of the three groups of participants' MVKS scores were all 0.200, which indicated normal distributions. The values of the Shapiro-Wilk tests of the three groups of participants' MVKS scores were 0.970, 0.974, and 0.970, respectively. The df values were all 35. The sig. values of Shapiro-Wilk tests of the three groups of participants' MVKS scores in the delayed post-tests were 0.45, 0.55, and 0.45, respectively. As they were greater than 0.05, the data were normally distributed.

In sum, the three groups of participants' MVKS scores in the pre-test were not normally distributed, but their scores in both immediate and delayed post-tests were normally distributed. These assumptions were examined by two types of statistical tests, the Kolmogorov-Smirnov and Shapiro-Wilk tests.

4.4 Test of Homogeneity of Variance

I also examined whether the three groups of participants' MVKS scores in the pre-test and immediate and delayed post-tests shared homogeneous variances. I conducted the Levene's test with the "task type" (i.e., Task 1, Task 2, and Task 3) as the grouping variable and the "three groups of participants' MVKS scores in the pre-test, immediate post-test, and delayed post-test" as the dependent variables.

For the three groups of participants' pre-test scores, the sig. value of the Levene's test was 0.53. As it is greater than 0.05, it can be concluded that the three groups of participants' scores in the pre-test shared homogeneous variances. For the participants' scores in the immediate post-test, the sig. value of the Levene's test was 0.24, which is greater than 0.05, meaning that there were no significant differences among the three groups of data. Thus, the three groups of participants' scores in the immediate post-test shared homogeneous variances. For the participants' scores in the delayed post-test, the sig. value of Levene's test was 0.91, which similarly indicated that there were no significant differences among the three groups of data. So, the three groups of participants' scores in the delayed post-test also shared homogeneous variances.

In sum, the three groups of participants' MVKS scores in the pre-test, immediate post-test, and delayed post-test all shared homogeneous variances, as verified by the Levene's tests. That is, although the three groups of participants' means of the immediate and delayed post-tests differed greatly, they passed the test of homogeneity of variance. As presented earlier in Section 4.3 Test of normality, the three groups of participants' MVKS scores in the immediate and delayed post-tests also met the requirement of normal distribution.

4.5 Task Effectiveness in Terms of Promoting Initial Learning and Retention of Target Vocabulary Knowledge

The post-test scores of the participants are compared in this section through four independent samples t-tests.

1) For the first independent samples t-test, I compared the immediate post-test scores of the participants who did Task 1 (Reading a text and answering reading comprehension questions on paper) and those who did Task 2 (Reading a text and answering reading comprehension questions in a digital game).
2) For the second independent samples t-test, I compared the immediate post-test scores of the participants who did Task 2 and those who did Task 3 (Reading a text and inferring meanings of the underlined words in a digital game).
3) For the third independent samples t-test, I compared the delayed post-test scores of the participants who did Task 1 and those who did Task 2.
4) For the fourth independent samples t-test, I compared the delayed post-test scores of the participants who did Task 2 and those who did Task 3.

The main foci of these independent samples t-tests are on two comparisons, (1) the comparison between Task 1 and Task 2, the participants of which did the same learning task in two different learning environments (i.e., one on paper and one in a digital game environment); and (2) the comparison between Task 2 and Task 3, the participants of which did two different tasks with the same involvement load (i.e., moderate need, search, and moderate evaluation) in the same digital game environment.

The results of the two comparisons can indicate (1) whether for the same learning task, learners performed better in initial learning and retention of target vocabulary knowledge in a digital game environment than in a conventional one; and (2) whether for two different learning tasks with the same involvement load, learners performed differently in a digital game environment from the perspective of their initial learning and retention of the target vocabulary knowledge. Many researchers agreed that the effectiveness of vocabulary learning tasks can be demonstrated by the differences between the learners' post-test scores (e.g., Folse, 2006; Laufer & Hulstijn, 2001; Keating, 2008; Zou, 2012; Hu & Nassaji, 2016).

4.5.1 Comparing Task 1 and Task 2 in Terms of Promoting Initial Learning

As presented in Table 24, the values in the "Equal variances assumed" row were the results of the independent samples t-test, as the sig. value of the Levene's test for equality of variances was 0.15, greater than 0.05. The sig. value of the sig. (2-tailed) was 0.003, which was smaller than the significance level of 0.05, indicating that these two sets of data were significantly different. Similarly,

the 95% confidence interval of the difference ranged from -3.46 to -0.76, excluding the value of 0, which also indicated significant difference between the mean scores of the participants who did Task 1 and Task 2 in the immediate posttest.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	*Sig.*	*t*	*Df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	Lower	Upper
Equal variances assumed	2.07	.15	-2.96	68	.004*	-2.00	.67	-3.34	-.65
Equal variances not assumed			-2.96	66.68	.004	-2.00	.67	-3.34	-.65

**p < 0.05*

Table 24: Results of the independent samples t-test of the participants' scores in the immediate posttest (Task 1 vs. Task 2)

Thus, in terms of promoting initial learning, Task 1 (Reading a text and answering reading comprehension questions on paper) was significantly less effective than Task 2 (Reading a text and answering reading comprehension questions in a digital game). This indicated that for the same learning task, learners performed better in initial vocabulary learning in a digital game environment than in a conventional one.

4.5.2 Comparing Task 2 and Task 3 in Terms of Promoting Initial Learning

As presented in Table 25, the values in the "Equal variances assumed" row were the results of the independent samples t-test, as the sig. value of the Levene's test for equality of variances was 0.14, greater than 0.05. The sig. value of the sig. (2-tailed) was smaller than the significance level of 0.05, indicating that these two sets of data were significantly different. Similarly, the 95% confidence interval of the difference ranged from -4.72 to -2.07, excluding the value of 0, which also indicated significant difference between the mean scores of the participants who did Task 2 and Task 3 in the immediate posttest.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	*Sig.*	*t*	*Df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	Lower	Upper
Equal variances assumed	2.23	.14	-5.12	68	.00*	-3.40	.66	-4.72	-2.07
Equal variances not assumed			-5.12	65.96	.00	-3.40	.66	-4.72	-2.07

**p < 0.05*

Table 25: Results of the independent samples t-test of the participants' scores in the immediate posttest (Task 2 vs. Task 3)

Therefore, in terms of promoting initial learning, Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) was significantly more effective than Task 2 (Reading a text and answering reading comprehension questions in a digital game). This indicated that for two different learning tasks with the same involvement load (i.e., moderate need, search, moderate evaluation), learners performed differently in a digital game environment.

4.5.3 Comparing Task 1 and Task 2 in Terms of Promoting Retention

In the context of this study, students' retention of vocabulary knowledge was measured by a delayed post-test one week after the treatment, as explained earlier in Section 3.3 Research design. Many similar studies applied this measurement approach, for example, Keating (2008), Kim (2008), and Zou (2012, 2016, 2017). Specifically, the effectiveness of the tasks in promoting retention of target vocabulary knowledge was demonstrated by the participants' scores in the delayed post-tests (Hulstijn, 2001; Zou, 2012; Hu & Nassaji, 2016). The participants' MVKS scores in the delayed post-tests were set as the test variables and the task type as the grouping variable.

As presented in Table 26, the values in the "Equal variances assumed" row were the results of the independent samples t-test, as the sig. value of the Levene's test for equality of variances was 0.70,

greater than 0.05. The sig. value of the sig. (2-tailed) was 0.03, which was smaller than the significance level of 0.05, indicating that these two sets of data were significantly different. Similarly, the 95% confidence interval of the difference ranged from -2.47 to -0.09, excluding the value of 0, which also indicated significant difference between the mean scores of the participants who did Task 1 and Task 2 in the delayed posttest.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	*Sig.*	*t*	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	Lower	Upper
Equal variances assumed	.14	.70	-2.15	68	.03*	-1.28	.59	-2.47	-.09
Equal variances not assumed			-2.15	67.82	.03	-1.28	.59	-2.47	-.09

**p < 0.05*

Table 26: Results of the independent samples t-test of the participants' scores in the delayed posttest (Task 1 vs. Task 2)

Thus, in terms of promoting retention of the target vocabulary knowledge, Task 1 (Reading a text and answering reading comprehension questions on paper) was significantly less effective than Task 2 (Reading a text and answering reading comprehension questions in a digital game). This indicated that for the same learning task, learners performed better in retention of the target vocabulary knowledge in a digital game environment than in a conventional one.

4.5.4 Comparing Task 2 and Task 3 in Terms of Promoting Retention

As presented in Table 27, concerning the delayed post-test scores of the participants who did Task 2 and Task 3, the values in the "Equal variances assumed" row were the results of the independent samples t-test, as the sig. value of the Levene's test for equality of variances was 0.72, greater than 0.05. The sig. value of the sig. (2-tailed) was 0.001, which was smaller than the significance level of 0.05, indicating that these two sets of data were significantly different. Similarly, the 95% confidence

interval of the difference ranged from -3.16 to -0.83, excluding the value of 0, which also indicated significant difference between the mean scores of the participants who did Task 2 and Task 3 in the delayed posttest.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	*Sig.*	*t*	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	Lower	Upper
Equal variances assumed	.12	.72	-3.42	68	.001*	-2.00	.58	-3.16	-.83
Equal variances not assumed			-3.42	67.34	.001	-2.00	.58	-3.16	-.83

**p < 0.05*

Table 27: Results of the independent samples t-test of the participants' scores in the delayed posttest (Task 2 vs. Task 3)

Thus, in terms of promoting retention of the target vocabulary knowledge, Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) was significantly more effective than Task 2 (Reading a text and answering reading comprehension questions in a digital game). This indicated that for two different learning tasks (i.e., Task 2 and Task 3) with the same involvement load (i.e., moderate need, search, moderate evaluation), learners performed differently in a digital game environment.

4.6 Loss of Vocabulary Knowledge

I also investigated students' loss of vocabulary knowledge by examining the attrition rates of their post-test scores, which refer to the change of the three groups of participants' mean scores from the immediate post-test to the delayed post-test. For example, if a student's immediate post-test score is 10, and her/ his delayed post-test score is 8, the attrition rate is 0.2, i.e., (10 - 8)/10 = 0.2. I used the attrition rates as the dependent variables and task type as the fixed factor and examined the differences among the three groups of participants' attritions through a one-way ANOVA. The data met the basic assumptions for ANOVA. They were drawn independently of each other, normally distributed, and

shared homogenous variances (see Table 29).

As presented in the following Table 28, the mean attrition of the participants who did Task 1 was 0.21 (*SD* = 0.23), which was lower than the mean attrition of the participants who did Task 2, 0.24 (*SD* = 0.27). The mean attrition of the participants who did Task 3 was the highest among the three groups of participants, which was 0.29 (*SD* = 0.18).

	N	Attrition			
		M	*SD*	*Max*	*Min*
Task 1	35	0.21	0.23	1	0
Task 2	35	0.24	0.27	1	0
Task 3	35	0.29	0.18	0.6	0
Total	105	0.24	0.23		

Table 28: Descriptive statistics of the attritions

		Levene Statistic	*df1*	*df2*	*Sig.*
Attrition	Based on Mean	3.04	2	102	0.052
	Based on Median	2.08	2	102	0.13
	Based on Median and with adjusted df	2.08	2	88.15	0.13
	Based on trimmed mean	2.62	2	102	0.07

Table 29: Test of Homogeneity of Variances

			Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	(Combined)		0.11	2	0.05	1.02	0.36
	Linear Term	Contrast	0.11	1	0.11	2.03	0.15
		Deviation	0.001	1	0.001	.01	0.90
Within Groups			5.53	102	0.05		
Total			5.64	104			

Table 30: ANOVA Results

The results of the ANOVA showed no statistically significant difference among the three groups of participants' attritions. As shown in Table 30, the Sig value was 0.36, which was greater

than 0.05. The results of the multiple comparisons, as presented in Table 31, also indicated that the three groups of participants' attritions were not significantly different, as the sig. values were all greater than 0.05.

(I) Task Type	(J) Task Type	*Mean Difference (I-J)*	*Std. Error*	*Sig.*	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-0.03	0.05	1.00	-0.16	0.10
	3.00	-0.07	0.05	0.47	-0.21	0.05
2.00	1.00	0.03	0.05	1.00	-0.10	0.16
	3.00	-0.04	0.05	1.00	-0.18	0.09
3.00	1.00	0.07	0.05	0.47	-0.05	0.21
	2.00	0.04	0.05	1.00	-0.09	0.18

Table 31: Results of Multiple Comparisons

Such results indicated that the three groups of students had similar rates concerning their loss of vocabulary knowledge after the learning sessions. Regardless of their learning environments and learning tasks, all students forgot the target vocabulary knowledge with similar rates.

4.7 Interview Results

As for the qualitative data, which were collected from 30 participants who were interviewed after the immediate post-test, I analyzed and coded them to identify important features of the three tasks (i.e., Task 1: Reading a text and answering reading comprehension questions on paper; Task 2: Reading a text and answering reading comprehension questions in a digital game; and Task 3: Reading a text and inferring meanings of the underlined words in a digital game). Specifically, the important features included the following aspects:

1. What features motivated the students' vocabulary learning;
2. What features demotivated the students' vocabulary learning;
3. What features engaged the students in the learning processes;
4. What features directed the students' attention to the target vocabulary knowledge;
5. What features distracted the students' vocabulary learning;
6. What features the students liked;
7. What features the students disliked;
8. What features the students considered useful for their learning of the target vocabulary knowledge;
9. What features the students considered useless for their learning of the target vocabulary knowledge.

The coding results showed that the digital game learning environment indeed led to a positive effect on learning motivation. As shown in Extract 1, the participant reported that he made some efforts to learn how to play the game and found that the game was quite motivating (…*The game was interesting and motivating*…). All extracts were transcribed from the audio interview data and literally translated to English when the interviewees spoke Chinese. The respondent in Extract 2 had experience in digital game playing and found it was easy-to-play, which reduced the degree of tediousness in conventional vocabulary acquisition (…*I enjoyed the game. In this way, acquiring vocabulary is not a tedious process anymore*…). Extract 3 showed that the digital game could also promote the motivation of the learner who was a non-game player in terms of task-driven form (...*The goal of the game task stimulated me to complete the learning process*…)

Extract 1: *"Although I was not familiar with the operation of digital games at the beginning, I understood how to play the game after guidance and demonstration from the assistant. The game was interesting and motivating. It is a good way to learn new words. However, the cost of this method seems to be very high."*

Extract 2: *"I usually play mobile games. Thus, playing the digital game is easy for me. I enjoyed the game. In this way, acquiring vocabulary is not a tedious process anymore. Hope that a mobile version of the game can be developed in the future."*

Extract 3: "*I am not a fan of digital games, but the use of the game is similar to reading an interesting story. The goal of the game task stimulated me to complete the learning process. I think this way of learning is very interesting, and I want more chances of learning in this way.*"

Similar comments on the game can be identified from almost all interviewees who learned the target vocabulary through playing the game. All 20 students (10 who did Task 2: Reading a text and answering reading comprehension questions in a digital game and 10 who did Task 3: Reading a text and inferring meanings of the underlined words in a digital game) reported that they loved this digital game-based approach to learning, feeling it interesting, creative, and innovative. A total of 16 out of 20 students believed that digital game-based learning could release their feelings of anxiety, which was conducive to the learning of target knowledge. Moreover, all 20 students said that they wanted to have more chances to learn in this way.

Extract 4: *"I felt immersed in the digital game and forgot about other things."*

Extract 5: *"I did not feel any anxiety."*

Extract 6: *"I enjoyed the digital game environment. It is relaxing."*

Concerning Task 1: Reading a text and answering reading comprehension questions on paper, the 10 participants who were interviewed generally believed that reading comprehension was a common approach to language learning, and they were very familiar with this task type. Similar comments concerning the popularity of reading comprehension as a common task type or means of learning can be found among the 10 participants who did Task 2: Reading a text and answering reading comprehension questions in a digital game.

It is noteworthy that a total of 18 out of 20 students (10 who did Task 1 and 10 who did Task 2) admitted that they felt vocabulary learning through reading comprehension not as effective as other common vocabulary learning task types, with which they were also very familiar and had practiced many times in the past years of language learning (e.g., cloze exercises, inferring meanings of target words, sentence writing, etc.). For explicit explanations concerning this perception, 16 interviewees explained that they felt that reading comprehension did not specifically directed or guided them to focus on the target vocabulary, so their attention was generally on the whole context for comprehension. They rarely attempted to remember the forms of the target vocabulary or figure out the exact meanings of the target words. Thus, they did not spend additional efforts on building clear form-meaning links for the target words. This might be the main reason why they found it difficult to remember the target words' meanings and forms after completing the reading comprehension tasks.

Extract 7: *"I think vocabulary learning through reading comprehension is not very effective, compared to other methods. I do not normally focus too much on certain words while reading. It's a bit unnatural, as the main target is the context, not the vocabulary I think."*

Extract 8: *"I did not pay much attention to the forms of the underlined words."*

Extract 9: *"I was not aware that the purpose is to memorize the underlined words, but I knew they are important to the reading comprehension."*

All 10 participants who did Task 3: Reading a text and inferring meanings of the underlined

words in a digital game showed positive attitudes towards this way of learning. They regarded it both interesting and effective. A total of 7 interviewees out of 10 suggested that this game-based inferring task should be more commonly applied in language education. One student even commented that the digital game could vividly contextualize the background setting, and the inferring task could effectively direct learners' attention to the target knowledge, so this task was very useful for learning.

Extract 10: *"This task is very interesting and effective for learning. I think digital game-based vocabulary learning is great. I like the idea of playing to learn."*

Extract 11: *"I like the idea of asking NPCs to guide me to guess the meanings of target words based on the context. This makes the learning fun."*

Extract 12: *"The game situated me in vivid contexts. I enjoy learning in this way."*

4.8 Summary

In this section, I firstly reported the experimental results of the study. Generally, the results showed that the three vocabulary learning tasks were effective in terms of promoting both initial learning and retention of vocabulary knowledge, considering that the participants had very little prior knowledge of the target vocabulary. Significant differences among the three tasks were identified. I then presented the interview results with extracts. Table 32 is a summary of the quantitative and qualitative results.

	Quantitative results	Qualitative results
Can the ILH, in its present form, predict differences between digital and non-digital environments?	For the same vocabulary learning task with the same involvement load, learners performed better in a digital game environment than in a conventional one.	1) Students considered digital game-based vocabulary learning interesting and effective and showed positive attitudes towards this way of learning. 2) The game environment could reduce students' feelings of anxiety.
	For two different learning tasks with the same involvement load according to ILH in its present form (i.e., moderate need, search, moderate evaluation), learners performed	3) Reading comprehension did not specifically direct or guide students to focus on the target vocabulary. Students' attention was generally on the whole context for comprehension. They rarely attempted to remember the forms of the target vocabulary or figure out the exact meanings of the target words.

	differently in a digital game environment.	4) The inferring task could effectively direct learners' attention to the target vocabulary knowledge.

Table 32: Summary of the quantitative and qualitative results

Concerning the first research question, "For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one?", the research results indicated that, for the same task of reading a text and answering reading comprehension questions, learners performed better in both immediate learning and retention of vocabulary knowledge in a digital game environment than in a conventional one. The interview results indicated that students considered digital game-based vocabulary learning interesting and effective and showed positive attitudes towards this way of learning. The game environment could also reduce students' feelings of anxiety.

Concerning the second research question, "For two different learning tasks, which are similarly effective in terms of promoting vocabulary learning in a conventional learning environment, do learners perform differently in a digital game environment?", the research results indicated that, for Task 2: Reading a text and answering reading comprehension questions in a digital game and Task 3: Reading a text and inferring meanings of the underlined words in a digital game, learners performed differently. Although the two tasks induce the same involvement load (moderate need, search, and evaluation) and tend to promote vocabulary learning with similar effectiveness according to the ILH, the research results showed that Task 3 was significantly more effective than Task 2. The interview results also indicated that reading comprehension did not specifically directed or guided students to focus on the target vocabulary, so students' attention was generally on the whole context for comprehension, and they rarely attempted to remember the forms of the target vocabulary or figure out the exact meanings of the target words. Whereas the inferring task could effectively direct learners' attention to the target knowledge, so students spent more additional efforts on building clear form-meaning links for the target words than those who learn with the reading comprehension task.

According to the ILH, it seemed that all three tasks induced the same involvement load. However, the empirical results showed statistically significant differences among the three tasks in terms of their effectiveness of promoting initial learning and retention of target vocabulary knowledge. One representative competing theory of the ILH is the TFA, which was proposed by Nation and Webb

(2011) on the basis of the ILH. Hu and Nassaji (2016) investigated two tasks, “reading a text and answering reading comprehension questions on paper” and “reading a text and inferring meanings of the underlined words on paper”. They considered that these two tasks had the same TFA score. Thus, it seemed that TFA also cannot explain the empirical results of this study well. In the following chapter, I will discuss possible reasons concerning why the three tasks appeared to induce the same involvement load and have the same TFA score but promoted vocabulary learning with significantly different effectiveness.

CHAPTER 5: DISCUSSION

5.1 Introduction

In this chapter, I discuss the research results from three main perspectives. Firstly, I analyze the effectiveness of digital game-based vocabulary learning in Section 5.2. Two main factors that play important roles are analyzed: increase of motivation and decrease of anxiety. Then I discuss the participants' engagement with vocabulary while doing the three vocabulary learning tasks in Section 5.3. After these analyses, I revisit the ILH and TFA from the perspective of digital game-based vocabulary learning in Section 5.4, attempting to suggest how these two representative hypotheses that attempt to evaluate the effectiveness of word-focused learning tasks can be applied in digital game environments.

5.2 Effectiveness of Digital Game-Based Vocabulary Learning

As reported in Section 4.5 Effectiveness of three vocabulary tasks, the effectiveness of all three tasks in terms of promoting initial learning and retention of target vocabulary knowledge were statistically significant. Task 2: Reading a text and answering reading comprehension questions in a digital game and Task 3: Reading a text and inferring meanings of the underlined words in a digital game were significantly more effective than Task 1: Reading a text and answering reading comprehension questions on paper. These results of the post-tests were consistent with many previous studies on digital game-based vocabulary learning (Zou et al., 2019). For example, Tsai and Tsai (2018) conducted a meta-analysis study on 26 empirical studies on digital game-based second-language vocabulary learning and found rich empirical evidence in support of the use of digital games in vocabulary learning. Firstly, when the experimental groups learned with video games, while the control groups learned with non-game-related activities, a large overall effect size was found. Secondly, when the experimental groups learned with video games with specific features added or changed, while the control groups learned with the base versions of the games, a medium overall effect size was reported. Thirdly, when the experimental groups learned through playing a digital game, while the control groups learned via conventional media, a medium to large overall effect size was found. Fourthly, when both experimental and control groups played the same digital game but were grouped by nongame-related variables, non-significant effect size was found.

Similarly, Chiu et al.'s (2012) meta-analysis of 14 studies on digital game-based vocabulary learning reported that digital games were significantly effective for vocabular knowledge development.

Hwang and Wang (2016) developed a digital role-playing game to help students learn English vocabulary about daily life and integrated flash cards, cloze, and multiple-choice exercises as supplementary exercises into the game. Their investigation on the students' vocabulary learning performance and behaviors also showed that vocabulary learning through playing a digital role-playing game could lead to better learning performance and higher engagement in the learning processes, compared to the conventional learning approach (Hwang & Wang, 2016). Pan (2017) also found that students who learned with a motion-sensing game performed significantly better than those who learned in a conventional non-game approach. Additionally, Sylven and Sundqvist (2012) found that gameplay could lead to improvement of language learners' proficiency levels.

5.3 Digital Game Environment and Motivation

Concerning the digital game-related features, which may play important roles in leading to effective vocabulary learning, the results of this study indicated that the digital game environment was conducive to the increase of learner motivation. As reported in Section 4.8 Interview results, almost all interviewees who learned through playing the digital role-playing game showed positive attitudes towards the digital game environment and felt motivated to learn in this way. Most students also described the game-based learning environment as interesting and creative. Similar results have been reported in many previous studies on digital game-based learning.

As reviewed previously in Chapter 2 Literature Review, there is rich empirical evidence in support of the effectiveness of game-based learning. Papastergiou (2009, p.10) has shown that a simple digital game, which is integrated with curricular contents, can be "more effective [at] promoting the knowledge acquisition of computer memory concepts and more motivational for students than the non-game approach." Papastergiou (2009) also pointed out that the digital game learning environment can be motivational and effective for all students despite the greater appreciation and experience on digital games of certain groups such as male students. Erhel and Jamet (2013) investigated the impact of different types of instructions and feedback in digital games and found that digital games with different instructions and feedback can promote learning and motivation. In addition, some recent research studies have further revealed the positive effects of digital game environments on motivation. Proulx et al. (2018) explained how a digital game promoted extrinsic and intrinsic motivation. Further, Liao et al. (2019) found that collaborative digital game-based learning can promote the intrinsic motivation of students. Zou et al. (2019) reviewed 21 studies for digital game-based vocabulary learning and found that most studies noted positive effects on motivation via digital game-based learning.

In sum, based on the interview data, I found one potential reason for the finding that the students who did Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) outperformed those who did Task 1 (Reading a text and answering reading comprehension questions on paper). The digital game environment was conducive to the increase of learner motivation. There is rich empirical evidence in support of this argument in the literature. Erhel and Jame (2013), Hung et al (2014), Liao et al. (2019), and Papastergiou (2009) all reported that the digital game learning environment can better motivate learners during their learning process than conventional learning environments.

5.4 Digital Game Environment and Anxiety

The interview data also suggested that the digital game environment could reduce students' anxiety of English learning. A common challenge for language learners is that many students feel anxious while learning foreign languages, and it is important to provide them with learning environments where the atmosphere and contexts are relaxing and comfortable (Yang et al., 2020). In this study, I found that around 80% of the interviewees who did Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) believed that the digital game environment could reduce their feelings of anxiety.

Similarly, many previous studies reported that their participants who learned with digital games had better performance than those who learned in conventional learning environments and experienced lower levels of anxiety. Young and Wang (2014) believed that digital game-based vocabulary learning was very helpful for speaking practices of students with low language proficiency levels, as they might suffer more from anxiety and benefit more from the game environments, compared to students with high language proficiency levels. Wei et al. (2018) also found that integrating personalized assistance strategies into games could create a low anxiety learning environment and lead to effective vocabulary learning. Similarly, Wu and Huang (2017) considered it important to release learner anxiety and found digital games useful in terms of reducing anxiety and enhancing students' engagement.

In sum, the results of this study indicated another potential reason for the finding that Task 2 and Task 3 were significantly more effective than Task 1. The digital game environment was conducive

to the decrease of learner anxiety. The results of many empirical studies also suggested that digital games could release learner anxiety, which was conducive to effective learning (e.g., Wei et al., 2018; Wu & Huang, 2017; Yang et al., 2020; Young & Wang, 2014).

5.5 Engagement with Vocabulary

As explained in Section 2.3 Vocabulary Learning Theories, many researchers have attempted to propose theories, hypothesis, or models to explain why certain tasks are more effective than others in terms of promoting vocabulary learning, and what features of vocabulary learning tasks tend to lead to successful and effective learning of vocabulary knowledge. Schmitt (2008) proposed the term "engagement with vocabulary" and summarized nine factors that would lead to engagement, including 1) increased frequency of exposure; 2) increased attention; 3) increased noticing; 4) increased intention to learn; 5) a requirement to learn; 6) a need to learn or use the target vocabulary; 7) increased manipulation of the target vocabulary and its properties; 8) increased amount of time on learning; and 9) increased amount of interaction with the target vocabulary (Schmitt, 2008).

I found that this concept "engagement with vocabulary" could well account for my research result that Task 2 (Reading a text and answering reading comprehension questions in a digital game) was significantly less effective than Task 3 (Reading a text and inferring meanings of the underlined words in a digital game). The interview data collected from the 20 participants who did Task 2 and Task 3 suggested that Task 3 induced more engagement with vocabulary than Task 2.

A total of eight out of 10 interviewees who did Task 2 reported that they did not re-read the text and check relevant parts when they answered the comprehension questions, as they felt it unnecessary. Seven out of 10 students also admitted that they did not notice that the reading comprehension questions were closely related to the underlined target words, and they were not aware that each question associated to one target word. Thus, they did not spend additional time or efforts on the words. All 10 students knew that the ten underlined words were important for the completion of the reading comprehension task, however, only two of them attempted to re-read the relevant text and answer the comprehension questions based on the associated contexts for the target words. Most students did not intentionally focus on the target vocabulary while completing the relevant comprehension exercises or feel the need of doing so.

Nevertheless, all 10 interviewees who did Task 3 knew that the 10 inferencing exercises were

for the 10 target words, as they were specifically asked what these words meant in the contexts of the reading text. Thus, they were clear about the learning goal of this task, the 10 target words' meanings and forms. Moreover, all students reported that they re-read the text and checked relevant parts when they did the inferencing exercises, as they felt it necessary and important. This is very different from what the students who did Task 2 reported. Moreover, a total of nine out of the 10 interviewees who did Task 3 mentioned that these inferencing exercises directed them to pay attention to the target vocabulary, the contexts where they were used in the reading text and attempted to identify clear meanings of the target words. On the contrary, seven out of the 10 interviewees who did Task 2 admitted that they only had very vague ideas about the meanings of the target words, as they did not feel the need of figuring out the exact meanings of the target words. It seemed that knowing what the contexts meant was sufficient for the completion of Task 2, and knowledge of the exact meanings of the target words was not a must.

In sum, the results indicated that Task 3: Reading a text and inferring meanings of the underlined words in a digital game induced more engagement with vocabulary than Task 2: Reading a text and answering reading comprehension questions in a digital game. Specifically, the students who did Task 3 were aware of the vocabulary learning goal. They tended to intentionally pay attention to the target vocabulary. Compared to the students who did Task 2 and only had vague ideas about the meanings of the target words, the students who did Task 3 attempted to identify clear meanings of the target words based on their contexts by re-reading the relevant texts. Moreover, the students who did Task 3 regarded knowing the exact meanings of the target words a must, while those who did Task 2 did not.

5.6 Flow theory

As reviewed in Section 2.5.1, students tend to be more concentrated on learning and are more willing to spend longer time engaged in learning when they experience flow, so the flow experience is likely to lead to effective learning (Wang & Chen, 2010; Wei et al., 2018). According to Salen and Zimmerman (2004), Schell (2008), and Song and Zhang (2008), three factors are essential for the creation of the flow experience: clear goals, immediate feedback, and challenge-skill balance. In this study, the learning goals of Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) were clear. The digital role-playing game also provided students with immediate, clear, and unambiguous feedback. Moreover, the interview results indicated that the participants' language skills

and the challenge of completing the reading comprehension and inferencing exercises were balanced. Thus, most participants experienced the flow states while learning through playing the games. However, the students who learned with Task 1 (Reading a text and answering reading comprehension questions on paper) did not receive immediate feedback while writing down the answers to the comprehension exercises on paper, which was different from those who submitted answers to the game system and received automatic feedback.

The interview results also indicated that the students who learned with Task 2 and 3 assumed roles while playing the games, had stronger sense of control, and felt more immersed in the learning than those who learned with Task 1 and answered reading comprehension questions on paper. Such sense of control and feelings of immersion also led to the flow states of the students who learned with Task 2 and 3.

The interview results did not suggest any major differences between the students who learned with Task 2 and 3 concerning their flow experience. The main differences were between these two groups of students and those who learned with Task 1.

5.7 Noticing and Form-focused Instruction

According to the noticing hypothesis, students learn a word better when s/he notices it than when s/he pays little attention to it (Schmidt, 1995). Also, as reviewed in Section 2.3.1, the concept of form-focused instruction contends that planned instructional activity leads to successful vocabulary learning (Ellis, 2001). From the perspectives of noticing and form-focused instruction, Task 3 (Reading a text and inferring meanings of the underlined words in a digital game) involves form-focused instruction and intentionally guides learners to notice the target words, while Task 2 (Reading a text and answering reading comprehension questions in a digital game) did not.

The interview results also revealed that the reading comprehension task did not specifically direct or guide students to focus on the target vocabulary, so their attention was generally on the whole context for comprehension. They rarely attempted to remember the forms of the target vocabulary or figure out the exact meanings of the target words. Thus, they did not spend additional efforts on building clear form-meaning links for the target words. Nevertheless, the inferencing exercises asked students to guess meanings of the target words from context, so more noticing was involved, and students intentionally pay attention to the context of the target words. These attempts might be the

main contributor to their learning of the words.

Therefore, the involvement of noticing and form-focused instruction in Task 2 and Task 3 seemed to be an important difference between them and Task 1. The students also reported in the interviews that they felt noticing and form-focused instruction helpful for effective vocabulary learning.

5.8 Revisiting ILH and TFA from the Perspectives of Digital Game-Based Vocabulary Learning

Both ILH and TFA are proposed for the evaluation of task effectiveness in conventional learning environments. They are from the dimension of task design, aiming to help teachers select appropriate task types for vocabulary teaching. They are not proposed from the perspective of learners, so learner factors are not the key elements when teachers evaluate the involvement load or technique features of a vocabulary learning task. In other words, for the evaluation of the involvement load or the analysis of the technique features of a vocabulary learning task, it is not necessary to take learner factors into account. Nevertheless, this does not necessarily mean that ILH and TFA regard learner factors unimportant for vocabulary learning; it is because they are proposed from the dimension of task design and focus mainly on task features.

As discussed in 4.7 Task Effectiveness in Terms of Promoting Retention and Section 4.9 Summary, neither ILH nor TFA could account for the results of this study. This is likely because the participants of this research did Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3 (Reading a text and inferring meanings of the underlined words in a digital game in game environments). The change of learning environments from the paper-based mode to the digital game-based mode seemed to induce different involvement loads and task features. Firstly, learner motivation generally tends to be higher in the digital game-based mode than the paper-based mode. Secondly, learner anxiety seems to be higher in the digital game-based mode than the paper-based mode.

5.8.1 Applying ILH to Digital Game-Based Vocabulary Learning

I suggest that teachers, learners, and researchers consider the following revisions when they apply ILH to digital game-based vocabulary learning. Firstly, consider adding one more degree of prominence to “need” when evaluating the involvement load of a task in digital game environments. Take Task 2 of this study as an example. I suggest that its task-induced involvement load should be adjusted from “moderate need, search, and moderate evaluation” to “strong need, search, and moderate evaluation”, because it is of the digital game-based mode. This being the case, the result that Task 2

promoted more effective learning than Task 1 can be explained, as the adjusted involvement load of Task 2 (2+1+1) is higher than Task 1 (1+1+1).

Secondly, consider adding one more degree of prominence to "search" when learners are required to infer exact meanings of target words based on the contexts. So, when learners infer meanings of the contexts where target words are integrated, the involvement load of search is moderate; while when learners infer meanings of the target words, the involvement load of search is strong. Take Task 3 of this study as an example. I suggest that its task-induced involvement load should be adjusted from "moderate need, search, and moderate evaluation" to "strong need, strong search, and moderate evaluation", as it is of the digital game-based mode, and it requires learners to infer meanings of target vocabularies. Consequently, the result that Task 3 promoted more effective learning than Task 2 can be explained, as the adjusted involvement load of Task 3 (2+2+1) is higher than Task 2 (2+1+1).

5.8.2 Applying TFA to Digital Game-Based Vocabulary Learning

Based on the empirical results of this study and the relevant literature on digital game-based vocabulary learning, I also propose one suggestion: adding one criterion in the category "motivation", "Is the activity game-based?"

Concerning the three tasks of this study, I suggest that their TFA scores should be 4, 5, and 7, respectively. The details are presented in Table 33. The first question, "Is there a clear vocabulary learning goal?" refers to whether the learning activity has a clear vocabulary learning goal or not. If there is a clear vocabulary learning goal in the learning activity, it is more likely to motivate learners to complete this learning activity than those learning activities without a clear vocabulary learning goal. All three tasks have clear vocabulary learning goals, as the participants who did these tasks in this study reported that they knew these were vocabulary learning tasks.

The second question, "Does the activity motivate learning?" refers to whether the format of the learning task contains motivational elements, which can be further categorized into (i) game-based elements such as games or puzzles; (ii) challenging elements: the learners feel that the learning activity is a challenge for them; and (iii) elements for awareness of learning achievements such as a progress bar or word cards. All three tasks are given one score. The three groups of students all felt the activity motivating, as the topic of the reading text "Procrastination" was closely related to their daily life, and

many interviewees said that they felt this reading text was written for them. Procrastination was a common problem among these students, and everyone tended to have great empathy with the story.

For the third question, "Do the learners select the words?", this criterion is given a score of zero if learners are required to learn the specified target words. The criterion is given a score of "1" if learners can select their preferred target words or the learners give attention to the words due to their interests. For example, the learner may make a note of some words when she is reading an interesting magazine. All three tasks are given zero score. The target words are selected by the task designer, me, not them.

For the fourth question, "Is the activity game-based?", Task 2 (Reading a text and answering reading comprehension questions in a digital game) and Task 3(Reading a text and inferring meanings of the underlined words in a digital game) are of the digital game-based mode, so they are given one score. However, Task 1: Reading a text and answering reading comprehension questions on paper is given zero score.

The fifth question, "Does the activity focus attention on the target words?" is in the dimension of Noticing. This criterion refers to whether there is an explicit noticing of the unknown target words during the learning process or not. For example, if the target words are in bold font, this criterion is given one point. Otherwise, a score of zero is given. Task 1 and Task 2 are given zero score, while Task 3 one score. As reported by the interviewees who did Task 1 and Task 2, they did not pay special attention to the target words as they felt it unnecessary. However, those who did Task 3 re-read the relevant text to infer the exact meanings of the target words, thus attention on the target words was induced by this activity.

Similarly, the sixth question asks, "Does the activity raise awareness of new vocabulary learning?". Different from the above criterion, which aims to check whether there is an explicit notice on the unknown target words, this criterion refers to the explicit notice of something about the learning of the target words. For example, if the learner identifies the words from a short essay or uses them in an original context, the learning activity is deemed to raise awareness of new vocabulary learning. Therefore, the score of this criterion will be given one point. Task 1 and Task 2 are given zero score, while Task 3 one score. The interview data indicated that only the students who did Task 3 were aware that the inferencing exercises were associated with new vocabulary learning, while those who did Task

1 and Task 2 did not know that the 10 reading comprehension questions were closely related to the 10 target words.

Component	Criteria	Task 1	Task 2	Task 3
Motivation	1) Is there a clear vocabulary learning goal?	1	1	1
	2) Does the activity motivate learning?	1	1	1
	3) Do the learners select the words?	0	0	0
	4) Is the activity game-based?	0	1	1
Noticing	5) Does the activity focus attention on the target words?	0	0	1
	6) Does the activity raise awareness of new vocabulary learning?	0	0	1
	7) Does the activity involve negotiation?	0	0	0
Retrieval	8) Does the activity involve retrieval of the word?	0	0	0
	9) Is it productive retrieval?	0	0	0
	10) Is it recall?	0	0	0
	11) Are there multiple retrievals of each word?	0	0	0
	12) Is there spacing between retrievals?	0	0	0
Generation	13) Does the activity involve generative use?	0	0	0
	14) Is it productive?	0	0	0
	15) Is there a marked change that involves the use of other words?	0	0	0
Retention	16) Does the activity ensure successful linking of form and meaning?	0	0	0
	17) Does the activity involve instantiation?	0	0	0
	18) Does the activity involve imaging?	0	0	0
	19) Does the activity avoid interference?	1	1	1
Total score		3	4	6

Table 33: My suggested TFA scores

The seventh question asks, “Does the activity involve negotiation?”. The criterion refers to whether the learning activity involves any negotiation about the target words. According to the research conducted by Newton (1995), negotiating the meaning of the target words was more effective for learning than target words without any negotiation. If the learning activity involves the negotiation of the target words, the score of this criterion will be given one point. The three tasks are reading based, and no negotiation is involved. All three tasks are given zero score.

In the dimension of retrieval, the eighth question asks, “Does the activity involve retrieval of the word?”. This criterion refers to whether there is an activity involving the retrieval of the word meaning or form. The word meaning and the word form correspond to the receptive and productive vocabulary knowledge, respectively. In other words, this criterion is essentially to check whether the activity involves the receptive/productive retrieval. The activities involved these two types of retrievals

are given one point for each criterion.

The ninth question asks, “Is it productive retrieval?”. As mentioned in the previous criterion, which aims to check whether the activity involves the receptive/productive retrieval, the current criterion is a follow-up step to confirm whether the involved retrieval type is productive as this type of retrieval has been found to be more difficult than simple receptive retrieval according to Yanagisawa (2016). Thus, one point is given for the previous and current criteria if the learning activity involves the productive retrieval. For example, the learning task "create an original sentence which uses the target words" involves productive retrieval.

The tenth question asks, “Is it recall?”. Recall refers to whether the learning activity involves the retrieval of the word meaning or form from the memory or not. Note that recall is different from recognition, which is a process to identify the meaning or form from the reading materials or the sentences rather than the learner’s memory. If the learning activity involves the recall of the target words, it is given one point.

The 11th question asks, “Are there multiple retrievals of each word?”. This criterion refers to whether the learning activity requires the learners to conduct more than one retrieval of the word meaning and form for the target words. For example, learning activities with multiple exercises, including multiple choices on the word meaning and word completion for each target word, involves multiple retrievals. If the activity has this characteristic, the score of this criterion is awarded one point.

The 12th question asks, “Is there spacing between retrievals?”. This criterion refers to the multiple retrieval of word form and meaning for each word in the learning activity. Here, the criterion checks whether there is a certain time interval between multiple retrievals in the learning task. According to Baddeley (1992), time intervals between multiple retrievals of target words can be more facilitative for promoting word retention than multiple retrievals without time intervals. If the learning activity involves time intervals between the multiple retrievals of target words, it is given one point.

In the dimension of generation, the 13th question asks, “Does the activity involve generative use?”. This criterion refers to whether there is generative use of the target words in the learning activity. The generative use of the target words improves the learning effectiveness and promotes word retention according to previous studies (Joe, 1998; Schmitt, 2008). There are two types of generative

use of target words, which are receptive generative use and productive generative use. The former denotes that the learner has met a target word that was used in a certain way, while the latter refers to the learner using a target word in a new way. If the learning activity involves one of the two types of generative use of the target words, it is given one point.

The 14th question asks, "Is it productive?". This criterion refers to the generative use of the target words in the learning activity. According to Joe (1998) and Laufer (2009), the productive generative use of target words requires more mental efforts, which leads to more effective learning of words than receptive generative use. Therefore, this follow-up criterion confirms whether the type of generative use is productive or not. If the learning activity involves productive generative use of the target words, one point is given.

The 15th question asks, "Is there a marked change that involves the use of other words?". This criterion is a further follow-up to the previous one and only applicable for productive generative use. For this criterion, the involvement of the use of other words indicates that the degree of prominence of generativity is high. For example, if the learning activity involves sentence production based on the target words, the score of this criterion will be given one point as the sentence production involves the use of other words in the productive generative use.

In the dimension of retention, the 16th question asks, "Does the activity ensure the successful linking of form and meaning?". This criterion checks whether the learning activity establishes a linkage between the form and meaning of the target words or not. The linkage between form and meaning can be enhanced by the successful retrievals of form or meaning but can be undermined by errors. If the learning activity is very likely to build the linkage between form and meaning, it is given one point. Specifically, some learning activities, such as consulting dictionaries, which require the retrieval of the form and meaning of target words at the same time, can get one point for this criterion, whereas other learning activities, such as fill-in-the-blanks, which separate the retrieval of form and meaning of target words, get a score of zero.

The 17th question asks, "Does the activity involve instantiation?". This criterion checks whether the learning activity involves learners noticing an instance of using the target words in a meaningful context or not. One of the potential reasons that instantiation is important for word retention is that the use of target words in a meaningful context can be associated with a visual situation. The visual

situation of the use of the target words can promote the retention of the target words. Therefore, if the learning activity involves instantiation of the target words, one point is given.

The 18th question asks, "Does the activity involve imaging?". This criterion checks whether the learning activity involves the learner seeing visual images of the meanings of target words or deliberately establishing a linkage between visual images and the meanings of target words or not. One obvious example of a learning activity that involves imaging, is showing pictures of the meanings of target words by using an image search engine like Google Image. If keywords (target words) are input into the search engine, the results contain many images, which reflect the semantic meanings of the input. If the learning activity involves imaging of the target words, it is given one point.

The 19th question asks, "Does the activity avoid interference?". This criterion checks whether the learning activity involves the learning of several unknown words that share similar/opposite meanings, or not. For example, if the learning activity puts a set of synonyms together, the learning difficulty will be significantly increased according to previous studies (Warring, 1997; Erten and Tekin, 2008). Therefore, it is important to ensure that vocabulary acquisition will not be interfered with by other words. If the learning activity avoids interference from other synonyms or words of opposite meaning during the acquisition of target words, it is given one point.

For the five questions in the category "retrieval", the three questions in the category "generation", and the first three questions in the category "retention", all three tasks are given zero scores. While for the last question of retention, "Does the activity avoid interference?" all three tasks are given one score.

Based on these TFA scores (i.e., 4 for Task 1, 5 for Task 2, and 7 for Task 3, respectively), it seems that the adjusted TFA could account for the research results appropriately. Compared to ILH, TFA is more detailed and comprehensive. Some researchers compared ILH and TFA and suggested that TFA was more accurate than ILH (e.g., Chaharlang & Farvardin, 2018; Gohar et al., 2018; Hu & Nassaji, 2016; Zou et al., 2020). This is likely because TFA includes more items than ILH, because of which, it appears a bit redundant, and some criteria seem overlapped.

5.9 Summary

In this chapter, I discussed the research results from several aspects. Firstly, the effectiveness

of digital game-based vocabulary learning. Then, I discussed the digital game environment and motivation and anxiety. I argue that digital games could lead to effective vocabulary learning as they are able to increase learner motivation but reduce learner anxiety. In Section 5.5, I discussed learner engagement with vocabulary and argued that the task of asking learners to infer the exact meanings of target words could increase learner engagement with vocabulary, compared to reading comprehension. I also discussed the results from the perspective of the flow theory, noticing and form-focused instruction in Section 5.6, 5.7, and 5.8, respectively.

Based on these analysis and discussions, I suggest that the application of ILH to digital game-based vocabulary learning should consider adding one more degree of prominence to "need" when evaluating the involvement load of a task in digital game environments. I also suggest that when learners are required to infer exact meanings of target words based on the contexts, the involvement load of search should be strong, rather than moderate. Concerning the application of TFA to digital game-based vocabulary learning, I suggest adding one criterion in the category "motivation", "Is the activity game-based?"

CHAPTER 6: IMPLICATIONS

6.1 Introduction

In this chapter, I discuss the implications of the findings based on the experimental results from two perspectives. The first perspective concerns the theoretical contributions of the current study to the field of digital game-based vocabulary learning and L2 vocabulary acquisition. The second perspective is the pedagogical implications that digital game-based learning has on vocabulary learning and L2 vocabulary acquisition.

The remaining sections of this chapter are structured as follows. Section 6.2 discusses the theoretical implications of the main findings (i.e., the three suggestions) in Chapter 5. In Section 6.3, the implications for vocabulary learning in terms of classroom pedagogical practices is introduced. Section 6.4 summarizes the implications discussed in this chapter.

6.2 Theoretical Implications

From the theoretical perspective, the findings have uncovered how the conventional allocation on the Need component of ILH for vocabulary learning tasks in a digital game learning environment is inappropriate. The Need component of ILH is based on the notion that "strong need is intrinsically motivated, while moderate need is imposed by an external agent" (Zou, 2012, p. 2). As research studies have typically required participants to attend to learning procedures, the Need component of ILH of conventional vocabulary learning tasks were generally considered as "moderate need." However, this seems inappropriate for digital game-based vocabulary learning. Like the results of this study, many previous studies (e.g., Hung et al., 2014; Erhel and Jamet, 2013; Liao et al., 2018; Papastergiou, 2009; Proulx et al., 2018; Zou et al., 2019) also argued that learners who learned through playing digital games were more motivated than those who learned in conventional ways. Thus, I suggest that digital game-based vocabulary learning tasks induce “strong need” (the 1st suggestion). I also suggest that when learners are required to infer exact meanings of target words based on the contexts, the involvement load of search should be strong, rather than moderate (the 2nd suggestion).

The third suggestion advocates the use of TFA for estimating the effectiveness of vocabulary learning tasks. Compared to ILH, it is more detailed and comprehensive, so the evaluation results seem accurate. Concerning the application of TFA to digital game-based vocabulary learning, I suggest adding one criterion in the category “motivation”, “Is the activity game-based?”

6.3 Pedagogical Implications

In this section, the pedagogical implications are discussed based on the perspectives of students, teachers, and educational game designers.

Implications for students. From the perspective of students, the findings of the current studies can be mainly used for their own vocabulary acquisition. Specifically, students should play digital games to learn new words as many studies (Erhel and Jame, 2013; Hung et al., 2014; Liao et al., 2019; Papastergiou, 2009; Proulx et al., 2018; Zou et al., 2019) have revealed that digital games can increase learning motivation. Furthermore, the results of the current study demonstrate that (i) tasks performed in a digital game environment are more effective than the same task performed on paper (T2 vs. T1; T2 > T1) and (ii) tasks performed in a digital game environment are more effective than the tasks performed with the same ILH involvement loads on paper (T3 vs. T1; T3 > T1) in promoting word retention and vocabulary acquisition. In addition, a new criterion, *"Is the activity game-based*?*"* in the revised version of TFA showed that learning tasks using digital game-based learning can be more effective than the same tasks without digital game-based learning in promoting word retention and vocabulary acquisition.

Students are also suggested to seek out digital role-playing games for their learning because digital games have been shown to be effective for vocabulary learning in several studies (Hwang & Wang, 2016; Sylvén & Sundqvist, 2012). Sylvén and Sundqvist (2012) found that there is a positive relationship between the frequency of playing online role-playing games and language proficiency development. Hwang and Wang (2016) compared two different guiding strategies (i.e., single-loop and double-loop learning) for English vocabulary acquisition in situated computer games developed by RPGMaker and found that both strategies were more effective than non-game learning tasks. The findings from the current study also demonstrate that role-playing games are an effective learning activity. Notably, the digital game environment in learning tasks T2 and T3 developed by RPGMaker was a role-playing game.

Implications for teachers. From the perspective of teachers, the findings of this study provide them with implications on how to adopt digital game-based vocabulary learning in classroom activities. The first suggestion is to employ a game-based teaching approach for vocabulary learning. According to the review conducted by Zou et al. (2019), the game-based teaching approach has been found to be more effective than conventional classroom teaching approaches for vocabulary learning

by all studies related to digital game-based vocabulary learning in terms of short-term word retention, while eight of 11 studies found that it is more effective in terms of long-term word retention. This finding is also supported by the current study, which showed that the tasks performed on paper are less effective than the similar tasks performed in a digital game environment. However, there are still some studies indicating that digital game-based vocabulary learning might not be more effective. For example, Hung et al. (2015) found that high-achieving students control (i.e., traditional learning) and experimental (i.e., digital game-based learning) groups performed similarly in terms of word retention, although the performance of low-achieving students in the experimental group outperformed the control group significantly. A possible explanation for this is that the high-achieving students were already self-motivated, and the digital game may not have been able to further improve their learning motivation. Two studies (Calvo-Ferrer, 2017; Huang & Huang, 2015) also observed that the students in the group that used DGBVL performed similarly to the group of students who employed the traditional learning approach. Surprisingly, Young and Wang (2014) found a negative result in long-term vocabulary acquisition for a group of students using DGBVL compared to a group using traditional learning approaches. A possible explanation is that the game embedded in Young and Wang's study (2014) was similar to an exercise rather than a digital game as there were no game elements such as game flow and storyline.

Another implication for teachers is to investigate game acceptance before conducting digital game-based vocabulary learning and provide sufficient support to guide students on how to play digital games. First, acceptance of game-based learning is an important factor for successful learning. Bourgonjon, De Grove, De Smet, Van Looy, Soetaert, and Valcke (2013) proposed an acceptance model for digital game-based learning for teachers. Cheng, Lou, Kuo, and Shih (2013) developed a similar model for student acceptance of digital games. If teachers or students do not accept the use of digital games as a learning tool, they may be reluctant to view digital game-based learning as a viable learning approach. To address this problem, a common practice is to involve students in the development process and conduct a pilot study as I did in the current study. Second, providing sufficient support to guide students on how to play games is an essential ingredient for successful digital game-based vocabulary learning. All, Castellar, and Van Looy (2016) showed that various types of support, including procedure support, technical support, and game-playing, can facilitate the effectiveness of digital game-based learning. In the current study, I arranged for an assistant to help the participants to address technical and game-playing problems and provide guidance on game-playing to train participants before conducting the actual learning activity. The qualitative data showed

that some participants really benefited from this step.

The final implication for teachers concerns the need to keep a high level of motivation among students. Digital game learning, by employing various strategies, including competition and diversifying game contents, achieves this objective. Although this implication may not be directly supported by the findings of the current study, some participants told me that they became a bit tired of the digital game when they repeatedly played it without any new content and playing strategies. Annetta (2010) reviewed various educational digital games and summarized six key features, including "identity, immersion, interactivity, increasing complexity, and instructional," for engaging and motivating players. Although digital games are directly relevant to these six dimensions, teachers can still adopt various strategies for enhancing these features. For example, teachers may organize competitions among different groups of students using a scoreboard to enhance interactivity. Teachers can also provide additional guidelines and instructions to improve the instructional components of digital games when they are insufficient.

Implications for educational game designers. From the perspective of educational game designers, there are also some implications regarding how to develop successful digital games for vocabulary learning and these can be categorized into two levels. The first level concerns how to design an engaging game. In the current study, the qualitative data revealed that participants have positive comments about the storylines. More generally, Ge and Ifenthaler (2017) proposed a three-dimension model of motivation and engagement for educational games, which included emotional, behavioral and cognitive dimensions. Specifically, there are 15 features, including fascinating storylines, conceptual play space, uniqueness and thrilling experiences, immediate positive feedback, challenges with more controls and fewer constraints, uncertainty or unexpectedness, relatedness, connections, emotional associations, incentives, and choices, controls, and decision making, for a motivating and engaging game design (Ge & Ifenthaler, 2017). Accordingly, digital game designers need to take these features into consideration when designing games.

The second level of implications concerns how to integrate vocabulary learning into a digital game. The findings of the present study revealed that the TFA provides a good theoretical framework for estimating the effectiveness of vocabulary learning tasks which can help digital game designers. In addition to the use of TFA, Calvo-Ferrer (2017, p. 265) argued that "cognitive engagement plays a more important part in L2 vocabulary acquisition than the feeling of fun during gameplay." Digital game designers may thus consider how to infuse cognitive engagement into their game designs.

According to the three-dimensional model proposed by Ge and Ifenthaler (2017), cognitive engagement is significantly improved if the learning tasks and game content can increase the degree of uncertainty and involve decision making. Furthermore, Zou et al. (2019) suggested that interactivity plays an essential role in digital game-based vocabulary learning. Accordingly, game designers should consider how to facilitate interactivity in vocabulary learning games, for example, by adding competitive functions.

6.4 Summary

In this section, the implications based on the findings of this study from two main aspects have been discussed. The first aspect concerns the theoretical implications, which are elaborated based on the three main suggestions described in Chapter 5. Specifically, the first implication, based on 1st suggestion, is that the conventional allocation on ILH neglected the fact that the learning tasks, which have been well-designed and integrated with some motivational elements (e.g., games and digital game, collaboration), may intrinsically motivate learners. The second implication based on 2nd suggestion is that we can generalize that TFA has better predictive power than the ILH from conventional learning to digital game-based learning for all vocabulary learning tasks. The third implication based on 3rd suggestion is the possibility of extending the original TFA theory by adding suitable criteria.

The second aspect concerns the pedagogical implications, which were discussed based on the perspectives of students, teachers, and educational game designers. From the perspective of the students, they are suggested to (i) employ digital games as one of their learning tools for vocabulary acquisition and (ii) select role-playing games as the digital game types for their learning. From the perspective of the teachers, they are suggested to (i) employ a game-based teaching approach for vocabulary learning; (ii) investigate the level of game acceptance among their students before conducting digital game-based vocabulary learning and also to provide sufficient support to guide students how to play the digital games; (iii) to keep a high-level of learning motivation by employing various strategies, including competition, diversifying game contents and so on. From the perspective of educational game designers, they are suggested to (i) take the features proposed by Ge and Ifenthaler (2017) into consideration during game design and (ii) examine the effectiveness of vocabulary learning tasks in digital games by using the revised version of TFA; and (iii) increase the degree of uncertainty, involve decision making and consider interactivity in the game design.

CHAPTER 7: CONCLUSION AND LIMITATIONS

7.1 Conclusion

This thesis has examined whether the ILH theory, a typical theoretical framework for evaluating involvement loads of vocabulary learning tasks, is still valid for tasks in digital game environments or not. Specifically, I attempted to answer the following two research questions in this study:

(1) For the same vocabulary learning task with the same involvement load, do learners perform better in a digital game environment than in a conventional one?
(2) For two different vocabulary learning tasks in the digital game environment, do learners perform differently even though the two tasks are considered to have the same involvement loads according to ILH in its present form?

To answer these questions, I developed a research framework by establishing three different groups. Specifically, the research framework contained two comparisons among three group, which were (i) "comparison between tasks performed on paper and digital games" (i.e., Group A vs. Group B) and (ii) "comparison between two different tasks which induce the same ILH loads according to the original version of the ILH" (i.e., Group B vs. Group C). The first comparison was to answer the first research question, and the second comparison was to answer the second research question. Further, other details in the experimental design were explained including the criteria of selecting the participants for the three groups, the specification of three vocabulary learning tasks, the procedure concerning how to conduct the pre-test, immediate post-test, and delayed post-test and the learning processes of vocabulary acquisition.

I reported the overall experimental results of the three groups of participants who completed the three tasks, i.e., Task 1: Reading a text and answering reading comprehension questions on paper, Task 2: Reading a text and answering reading comprehension questions in a digital game, and Task 3: Reading a text and inferring meanings of the underlined words in a digital game.

The research results indicated that:

(1) The three vocabulary learning tasks were effective in terms of promoting initial learning and retention of target vocabulary knowledge;

(2) Task 3 was more effective than Task 2 and Task 1 in terms of promoting initial learning and retention of target vocabulary knowledge;
(3) Task 2 was more effective than Task 1 in terms of promoting initial learning and retention of target vocabulary knowledge.

Accordingly, I proposed three suggestions for the application of ILH and TFA in digital game-based vocabulary learning and discussed the theoretical and pedagogical implications. Firstly, I suggest that the application of ILH to digital game-based vocabulary learning should consider adding one more degree of prominence to "need" when evaluating the involvement load of a task in digital game environments. I also suggest that when learners are required to infer exact meanings of target words based on the contexts, the involvement load of search should be strong, rather than moderate. Concerning the application of TFA to digital game-based vocabulary learning, I suggest adding one criterion in the category "motivation", "Is the activity game-based?"

The pedagogical implications were discussed based on the perspectives of students, teachers, and educational game designers.

(1) Students are suggested to:
- employ digital games as one of their learning tools for vocabulary acquisition; and
- select digital role-playing games as the game types for their learning.

(2) Teachers are suggested to:
- employ a game-based teaching approach for vocabulary learning;
- investigate the level of game acceptance of their students before conducting digital game-based vocabulary learning and provide sufficient support to guide students on how to play digital games;
- keep a high-level of learning motivation by employing various strategies, including competition and diversifying game contents.

(3) Educational game designers are suggested to:
- take the features proposed by Ge and Ifenthaler (2017) into consideration during game design;
- examine the effectiveness of vocabulary learning tasks in digital games by using the revised version of TFA; and
- increase the degree of uncertainty, involve decision making, and consider interactivity in game design.

In summary, the research frameworks, research questions, findings, experimental results, and implications further the understanding of digital game-based vocabulary learning. In the next section, the limitations of the study and future research directions are discussed.

7.2 Limitations

However, due to limited resources, the current study has limitations in the following aspects. Firstly, I only developed one type of digital games (i.e., the digital role-playing game) using the RPGMaker because of a limited development budget. There are many types of digital games, for example, simulation games, motion-sensing games, adventure games, digital card games, and digital board games. As the results of this study were obtained from digital role-playing game-based vocabulary, and the effects of game types on game-based vocabulary learning were not investigated, the generalizability of the research findings to other game types for vocabulary learning tends to be limited. Therefore, future research may consider comparing different types of games and examining and discussing whether game types place significant influences on the effectiveness of digital game-based vocabulary learning.

Moreover, a range of game features may be involved in a type of game, while in this study, my foci were not specific game features separately but the game as a whole part. Thus, I did not discuss whether it was the multimedia enhanced game scenarios or the interaction with the NPCs that led to the effective learning. I considered all features of the game together and discussed their total effects on vocabulary learning, rather than separate contribution to the task effectiveness. Future studies may consider comparing different game features (i.e., interactivity, challenge levels, numbers of players, game flow, etc.) and investigating how much they lead to successful vocabulary learning, respectively.

Another limitation is the limited number of task types that I investigated in this study. Only two task types were investigated: (1) Reading a text and answering reading comprehension questions, and (2) Reading a text and inferring meanings of the underlined words. The results arising from the limited types of learning tasks cannot be generalized to other types of learning tasks. Accordingly, future studies may consider investigating the effectiveness of different types of learning tasks in digital game environments in terms of promoting initial learning and retention of target vocabulary knowledge.

Concerning the ILH, its assumptions are limited from three aspects. Its three components, need,

search, and evaluation do not necessarily contribute to vocabulary learning equally, thus it is problematic to give them the same weight. However, it is uncertain what weight should be given to the three components respectively, and this is the first limitation of the ILH (Hazrat & Read, 2021). Secondly, according to the ILH, tasks with the same involvement load ought to be similarly effective despite of its distribution of the components. Nevertheless, my research and others' (e.g., Hu & Nassaji, 2016; Laufer, 2019; Kim, 2011) indicated that tasks with the same involvement load but different distribution of the components are not equally effective. Thirdly, the range of involvement load scores available to assign to tasks is limited, so it seems necessary to add more degrees of prominence to the different components. Moreover, the predictive ability of the ILH is influenced by students' proficiency levels, the frequency of their exposure to target words, and time-on-task (Hazrat & Read, 2021). Lastly, the number of studies on ILH, which are conducted in game environments, is limited. Therefore, future research may consider focusing more on the applications of ILH in game-based vocabulary learning.

Additionally, based on the review of the literature on digital game-based vocabulary learning, it can be seen that the number of Exergames is comparatively smaller than other types of games. This indicated that the AR/VR technologies for vocabulary education was not yet mature enough, and the application of AR and VR technologies in the field of vocabulary learning was limited. Another possible reason for this is the knowledge gap between AR/VR developers and educators. The development of AR/VR games involves sophisticated game design and technologies, and most available tools for AR/VR game development demand at least some knowledge of computer science (Jiang, Zhang, Shang, & Jong, 2017); yet it is likely that the majority of language teachers or researchers feel it is difficult to make use of these tools, so it is important to develop some easy-to-use tools for language educators to develop their own games. If the threshold of AR/VR game development can be lowered, the number of this type of games is likely to increase fast, given that language learners and teachers have great interest in game-based learning in virtual contexts (Lan, 2014). Moreover, there had been a call for contextualized augmented or virtual environments where learners can immerse themselves in language learning contexts, the use of games for the realization of which could make the immersion more authentic and increase learner engagement (Lan, 2015). Many researchers and educators believed in the essential role of omni-environments (either real or virtual) in promoting effective language learning in multimodality (Ozcelik & Acarturk, 2011), so it is suggested that future research focus more on AR/VR vocabulary learning games and the development of simpler tools that can be easily mastered by language teachers and educators, which will tend to engender meaningful

advancement in the field of game-based language learning.

REFERENCES

Ahmad, S. R. (2016). Importance of English communication skills. *International Journal of Applied Research, 2*(3), 478–480.

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology, 14*(2), 105–112.

Alemi, M. (2010). Educational games as a vehicle to teaching vocabulary. *The Modern Journal of Applied Linguistics, 2*(6), 425–438.

Ali Mohsen, M. (2016). The use of computer-based simulation to aid comprehension and incidental vocabulary learning. *Journal of Educational Computing Research, 54*(6), 863–884.

All, A., Castellar, E. P. N., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education, 92*, 90–103.

Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education, 3*(3), 21–34.

Asgari, A., & Mustapha, G. (2011). The type of vocabulary learning strategies used by ESL students in University Putra Malaysia. *English Language Teaching, 4*(2), 84–90.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice, 20*(1), 50–57.

Arlov, P. (2004). Wordsmith: A guide to college writing. Upper Saddler River, New Jersey: Prentice Hall.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice*. Hove (UK): Psychology Press.

Bagnole, J. W. (1993). The magic and mystery of motivation in TEFL and language learning. *TESOL Matters*, February/March, 10.

Barbosa, J. F. R., & Madeira, C. A. G. (2019). *PlayEduc: A Conceptual Framework for the Development of Digital Educational Games*. Paper presented at the *19th IEEE International Conference on Advanced Learning Technologies (ICALT)*, Maceio, Brazil.

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

Bax, S. (2003). CALL—past, present and future. *System, 31*(1), 13–28.

Bell, R. C., & Hay, J. A. (1987). Differences and biases in English language examination formats. *British Journal of Educational Psychology, 57*(2), 212–220.

Blake, R. J. (2011). Current trends in online language learning. *Annual Rreview of Applied Linguistics, 31*(1), 19–35.

Blina, F., & Munrob, M. (2008). Why hasn't technology disrupted academics' teaching practices?

Understanding resistance to change through the lens of activity theory. *Computers & Education, 50*(2), 475–490.

Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary Acquisition from Listening to Stories and Explanations of Target Words. *The Elementary School Journal, 96*(4), 415–422.

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Brown, S., & Craik, F. (2000). Encoding and retrieval of information. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory* (pp. 93–107). Oxford: Oxford University Press.

Boers, F., Warren, P., Grimshaw, G., and Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning, 30*(7), 709–725.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94,* 178–192.

Bourgonjon, J., De Grove, F., De Smet, C., Van Looy, J., Soetaert, R., & Valcke, M. (2013). Acceptance of game-based learning by secondary school teachers. *Computers & Education, 67,* 21–35.

Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies?. *Computers & Education, 88,* 343–353.

Buchem, I., Attwell, G., & Torres, R. (2011). *Understanding personal learning environments: Literature review and synthesis through the activity theory lens*. Paper presented at the *Proceedings of the Personal Learning Environment Conference* (Vol. 1, pp. 1–33), Southampton, UK.

Butler, Y. G. (2015). The use of computer games as foreign language learning tasks for digital natives. *System, 54,* 91–102.

Butler, Y. G. (2017). Motivational elements of digital instructional games: A study of young L2 learners' game designs. *Language Teaching Research, 21*(6), 735–750.

Calabrich, S. L. (2016). Learners' Perceptions of the Use of Mobile Technology in a Task-Based Language Teaching Experience. *International Education Studies, 9*(12), 120–136.

Calvo-Ferrer, J. R. (2017). Educational games as stand-alone learning tools and their motivational effect on L2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology, 48*(2), 264–278.

Chaharlang, N., & Farvardin, M. T. (2018). Predictive power of Involvement Load Hypothesis and Technique Feature Analysis across L2 vocabulary learning tasks. *International Journal of Foreign Language Teaching and Research, 6*(24), 127–141.

Chang, C. C., Liang, C., Chou, P. N., & Lin, G. Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? Perspective from multimedia and media richness. *Computers in Human Behavior*, *71*, 218–227.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and instruction, *8*(4), 293–332.

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. UK: Cambridge University Press.

Chen, C. M., Liu, H., & Huang, H. B. (2019). Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance: A case study of Taiwanese EFL learners. *ReCALL, 31*(2), 170–188.

Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game‐based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology, 49*(1), 69–77.

Chen, X., Zou, D., Xie, H., & Su, F. (2021). Twenty-five years of computer-assisted language learning: a topic modeling analysis. *Language Learning & Technology, 25*(3), 151-185.

Chen, Z. H., & Lee, S. Y. (2018). Application-driven Educational game to assist young children in learning English vocabulary. *Journal of Educational Technology & Society, 21*(1), 70–81.

Cheng, Y. M., Lou, S. J., Kuo, S. H., & Shih, R. C. (2013). Investigating elementary school students' technology acceptance by applying digital game-based learning to environmental education. *Australasian Journal of Educational Technology, 29*(1), 1–15.

Chiu, Y. H., Kao, C. W., & Reynolds, B. L. (2012). The relative effectiveness of digital game-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology, 43*(4), 104–107.

Chou, J. C., Hung, C., & Hung, Y. (2014). *Design factors of mobile game for increasing gamer's flow experience*. Paper presented at the IEEE International Conference on Management of Innovation and Technology (ICMIT), Singapore.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294.

Christ, T., Wang, X. C., & Chiu, M. M. (2017). Exploring factors related to young children's word-meaning derivations during read-alouds. *Reading Psychology, 38*(1), 1–38.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149–210.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661–686.

Cornillie, F., Thorne, S. L., &Desmet, P. (2012). ReCALL special issue: Digital games for language learning: challenges and opportunities: Editorial Digital games for language learning: From hype to insight? *ReCALL*, *24*(3), 243–256.

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.

Crawford, C. (2003). *Chris Crawford on game design*. Indiana: New Riders.

Criswell, C. (2009). Can Video Games Be Educational? *Teaching Music, 16*(6), 24.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Currie, M., & Chiramanee, T. (2010). The effect of the multiple-choice item format on the measurement of knowledge of language structure. *Language Testing*, *27*(4), 471–491.

Crystal, D. (2008). Two thousand million?. *English Today*, *24*(1), 3–6.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18.

Daniels, H., Edwards, A., Engeström, Y., Gallagher, T., & Ludvigsen, S. R. (2009). *Activity theory in practice: Promoting learning across boundaries and agencies*. Boca Raton: Taylor & Francis, Routledge.

Day, R. (1991). Incidental EFL Vocabulary Learning and Reading. *Reading in a Foreign Language*, *7*(2), 541–551.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and linguistic computing*, *25*(4), 447-464.

DeHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94.

De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24*(1), 81–12.

Deterding, S., Dixon, D., Khaled, R., &Nacke, L. (2011). From game design elements to gamefulness: defining gamification. Paper presented at the 15th international academic MindTrek conference: Envisioning future media environments, ACM.Tampere, Finland.

Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, *42*(3), 456–469.

Dörnyei, Z. (1994). Understanding L2 motivation: On with the challenge!. *The Modern Language Journal*, *78*(4), 515–523.

Dondi, C., & Moretti, M. (2007). A methodological proposal for learning games selection and quality assessment. *British Journal of Educational Technology*, *38*(3), 502–512.

Dore, R. A., Shirilla, M., Hopkins, E., Collins, M., Scott, M., Schatz, J., J., Lawson-Adams, J., Valladares, T., Foster, L., Puttre, H. & Hirsh-Pasek, K. (2019). Education in the app store: Using a mobile game to support US preschoolers' vocabulary learning. *Journal of Children and Media*, *13*(4), 452–471.

Ebrahimzadeh, M. (2017). Readers, Players, and Watchers: EFL students' vocabulary acquisition through digital video games. *English Language Teaching, 10*(2), 1-18.

EdUHK (2019). *Postgraduate Programmes - 2020 Entry: English Language Requirements*. Retrieved from https://www.eduhk.hk/gradsch/pg_prospective/application-admission.html (Accessed: 7 Oct 2019).

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge: Cambridge University Press.

Ellis, R. (2001). Introduction: Investigating form‐focused instruction. *Language Learning, 51,* 1-46.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164–194.

Elmaadaway, M. A. N. (2018). The effects of a flipped classroom approach on class engagement and skill performance in a blackboard course. *British Journal of Educational Technology, 49*(3), 479-491.

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.

Ethnologue (2019a). *Ethnologue: Languages of the World, 22nd edition website.* Retrieved from https://www.ethnologue.com/ (Accessed: 7 Oct 2019).

Ethnologue (2019b). *Ethnologue : Languages of the World - English, 22nd edition* website. Retrieved from https://www.ethnologue.com/language/eng (Accessed: 7 Oct 2019).

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167.

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*, 106–114.

Erten, İ. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*(3), 407–422.

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

Fan, M. Y (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *Modern Language Journal*, *87*(2), 222–241.

Fehr, E., Fischbacher, U., Von Rosenbladt, B., Schupp, J., & Wagner, G. G. (2003). A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative survey. *Journal of Applied Social Science Studies, 122* (4), 519-542

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

Franciosi, S. J. (2017). The effect of computer game-based learning on FL vocabulary transferability. *Journal of Educational Technology & Society*, *20*(1), 123–133.

Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101–112.

GameDesignining (2019). *The Top 10 Free Game Dev Tools*. Retrieved from https://www.gamedesigning.org/gaming/game-development-tools/ (Accessed: 30 April 2019).

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum*, *8*, 9–20.

Gardner, R. C., & MacIntyre, P. D. (1991). An instrumental motivation in language study: Who says it isn't effective?. *Studies in second language acquisition*, *13*(1), 57-72.

Gee, J. P. (2009). Literacy, video games, and popular culture. In D. R. Olson, & N. Torrance (Eds), *The Cambridge handbook of literacy* (pp. 313–325). Cambridge, England: Cambridge University Press.

Ge, X., & Ifenthaler, D. (2017). Designing engaging educational games and assessing engagement in game-based learning. In R. Zheng, & M. K. Gardner (Eds.). *Handbook of Research on Serious Games for Educational Applications* (pp. 253–270). Hershey, PA: IGI Global.

Gillespie, J. (2020). CALL research: Where are we now? *ReCALL*, *32*(2), 127–144.

Gohar, M. J., Rahmanian, M., & Soleimani, H. (2018). Technique feature analysis or involvement load hypothesis: Estimating their predictive power in vocabulary learning. *Journal of Psycholinguistic Research*, *47*(4), 859–869.

Godwin-Jones, R. (2016). Emerging technologies augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Language Learning & Technology*, *20*(3), 9–19.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105.

Gu, Y. (2010). Learning strategies for vocabulary development. *Reflections on English Language Teaching*, *9*(2), 105–118

Gu, Y, & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, *46*(4), 643–679.

Gu, P. Y. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *TESL-EJ*, *7*(2), 1–25.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179.

Haratmeh, M. S. (2012). Involvement load and task type in task effectiveness: Two aspects of vocabulary knowledge. *International Journal of Academic Research*, *4*(4).

Hansson, T. (2012). ICT, learning objects and activity theory. In A. Lopez-Varela (Ed.), *Social sciences and cultural studies – Issues of language, public opinion, education and welfare* (pp. 411–438). New York: Intech.

Hao, Y., Lee, K. S., Chen, S. T., & Sim, S. C. (2019). An evaluative study of a mobile application for middle school students struggling with English vocabulary learning. *Computers in Human Behavior*, *95*, 208–216.

Hazrat, M., & Read, J. (2021). Enhancing the involvement load hypothesis as a tool for classroom vocabulary research. *TESOL Quarterly*, doi: 10.1002/tesq.3051.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*(2), 689–696.

HKEAA (2013). *Benchmarking Study between IELTS and HKDSE English*. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/IR/FAQs_IELTS_eng_Aug2013.pdf (Accessed: 6 May 2018).

Hodgson, P., Man, D., & Leung, J. (2010). *Managing the development of digital educational games*. Paper presented at the 3rd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, Kaohsiung, Taiwan.

Hong, J. C., Tai, K. H., & Ye, J. H. (2019). Playing a Chinese remote-associated game: The correlation among flow, self-efficacy, collective self-esteem and competitive anxiety. *British Journal of Educational Technology*, *50*(5), 2720–2735.

Hogg, R. V., & Tanis, E. A. (2009). *Probability and statistical inference*. Pearson Educational International.

Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? *Computers & Education*, *106*, 137–149.

Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System*, *56*, 28–39.

Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A Meta-Analytic Study. *The Modern Language Journal*, *96*(4), 544–557.

Huang, Y. M., & Huang, Y. M. (2015). A scaffolding strategy to develop handheld sensor-based vocabulary games for improving students' learning motivation and performance. *Educational Technology Research and Development*, *63*(5), 691–708.

Huang, Y. M., Huang, Y. M., Huang, S. H., & Lin, Y. T. (2012). A ubiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. *Computers & Education*, *58*(1), 273–282.

Huang, Y. M., Liang, T. H., Su, Y. N., & Chen, N. S. (2012). Empowering personalized learning with an interactive e-book learning system for elementary school students. *Educational Technology Research and Development*, *60*(4), 703–722.

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). Palgrave Macmillan, London.

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.

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

Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., & Wang, C. C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, *126*, 89–104.

Hung, H. C., Young, S. S. C., & Lin, C. P. (2015). No student left behind: A collaborative and competitive game-based learning environment to reduce the achievement gap of EFL students in Taiwan. *Technology, Pedagogy and Education*, *24*(1), 35–49.

Hung, C. M., Huang, I., & Hwang, G. J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, *1*(2), 151–166.

Hwang, G. J., & Wang, S. Y. (2016). Single loop or double loop learning: English vocabulary learning performance and behavior of students in situated computer games with different guiding strategies. *Computers & Education*, *102*, 188–201.

Itch (2019). *Search results of games created by RPGMaker*. Retrieved from https://itch.io/games/made-with-rpg-maker (Accessed, 20 March 2019)

Jahangard, A., & Movassagh, H. (2011). High vs. low load vocabulary learning tasks: A Case for intentional learning. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *2*(4), 48–55.

Jian, M. S., Shen, J. H., Huang, T. C., Chen, Y. C., & Chen, J. L. (2015). Language learning in cloud: Modular role player game-distance-learning system based on voice recognition. In J. Park, Y. Pan, C. Kim, & Y. Yang (Eds), *Future Information Technology-II* (pp. 129-135). Springer, Dordrecht.

Jiang, Y., Zhang, L., Shang, J. J., & Jong, M. S. Y. (2017). Game-based learning: Design and application. In H. Niemi, & J. Jia (Eds.), *New ways to teach and learn in China and Finland* (pp. 195–211). New York: Peter Lang.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?. *Applied linguistics*, *19*(3), 357–377.

Jonassen, D. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65–94.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly*, *40*(1), 157-181.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity.* Oxford: Oxford University Press.

Jung, J., & Graf, S. (2008). An approach for personalized web-based vocabulary learning through word association games. In *2008 International Symposium on Applications and the Internet* (pp. 325-328). IEEE.

Kachru, B. (2006). English: World Englishes. In B. Keith (Ed.), *Encyclopedia of. language & linguistics* (pp. 195–202). Elsevier.

Karakostas, A., & Demetriadis, S. (2011). Enhancing collaborative learning through dynamic forms of support: The impact of an adaptive domain-specific support strategy. *Journal of Computer Assisted Learning, 27*(3), 243–258.

Karpicke, J. D. (2017). *Retrieval-based learning: A decade of progress*. United states: Purdue University. Retrieved from https://files.eric.ed.gov/fulltext/ED599273.pdf

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968.

Ke, F., &Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology, 44*(2), 225–242.

Keating, G. D. (2008). Task effectiveness and vocabulary learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*(3), 365–386.

Khenissi, M. A., Essalmi, F., Jemni, M., Graf, S., & Chen, N. S. (2016). Relationship between learning styles and genres of games. *Computers & Education, 101*, 1–14.

Kiili, K. (2005). *On educational game design: Building blocks of flow experience* (Doctoral thesis, Tampere University of Technology).

Kim, Y. (2008). The role of task–induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285–325.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61*(1), 100–140.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. In P. Robinson (Ed.), *Task-Based Language Learning* (pp. 100–140). Malden, MA: Wiley-Blackwell.

Kinginger, C., & Wu, Q. (2018). Learning Chinese through contextualized language practices in study abroad residence halls: Two case studies. *Annual Review of Applied Linguistics, 38*, 102–121.

Konopak, B., Sheard, C., Longman, D., Lyman, B., Slaton, E., Atkinson, R., & Thames, D. (1987). Incidental versus intentional vocabulary learning from context. *Reading Psychology: An International Quarterly, 8*(1), 7–21.

Klimesch, W. (1994). *The Structure of long-term memory: A connectivity model of semantic processing*. Hillsdale, NJ: Lawrence Erlbaum.

Krashen, S. (1982). Principles and practice in second language acquisition.

Kucirkova, N. (2017). An integrative framework for studying, designing and conceptualising interactivity in children's digital books. *British Educational Research Journal, 43*(6), 1168–1185.

Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology, 49*(2), 207–218.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65–85.

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.

Laufer, B. (1990). Sequence and order in the development of L2 lexis: Some evidence from lexical confusions. *Applied Linguistics, 11*(3), 281–296.

Laufer, B., & Girsai, N. (2008). Form–focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694–716.

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching, 42*(3), 341–354.

Laufer, B. (2019). Evaluating exercises for learning vocabulary. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 351–368). Taylor & Francis Ltd.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26.

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language, *Prospect, 16*(3), 44–54.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59(*4), 567–587.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19*(6), 687–711.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391–411.

Lan, Y. J. (2014). Does Second Life improve Mandarin learning by overseas Chinese students. *Language Learning & Technology, 18*(2), 36–56.

Lan, Y. J. (2015). Contextual EFL learning in a 3D virtual environment. *Language Learning & Technology, 19*(2), 16–31.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3*(3), 299–321.

Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for vocabulary

knowledge. *Scientific Studies of Reading*, *15*(1), 92–108.

Lantolf, J. P. and Thorne, S. L. (2008) Sociocultural theory and second language learning. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 201–224). New York, NY: Routledge.

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2019). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning*, DOI: 10.1080/09588221.2019.1619585, 1–22.

Liao, C. W., Chen, C. H., & Shih, S. J. (2019). The interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behavior patterns in a digital game-based learning environment. *Computers & Education*, *133*, 43–55.

Lightbown, P. M., & Spada, N. (2006). *How languages are learned* (3rd ed.). Oxford: Oxford University Press.

Lin, C. J., Hwang, G. J., Fu, Q. K., & Chen, J. F. (2018). A flipped contextual game-based learning approach to enhancing EFL students' English business writing performance and reflective behaviors. *Journal of Educational Technology & Society, 21*(3), 117–131.

Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, *16*(1), 33–42.

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/Revue canadienne de psychologie, 44(*1), 87–112.

Long, M. H. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences*, *379*(1), 259–278.

Luppescu, S., & Day, R. (1993). Reading, Dictionaries, and Vocabulary Learning. *Language Learning, 43*(2), 263–279.

Ludwig, J. (1984). Vocabulary acquisition as a function of word characteristics. *The Canadian Modern Language Review*, *40*(4), 552–562.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405–430.

Mármol, G. A., & Sánchez-Lafuente, Á. A. (2013). The involvement load hipothesis: The effect on vocabulary learning in primary educaion. *Revista española de lingüística aplicada*, (26), 11–24.

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3-4), 422–429.

Mavridis, A., & Tsiatsos, T. (2017). Game-based assessment: investigating the impact on test anxiety and exam performance. *Journal of Computer Assisted Learning*, *33*(2), 137–150.

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, *8*(4), 357–371.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19.

Mayer, R. E. (1999). Research-based principles for the design of instructional messages: The case of multimedia explanations. *Document Design, 1*(1), 7–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction, 13*(2),125–139.

Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). IAP Information Age Publishing.

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction, 29*, 171–173.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.

McGraw, I., Yoshimoto, B., & Seneff, S. (2009). Speech-enabled card games for incidental vocabulary acquisition in a foreign language. *Speech Communication, 51*(10), 1006–1023.

McGregor, K. K., Marshall, B. A., Julian, S. K., & Oleson, J. (2019). Learning while playing: A randomized trial of serious games as a tool for word mastery. *Language, Speech, and Hearing Services in Schools, 50*(4), 596–608.

MeliorGames (2019). *2D Vs 3D Games: Differences, Benefits and Costs*. Retrieved from https://meliorgames.com/game-development/2d-vs-3d-games-differences-benefits-and-costs/ (Accessed, 20 March 2019)

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education, 3*(4), 311–328.

Milton, J. (2008). Vocabulary uptake from informal learning tasks. *Language Learning Journal*, *36*(2), 227–237.

Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning, 29*(7), 1220–1237.

Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior*, *24*(6), 2530–2540.

Müller, A., Son, J. B., Nozawa, K., & Dashtestani, R. (2018). Learning English idioms with a web-based educational game. *Journal of Educational Computing Research*, *56*(6), 848–865.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237–270.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, *20*(1), 3–20.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273.

Nardi, B. (2003). A brief introduction to activity theory. In K. Gaßner, & O. Schröder (Eds.), Wissensmodellierung und Wissenskommunikation in Lernszenarien (special issue). *Künstliche Intelligenz*, *17*(1), 35–36.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Massachusetts: Newbury House.

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

Nation, P. (2003). The role of the first language in foreign language learning.

Nation, P. (2006). Vocabulary: Second language (2nd ed.). In B. Keith (Ed.), *Encyclopedia of Language & Linguistics* (pp. 448–454), Oxford: Elsevier.

Nation, I. S. (2013). *Teaching & learning vocabulary*. Boston: Heinle Cengage Learning.

Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading, 1*, 5-16.

Nation, P., & Newton, J. (1997). Second language vocabulary acquisition: A rationale for pedagogy. In C. Jame & H. Thomas (Eds.), *Teaching vocabulary* (pp. 238–254). UK: Cambridge University Press.

Newton, J. M., & Nation, I. S. P. (2020). *Teaching ESL/EFL listening and speaking*. Routledge.

Nation, P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

Neville, D. O., Shelton, B. E., & McInnis, B. (2009). Cybertext redux: Using digital game-based learning to teach L2 vocabulary, reading, and culture. *Computer Assisted Language Learning*, *22*(5), 409–424.

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, *11*(2), 159–176.

Nuyen, N. T. T., & Nga, K. T. T. (2003). Learning vocabulary through games. *Asian EFL Journal, 5*(4), 90–105.

Ozcelik, E., &Acarturk, C. (2011). Reducing the spatial distance between printed and online information sources by means of mobile technology enhances learning: Using 2D barcodes. *Computers & Education, 57*(3), 2077–2085.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1–12.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. 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: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21*(2), 195–224.

Paivio, A. (1990). *Mental representations: A dual coding approach* (Vol. 9). Oxford University Press.

Pan, W. F. (2017). The effects of using the kinect motion-sensing interactive system to enhance English learning for elementary students. *Educational Technology & Society, 20*(2), 188–200.

Peña-Ayala, A., Sossa, H., & Méndez, I. (2014). Activity theory as a framework for building adaptive e-learning systems: A case to provide empirical evidence. *Computers in Human Behavior*, 30, 131–145.

Peterson, M. (2010) Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation & Gaming, 41*(1), 72–93.

Peterson, M. (2013). *Computer games and language learning*. New York: Palgrave Macmillan.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L.T. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 67–86). John Benjamins Publishing.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning, 53*(2), 233–284.

Ranalli, J. (2008). Learning English with The Sims: exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, *21*(5), 441–455.

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

Reiners, T., & Wood, L. C. (2015). *Gamification in Education and Business*. Switzerland: Springer Science.

Reynolds, B. L. (2017). Evidence for the task-induced involvement construct in incidental vocabulary acquisition through digital gaming. *The Language Learning Journal*, *45*(4), 466–484.

Roediger, H. L. (2000). Why retrieval is the key process in understanding human memory. In *Memory, consciousness, and the brain*: *The Tallinn Conference* (pp. 52-75). Philadelphia, ePA, PA: Psychology Press.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25–41.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*(4), 183–186.

Rupley, W., & Nichols, W. (2005). Vocabulary instruction for the struggling reader. *Reading & Writing Quarterly*, *21*(3), 239–260.

Ryu, D. (2013). Play to learn, learn to play: Language learning through gaming culture. *ReCALL*, *25*(2), 286–301.

Saeman, J. F. (1970). Wood Research and the Environment. *Journal of Forestry, 68*(7), 396–398.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal*, *79*(1), 15–28.

Sandberg, J., Maris, M., & Hoogendoorn, P. (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. C*omputers & Education*, *76*, 119–130.

Salen, K., & Zimmerman, E. (2004). Rules of Play: Game Design Fundamentals. Cambridge, Massachusetts: The MIT Press.

Schell, J. (2008). The art of game design : a book of lenses. Amsterdam ;Boston: Elsevier/Morgan Kaufmann.

Schellekens, J. M., Sijtsma, G. J., Vegter, E., & Meijman, T. F. (2000). Immediate and delayed after-effects of long lasting mentally demanding work. *Biological Psychology*, *53*(1), 37–56.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention

and awareness in learning. *Attention and Awareness in Foreign Language Learning, 9,* 1–63.

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

Schmitt, N., and Schmitt, D. (1993). Identifying and assessing vocabulary learning strategies. *Thai TESOL Bulletin*. *5*(4), 27–33.

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

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

Schnotz, W., Böckheler, J., & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, *14*(2), 245–265.

Schnotz, W., Picard, E., & Henninger, M. (1994). The use of graphics and texts in constructing mental models. *Advances in Psychology*, *108,* 185–205.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal*, *59(*4), 339–341.

Shadiev, R., Hwang, W. Y., & Liu, T. Y. (2018). A study of the use of wearable devices for healthy and enjoyable English as a foreign language learning in authentic contexts. *Journal of Educational Technology & Society*, *21*(4), 217–231.

Sharifian, F. (2013). Globalisation and developing metacultural competence in learning English as an International Language. *Multilingual Education*, *3*(1), 1–11.

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior*, *22*(1), 71–97.

Shin, J. A. (2010). The Effects of Prior Knowledge of the Target Form on Noticing during Output Production. *English Teaching*, *65*(1), 241-265.

Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*(2), 253–279. SIL International (2019). *SIL International official website*. Retrieved from https://www.sil.org/ (Accessed: 7 Oct 2019).

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent vocabulary knowledge: A cross-linguistic investigation. *Developmental Science*, *17*(1), 94–109.

Smeets, D. J. H., & Bus, A. G. (2015). The interactive animated e-book as a vocabulary learning device for kindergartners. *Applied Psycholinguistics*, *36*(4), 899–920.

Someren, M. W., Barnard Y. F., & Sandberg J. (1994). *The think aloud method: a practical approach to modelling cognitive*. London: Academic Press.

Song, M., & Zhang, S. (2008). EFM: A model for educational game design. In Z. Pan, X. Zhang, A. El Rhalibi, W. Woo, & Y. Li (Eds.), *Technologies for e-learning and digital entertainment*.

Berlin: Springer.

Soderstrom, N. C., & McCabe, D. P. (2011). Are survival processing memory advantages based on ancestral priorities?. *Psychonomic Bulletin & Review*, *18*(3), 564–569.

Sun, Y., & Dong, Q. (2004). An experiment on supporting children's English vocabulary learning in multimedia context. *Computer Assisted Language Learning, 17*(2), 131–147.

Sung, H. Y., & Hwang, G. J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51.

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System*, *51*, 65–76.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. *Sociocultural Theory and Second Language Learning, 97*(1), 97–114.

Swanborn, M., & De Glopper, K. (1999). Incidental vocabulary learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261–285.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment (CIE), 3*(3), 1–24.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

Sweller, J. (2002). Visualisation and instructional design. In R. Ploetzner (Ed.), *International Workshop on Dynamic Visualizations and Learning.* Tubingen, Germany: Knowledge Media Research Center.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302–321.

Toyota, H., & Kikuchi, Y. (2005). Encoding richness of self-generated elaboration and spacing effects on incidental memory. *Perceptual and Motor Skills*, *101*(2), 621–627.

Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education, 125*, 345–357.

Tsai, F. H., Tsai, C. C., & Lin, K. Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education, 81*, 259–269.

Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1–25.

Ulanoff, S. H., & Pucci, S. L. (1999). Learning words from books: The effects of read-aloud on second

language vocabulary acquisition. *Bilingual Research Journal*, *23*(4), 409–422.

University Grant Council (2014). *Common English Proficiency Assessment Scheme*. Retrieved from https://www.ugc.edu.hk/eng/ugc/about/press_speech_other/press/2014/pr26092014.html (Accessed: 16 March 2019).

U.S. Department of Education, Office of Educational Technology (2017). *Reimagining the role of technology in education: 2017 national education technology plan update*. Retrieved from https://tech.ed.gov/files/2017/01/NETP17.pdf (Accessed: 6 May 2018).

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, *29*(6), 2568–2572.

Verga, L., & Kotz, S. A. (2017). Help me if I can't: Social interaction effects in adult contextual vocabulary learning. *Cognition*, 168, 76–90.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management, 16*(4), 361-385.

Wang, L. C., & Chen, M. P. (2010). The effects of game strategy and preference-matching on flow experience and programming performance in game‐based learning. *Innovations in Education and Teaching International*, *47*(1), 39-52.

Warring, R. (1997). The negative effects of learning words in semantic sets. *System*, *25*(2), 261–274.

Wei, C. W., Kao, H. Y., Lu, H. H., & Liu, Y. C. (2018). The effects of competitive gaming scenarios and personalized assistance strategies on English vocabulary learning. *Journal of Educational Technology & Society*, *21*(3), 146–158.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on vocabulary knowledge. *Studies in second language acquisition*, *27*(1), 33–2.

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*(1), 63–81.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13–40.

Wesiak, G., Steiner, C. M., Moore, A., Dagger, D., Power, G., Berthold, M., Albert, D. and Conlan, O. (2014). Iterative augmentation of a medical training simulator: Effects of affective metacognitive scaffolding. *Computers & Education*, *76*, 13–29.

Werbach, K. (2014). (Re)defining gamification: A process approach. In A. Spagnolli, L. Chittaro, & L. Gamberini (Eds), *Persuaive Technology* (pp. 266–272). Springer, Cham.

Wittrock, M. C. (1991). Generative teaching of comprehension. *The Elementary School Journal*, *92*(2), 169-184.

Whitehead, J. (2008). *Introduction to game design in the large classroom*. Paper presented at the 3rd International Conference on Game Development in Computer Science Education (pp. 61–65). New York: United States.

Whittemore, I. C. (1924). The influence of competition on performance: An experimental study. *Journal of Abnormal and Social Psychology*, *19*(3), 236–253.

Wieland, K. (2008). *Toward a unified theory of contextual vocabulary acquisition* (Unpublished doctoral dissertation, State University of New York at Buffalo). Retrieved from http://search.proquest.com/docview/193647087?accountid=14169. (193647087)

Wilkins, D. (1972). *Linguistics in language teaching*. London: ArnoldWolf, M. J. (2001). Genre and the video game. *The Medium of the Video Game*, *1*,113-134.Wu, T. T., & Huang, Y. M. (2017). A mobile game-based English vocabulary practice system based on portfolio analysis. *Educational Technology & Society*, *20*(2), 265–277.

Xie, H., Zou, D., Lau, R. Y., Wang, F. L., & Wong, T. L. (2016). Generating incidental word-learning tasks via topic-based and load-based profiles. *IEEE Multimedia*, *23*(1), 60–70.

Yanagisawa, A. (2016). The Effects of Receptive and Productive Word Retrieval Practice on Second Language Vocabulary Learning. *KATE Journal*, *30*, 139–152.

Yang, Y. T. C., & Chang, C. H. (2013). Empowering students through digital game authorship: Enhancing concentration, critical thinking, and academic achievement. *Computers & Education*, *68*, 334–344.

Yang, Q. F., Chang, S. C., Hwang, G. J., & Zou, D. (2020). Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. *Computers & Education*, *148*, 103808.

Yaqubi, B., Rayati, R. A., & Allemzade 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*, *29*(1), 145–163.

Young, S. S. C., & Wang, Y. H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Educational Technology & Society*, *17*(3), 239–251.

Yip, F. W., & Kwan, A. C. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International*, *43*(3), 233–-249.

Yu, Z. (2018). Differences in serious game-aided and traditional English vocabulary acquisition.

Computers & Education, *127*, 214–232.

Zou, D. (2012). *A study of the components of the involvement load hypothesis: how involvement load should be allocated to" Search" and" Evaluation"* (Doctoral dissertation, City University of Hong Kong). Retrieved from http://lbms03.cityu.edu.hk/theses/abt/phd-en-b46907701a.pdf

Zou, D. (2016). Comparing dictionary-induced vocabulary learning and inferencing in the context of reading. *Lexikos*, *26*(1), 372–390.

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence writing and composition writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, *27*(1), 54–57.

Zou, D., Huang, Y., & Xie, H. (2019). Digital game-based vocabulary learning: where are we and where are we going?. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2019.1640745

Zou, D., & Xie, H. (2018). Personalized word-learning based on technique feature analysis and learning analytics. *Journal of Educational Technology & Society*, *21*(2), 233–244.

Zou, D., Xie, H., Li, Q., Wang, F. L., & Chen, W. (2014). The load-based learner profile for incidental vocabulary learning task generation. In E. Popescu, R.W.H. Lau, K. Pata, H. Leung, & M. Laanpere (Eds), Advances in Web-Based Learning-International Conference on Web-Based Learning (pp. 190–200). Springer, Cham.

Zou, D., Xie, H., Rao, Y., Wong, T. L., Wang, F. L., & Wu, Q. (2017). A comparative study on various vocabulary knowledge scales for predicting vocabulary pre-knowledge. *International Journal of Distance Education Technologies (IJDET)*, *15*(1), 69–81.

Zou, D., Wang, M., Xie, H., Cheng, G., Wang, F. L., & Lee, L. K. (2020). A comparative study on linguistic theories for modeling EFL learners: facilitating personalized vocabulary learning via task recommendations. *Interactive Learning Environments*, 1–13. https://doi.org/10.1080/10494820.2020.1789178

APPENDICES

Appendix 1. Task 1 - Reading a text and answering reading comprehension questions on paper (adapted from Zou, 2012)

Instructions:

Please read the text and answer the ten comprehension questions.

Dealing with Procrastination

Procrastination refers to the act of delaying the work you should do to a later time. It is wasting time when you have some work to do but choose not to do it early.

Most procrastinators do not feel that they are doing this on purpose. Instead, they feel that they really tried to do the work. But they could not start because there are too many things out of their control. So, they have a long list of **ostensible** reasons, such as "I did not have time", "I had to attend a wedding" or "I had other important things to do". However, these surface reasons are not true.

When procrastination becomes a habit, it is **pernicious**. If you procrastinate, you may often find yourself not having enough time to do a satisfying work. This can make other people unhappy and get a bad impression of you. Habitual procrastination can even damage your friendships. As you always **renege** on your promises to complete work on time, your friends may no longer trust you. Thus, it is very necessary to know the causes of procrastination and learn to deal with it.

Procrastination is often caused by a real or imagined fear or worry. For instance, you might delay preparing for an oral presentation, because you are **apprehensive** that you will not be able to remember the entire speech. You may be so worried about doing a bad job that you decide not to work on it until the last minute.

Being a perfectionist is a main **trait** that causes fear and anxiety. When you imagine yourself making an English presentation, are you comparing yourself to great speakers? Or are you picturing that others **taunt** you about your accent? Rather than worry yourself with these thoughts, think of specific ways to improve the performance, may help to lessen performance anxiety.

"Lack of motivation" may also cause procrastination. If you are forced to learn a subject you are not interested in, you may find yourself wasting time instead of being **assiduous**. Even if you know that the subject can help you get a good job, you will not work hard. It is not easy to think carefully about something you have no interest in.

Feelings of **lassitude** can cause procrastination, too. This often happens when you keep on pushing yourself very hard without getting any rest. If so, you may experience a state of tiredness and feel

unable to focus on any work. Learning to balance your time can be helpful in preventing this.

Sometimes you put off doing something because you do not know how to do it. For example, if you start doing a job that requires collecting data and creating graphics, having the right skills is **paramount**. Knowing how to do a task before you begin it is very important. Sometimes it is difficult to ask for help and sometimes it is even harder to realize that you need help. Being able to **divulge** personal limitations and ask for help is a skill we need to learn.

1. What is the main idea of the text?

A. Procrastinators do not work hard.

B. Procrastinators need to get motivated.

C. Perfectionists often procrastinate.

D. Procrastination can be understood and controlled.

2. What can be causes of procrastination? Choose the incorrect answer.

A. Lack of motivation

B. Lack of necessary skills.

C. Not knowing the importance of completing work on time.

D. Feelings of tiredness

3. What do procrastinators feel?

A. They feel that they are procrastinating on purpose.

B. They feel that they are good enough.

C. They feel that they have not tried their best to do the work.

D. They feel that there are too many things out of their control.

4. What may happen when procrastination becomes a habit?

A. You may often find yourself having enough time to do a satisfying work.

B. You may break your promises to complete work on time.

C. Your friends are happy with your procrastination.

5. What will not cause procrastination?

A. Your real fear or worry.

B. Your imagined fear or worry.

C. Your determination of completing the work on time.

6. Why is it necessary to know the causes of procrastination? Choose the incorrect answer.

A. This can help you improve your English.

B. The habit of procrastination is harmful.

C. Knowledge of the causes of procrastination can help you learn to deal with it.

7. What may happen when you keep on pushing yourself very had without getting any rest? Choose the incorrect answer.

A. You will feel very tired.

B. You will feel unable to focus on any work.

C. Your feelings of tiredness will cause procrastination.

D. You will work more efficiently.

8. Which of the following cannot help you deal with procrastination?

A. Know how to do a task before you begin.

B. Know when you need help.

C. Know how to ask for help.

D. Know the importance of being perfect.

9. Which of the following is true?

A. Some of our fear is not real. We imagine it.

B. All our fear and anxiety is real.

10. What are suggested as ways to deal with procrastination?

A. Set strict rules and force yourself to work hard.

B. Do not compare yourself to great speakers.

C. Have great dreams to encourage yourself.

D. Ask for help when it is necessary.

Appendix 2. Task 2 - Reading a text and answering reading comprehension questions in a digital game (adapted from Zou, 2012)

Instructions:

Please read the text in the task card and then answer the ten comprehension questions asked by the monsters (i.e., the NPCs of the game).

Dealing with Procrastination

Procrastination refers to the act of delaying the work you should do to a later time. It is wasting time when you have some work to do, but choose not to do it early.

Most procrastinators do not feel that they are doing this on purpose. Instead, they feel that they really tried to do the work. But they could not start because there are too many things out of their control. So they have a long list of **ostensible** reasons, such as "I did not have time", "I had to attend a wedding" or "I had other important things to do". However, these surface reasons are not true.

When procrastination becomes a habit, it is **pernicious**. If you procrastinate, you may often find yourself not having enough time to do a satisfying work. This can make other people unhappy and get a bad impression of you. Habitual procrastination can even damage your friendships. As you always **renege** on your promises to complete work on time, your friends may no longer trust you. Thus it is very necessary to know the causes of procrastination and learn to deal with it.

Procrastination is often caused by a real or imagined fear or worry. For instance, you might delay preparing for an oral presentation, because you are **apprehensive** that you will not be able to remember the entire speech. You may be so worried about doing a bad job that you decide not to work on it until the last minute.

Being a perfectionist is a main **trait** that causes fear and anxiety. When you imagine yourself making an English presentation, are you comparing yourself to great speakers? Or are you picturing that others **taunt** you about your accent? Rather than worry yourself with these thoughts, think of specific ways to improve the performance, may help to lessen performance anxiety.

"Lack of motivation" may also cause procrastination. If you are forced to learn a subject you are not interested in, you may find yourself wasting time instead of being **assiduous**. Even if you know that the subject can help you get a good job, you will not work hard. It is not easy to think carefully about something you have no interest in.

Feelings of **lassitude** can cause procrastination, too. This often happens when you keep on pushing yourself very hard without getting any rest. If so, you may experience a state of tiredness and feel

unable to focus on any work. Learning to balance your time can be helpful in preventing this.

Sometimes you put off doing something because you do not know how to do it. For example, if you start doing a job that requires collecting data and creating graphics, having the right skills is **paramount**. Knowing how to do a task before you begin it is very important. Sometimes it is difficult to ask for help and sometimes it is even harder to realize that you need help. Being able to **divulge** personal limitations and ask for help is a skill we need to learn.

1. What is the main idea of the text?

A. Procrastinators do not work hard.

B. Procrastinators need to get motivated.

C. Perfectionists often procrastinate.

D. Procrastination can be understood and controlled.

E. Procrastination

2. What can be causes of procrastination? Choose the incorrect answer.

A. Lack of motivation

B. Lack of necessary skills.

C. Not knowing the importance of completing work on time.

D. Feelings of tiredness

3. What do procrastinators feel?

A. They feel that they are procrastinating on purpose.

B. They feel that they are good enough.

C. They feel that they have not tried their best to do the work.

D. They feel that there are too many things out of their control.

4. What may happen when procrastination becomes a habit?

A. You may often find yourself having enough time to do a satisfying work.

B. You may break your promises to complete work on time.

C. Your friends are happy with your procrastination.

5. What will not cause procrastination?

A. Your real fear or worry.

B. Your imagined fear or worry.

C. Your determination of completing the work on time.

6. Why is it necessary to know the causes of procrastination? Choose the incorrect answer.

A. This can help you improve your English.

B. The habit of procrastination is harmful.

C. Knowledge of the causes of procrastination can help you learn to deal with it.

7. What may happen when you keep on pushing yourself very had without getting any rest? Choose the incorrect answer.

A. You will feel very tired.

B. You will feel unable to focus on any work.

C. Your feelings of tiredness will cause procrastination.

D. You will work more efficiently.

8. Which of the following cannot help you deal with procrastination?

A. Know how to do a task before you begin.

B. Know when you need help.

C. Know how to ask for help.

D. Know the importance of being perfect.

9. Which of the following is true?

A. Some of our fear is not real. We imagine it.

B. All our fear and anxiety is real.

10. What are suggested as ways to deal with procrastination?

A. Set strict rules and force yourself to work hard.

B. Do not compare yourself to great speakers.

C. Have great dreams to encourage yourself.

D. Ask for help when it is necessary.

Appendix 3. Task 3 - Reading a text and inferring meanings of the underlined words in a digital game (adapted from Zou, 2012)

Instructions:

Please read the text in the task card, infer the meanings of the underlined words, and answer the questions asked by the monsters (i.e., the NPCs of the game).

Dealing with Procrastination

Procrastination refers to the act of delaying the work you should do to a later time. It is wasting time when you have some work to do but choose not to do it early.

Most procrastinators do not feel that they are doing this on purpose. Instead, they feel that they really tried to do the work. But they could not start because there are too many things out of their control. So, they have a long list of **ostensible** reasons, such as "I did not have time", "I had to attend a wedding" or "I had other important things to do". However, these surface reasons are not true.

When procrastination becomes a habit, it is **pernicious**. If you procrastinate, you may often find yourself not having enough time to do a satisfying work. This can make other people unhappy and get a bad impression of you. Habitual procrastination can even damage your friendships. As you always **renege** on your promises to complete work on time, your friends may no longer trust you. Thus, it is very necessary to know the causes of procrastination and learn to deal with it.

Procrastination is often caused by a real or imagined fear or worry. For instance, you might delay preparing for an oral presentation, because you are **apprehensive** that you will not be able to remember the entire speech. You may be so worried about doing a bad job that you decide not to work on it until the last minute.

Being a perfectionist is a main **trait** that causes fear and anxiety. When you imagine yourself making an English presentation, are you comparing yourself to great speakers? Or are you picturing that others **taunt** you about your accent? Rather than worry yourself with these thoughts, think of specific ways to improve the performance, may help to lessen performance anxiety.

"Lack of motivation" may also cause procrastination. If you are forced to learn a subject you are not interested in, you may find yourself wasting time instead of being **assiduous**. Even if you know that the subject can help you get a good job, you will not work hard. It is not easy to think carefully about something you have no interest in.

Feelings of **lassitude** can cause procrastination, too. This often happens when you keep on pushing yourself very hard without getting any rest. If so, you may experience a state of tiredness and feel

unable to focus on any work. Learning to balance your time can be helpful in preventing this.

Sometimes you put off doing something because you do not know how to do it. For example, if you start doing a job that requires collecting data and creating graphics, having the right skills is **paramount**. Knowing how to do a task before you begin it is very important. Sometimes it is difficult to ask for help and sometimes it is even harder to realize that you need help. Being able to **divulge** personal limitations and ask for help is a skill we need to learn.

1. **ostensible** means ( )

A. seeming to be doubtful, but perhaps not

B. seeming to be strong, but perhaps not

C. seeming to be true, but perhaps not

2. **pernicious** means ( )

A. fully intentional

B. not changeable

C. very harmful

3. **renege on promises** means ( )

A. to make promises

B. to break promises

C. to keep promises

4. **apprehensive** means ( )

A. decided

B. worried

C. comprehensive

5. **trait** means ( )

A. character or personality

B. strategy or guideline

C. trouble or problem

6. **taunt** means ( )

A. to respect sb. by learning from him

B. to encourage sb. by praising him

C. to upset sb. by laughing at him

7. **assiduous** means ( )

A. associative and creative

B. careful and hardworking

C. confident and interested

8. **lassitude** means ( )

A. tiredness

B. sickness

C. easiness

9. **paramount** means ( )

A. more important than anything else

B. more advanced than anything else

C. more difficult than anything else

10. **divulge** means ( )

A. to give away information that should be secret

B. to develop information that should be secret

C. to keep information that should be secret

Instructions:

Please read the text in the task card, infer the meanings of the underlined words, and answer the questions asked by the monsters (i.e., the NPCs of the game).

Dealing with Procrastination

Procrastination refers to the act of delaying the work you should do to a later time. It is wasting time when you have some work to do but choose not to do it early.

Appendix 4. The Modified Vocabulary Knowledge Scale Used in the Pre-test, Immediate Post-Test, and Delayed Post-Test (Folse, 2006; Zou, 2012)

Instructions:

1. If you cannot remember the meaning of the word, please tick (√) in Column [A].
2. If you can remember the meaning of the word, please write it in Column [B]. You can use either English or Chinese.
3. If you can write a sentence by using this word, please write one in Column [C]. Your sentence should include at least seven words.

		[A]	[B]	[C]
		I cannot remember the meaning of the word.	I can remember the meaning of the word, it means:	I can write the sentence by using this word, for example
1	trait			
2	apprehensive			
3	taunt			
4	divulge			
5	lassitude			
6	assiduous			
7	pernicious			
8	indispensable			
9	renege			
10	ostensible			

Appendix 5. Entries of the Target Words

These entries are extracted from the Longman Dictionary of Contemporary English 5 and online corpus (www.ldoceonline.com/dictionary/).

1. trait

/treɪ, treɪt $ treɪt/ **noun [countable]**

formal a particular quality in someone's character

personality/character traits

a mental illness associated with particular personality traits

genetic/inherited traits

trait

• Does Bryce have any bad traits?

• She shows both traits assigned to the men, immorality and dishonesty.

• Our entrepreneurial drive has long been our distinguishing trait.

• One should select the essential trait and reproduce it-or, even better, produce it.

• Pride seems to be one of our family traits.

• a genetic trait

• It's a human trait to joke about subjects that make us uncomfortable.

• The mutation has no functional significance and controls no traits, researchers say.

• His most noticeable trait was his charm, which he could seemingly turn on at will.

• Natural selection is of traits favourable to the survival, not of individuals, but of successive generations.

• Certain personality traits make people more likely to become victims of violent crime.

2. apprehensive

/ˌæprɪˈhensɪv/ **adjective**

worried or nervous about something that you are going to do, or about the future

apprehensive about/of

We'd been a little apprehensive about their visit.

apprehensive that

I was apprehensive that something would go wrong.

Some had apprehensive looks on their faces.

-apprehensively adverb

‘What’s wrong?’ I asked apprehensively.

apprehensive

• Twenty minutes in his company had left Merrill feeling stretched and apprehensive.

• I must admit that before my baby was born I was very apprehensive about motherhood.

• No one need be apprehensive about their personal safety; everything is under control.

• Afterward the architects agreed that they had been very apprehensive about what it would look like.

• What were the neighbors and zoning board apprehensive about?

• And she felt just as apprehensive as she always did here.

• Although a little apprehensive at first of steering such a large boat, we settled into it remarkably quickly.

• Anne waved her off, watching, nervous and apprehensive, from the upstairs window.

• Dr Gottlieb reassures apprehensive patients that the operation is a simple procedure.

• The Secret Service gets apprehensive when people even walk on this part of the colonnade.

3. taunt

/tɔːnt $ tɒːnt/ **verb [transitive]**

to try to make someone angry or upset by saying unkind things to them → **tease**

taunt somebody about something

The other children taunted him about his weight.

taunt somebody with something

They taunted him with the nickname ‘Fatso’.

And he’ll believe you, will he?’ Maria taunted.

-tauntingly adverb

taunt

• Of course he wasn't, an inner voice taunted.

• They were accosted by three white youths who taunted and then attacked them.

• The older boys taunted Chris and called him a girl.

• Or maybe, as she'd taunted earlier, his actions were governed by boredom.

• She was held in jail overnight, and she alleges in her lawsuit that guards taunted her with ethnic slurs.

• Now the telephone had acquired a personality, sat on the shelf so smug, taunting her with its silence.

• He couldn't forget how they had taunted him about his appearance. • She went on taunting him until he lost his temper. • They taunt me and beat me. • When I didn't want to fight he would taunt me repeatedly. "Coward, " he would say, "coward, coward, coward...." • You can blast your buddies and taunt them verbally at the same time.
4. divulge /daɪˈvʌldʒ, də-/ verb [transitive] formal to give someone information that should be secret [SYN] reveal **divulge information/secrets/details etc (to somebody)** It is not company policy to divulge personal details of employees. **divulge that** Clare divulged that she was recovering from a nervous breakdown. **divulge what/where etc** The Pentagon refused to divulge what type of plane it was. divulge • She hinted of an important secret still to be divulged. • There also are secret ingredients that she will not divulge. • The contract forbids employees to divulge details of this work to anyone outside the company. • Yet the Committees can not force ministers and civil servants to divulge information. • A spokeswoman for the company would not divulge the salaries paid to top managers. • The other three companies refused to divulge their plans. • I thought - I thought the case would be solved without my needing to divulge this information. • She would never divulge to Mattie that she had been second choice when Judge Tembleton could not do it. • I'm afraid I cannot divulge what Jameson said to me.
5. lassitude /ˈlæsɪtjuːd $ ˈlæsətjuːd, -tuːd/ noun [uncountable] formal tiredness and lack of energy or interest [SYN] weariness lassitude • The candidates have been trying to lift voters out of their lassitude.
6. assiduous /əˈsɪdjuəs $ -dʒuəs/ adjective formal very careful to make sure that something is done properly or completely [SYN] meticulous

assiduous in

He was assiduous in his attendance at church.

assiduously adverb

Even young children worked assiduously for a reward.

assiduity /ˌæsəˈdjuːəti $ -ˈduː-/ noun [uncountable]

assiduous

• Such assiduous attention has never been directed at this market before.

• Members of congress increasingly came to believe that they could insulate themselves against electoral defeat by assiduous attention to constituency casework.

• But even in the midst of these minefields, assiduous efforts have been made to fashion a safe passageway.

• Each patient had three or four attendants that night devoted to providing him or her the most assiduous service.

• Maybe I was too young because I fear I was not an assiduous student.

• an assiduous study of Austen's writings

7. pernicious

/pəˈnɪʃəs $ pər-/ adjective formal

very harmful or evil, often in a way that you do not notice easily

the pernicious effects of poverty

the media's pernicious influence

perniciously adverb

pernicious

• In addition, Roycayrol and Cattan studied fundic biopsy specimens of 18 totally achlorhydric patients, 12 of whom had pernicious anaemia.

• Gastrin levels may also be elevated by pernicious anemia. duodenal ulcers, and after a meal.

• In 1937, she had felt she had had enough of the pernicious course of human history.

• How can a struggling organization escape this pernicious cycle of perpetual reaction and strife?

• the pernicious effects of advertising

• a pernicious lie

• And how did this pernicious self-image arise?

• We are so easily led to pernicious solutions.

• Unfortunately in 1971 I was not protected from that most pernicious word by those inverted commas.

8. indispensable

/ˌɪndɪˈspensəbəl/ adjective

someone or something that is indispensable is so important or useful that it is impossible to manage without them [SYN] essential

indispensable to

This book is indispensable to anyone interested in space exploration.

indispensable for/in (doing) something

Meat is not indispensable for maintaining a healthy diet.

Mobile phones have become an indispensable part of our lives.

indispensable

• For mountain-climbing a really good sleeping-bag is indispensable.

• If you're planning on going sightseeing around the old city, a guide is indispensable.

• Police dogs have proved indispensable in the war on drugs.

• She soon became an indispensable member of staff.

• A knowledge of classical music is indispensable to anyone who wants to apply for this job.

• The book will be indispensable to anyone who wishes to learn more about the British Royal Family.

9. renege

/rɪˈniːg, rɪˈneɪg $ rɪˈnɪg, rɪˈniːg/ verb [intransitive]

renege on an agreement/deal/promise etc

renege

• The agency says it relied on a government commitment to provide liquidity, but the government reneged.

• Lewis was perfectly correct, even politically correct, to insist that Bowe had reneged on a pledge to fight him first.

• The House of Lords ruled that the mutual insurer was wrong to renege on guarantees offered to about 90,000 pension policyholders.

• As a result, he wanted to renege on his binding letter of intent, which he signed.

• At the same time, Landau persuaded the Boss to renege on statements that he never would play arenas or stadiums.

• As the developer lurched toward bankruptcy, Prudential tried to renege on the deal.

• An early government commitment to keep conscripts away from frontline fighting was swiftly reneged upon last year.

10. os·ten·si·ble

/ɒˈstensəbəl $ ɑ:-/ adjective

seeming to be the reason for or the purpose of something, but usually hiding the real reason or purpose

ostensible reason/purpose/aim

The ostensible reason for his resignation was ill health.

ostensible

• The war was fought to remove a cruel dictator - at least that was the ostensible aim.

• Does Astrophil, who is the ostensible author, mean he loves truth or that it is true he is in love?

• These experiences left him skeptical of the ostensible benefits of closing the glass subsidiary.

• Ongoing problems seem to arise of their own accord, and then to spread through the ranks with no ostensible cause.

• He similarly assumes that exile players are only qualified for the clubs' ostensible countries of origin.

• Intel carefully skirted the issues of Pentium's price and delivery at its ostensible introduction last month.

• The ostensible reason for this outlay was that kerosene was the lighting fuel of the poor.

Appendix 6. The Ethical Review Approval Letter from The Education University of Hong Kong

香港教育大學
The Education University of Hong Kong

20 June 2019

Dr XIE Haoran
Assistant Professor
Department of Mathematics and Information Technology

Dear Dr Xie,

Application for Ethical Review <Ref. no. 2018-2019-0362>

I am pleased to inform you that approval has been given by the Human Research Ethics Committee (HREC) for your research project:

Project title: Digital Game-based Vocabulary Learning for English Foreign Language Learners

Ethical approval is granted for the project period from 20 June 2019 to 31 December 2020. If a project extension is applied for lasting more than 3 months, HREC should be contacted with information regarding the nature of and the reason for the extension. If any substantial changes have been made to the project, a new HREC application will be required.

Please note that you are responsible for informing the HREC in advance of any proposed substantive changes to the research proposal or procedures which may affect the validity of this ethical approval. You will receive separate notification should a fresh approval be required.

Thank you for your kind attention and we wish you well with your research.

Yours sincerely,

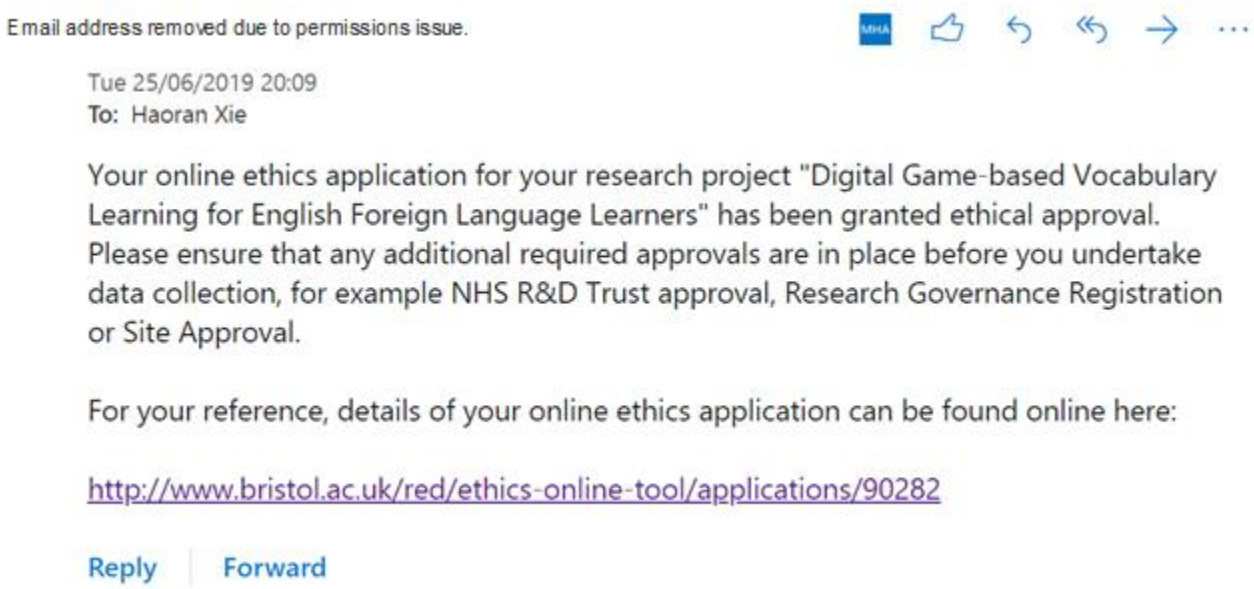
Signature removed due to permissions issue.

Patsy Chung (Ms)
Secretary
Human Research Ethics Committee

c.c. Prof CHOU Kee Lee, Chairperson, Human Research Ethics Committee

香港新界大埔露屏路十號
10 Lo Ping Road, Tai Po, New Territories, Hong Kong
T (852) 2948 8888 F (852) 2948 6000 www.eduhk.hk

Appendix 7. The Ethical Review Approval Letter from University of Bristol



Email address removed due to permissions issue.

Tue 25/06/2019 20:09
To: Haoran Xie

Your online ethics application for your research project "Digital Game-based Vocabulary Learning for English Foreign Language Learners" has been granted ethical approval. Please ensure that any additional required approvals are in place before you undertake data collection, for example NHS R&D Trust approval, Research Governance Registration or Site Approval.

For your reference, details of your online ethics application can be found online here:

http://www.bristol.ac.uk/red/ethics-online-tool/applications/90282

Reply | Forward

Appendix 8. SoE Research Ethics Form - University of Bristol

Name(s): Dr. XIE Haoran

Proposed research project: Digital Game-based Vocabulary Learning for English Foreign Language Learners

Proposed funder(s): Nil

Discussant for the ethics meeting: Prof. WONG Tak-Lam

Name of supervisor: Prof. YU Guoxing

Has your supervisor seen this submitted draft of your ethics application? Y

Please include an outline of the project or append a short (1 page) summary:

This study aims to identify the best learning strategies to enhance word learning for foreign language learners. There is a research gap between the literature and an important research issue of "the effectiveness and involvement load of a vocabulary learning task in digital games from the perspective of the ILH". Furthermore, all the three components (i.e., *Need*, *Search* and *Evaluation*) can be induced in vocabulary learning tasks according to through digital games to some existing studies. Therefore, we attempt to bridge the above research gap in this study. More specifically, the effectiveness of promoting the word retention by the same vocabulary learning task given in the paper-based form and the digital game form will be compared firstly.

The objective of this research is to fill the research gap of the unknown effectiveness of a vocabulary learning task or the different tasks with the same involvement load in digital games from the perspective of the ILH. In this study, the following two research questions will be addressed:

1) Is the same vocabulary learning task in a digital game learning environment more effective in word retention than conventional learning environment (i.e., paper-based instructions)?
2) Do two different vocabulary tasks (i.e., "reading a text and multiple-choice items on text" and "reading a text and choosing definitions") with the same involvement load in a digital game learning environment have similar effectiveness in word retention?

Ethical issues discussed and decisions taken (see list of prompts overleaf):

1. Researcher access /exit

The research will be conducted in a classroom at the university campus in Hong Kong. Students will be invited in the following channels:

a. I will have some posters on the university campus to recruit students.

b. I will promote this study via other social network platforms such as Facebook, WeChat and so on to recruit around students.

Students can contact the researcher through the above two channels to participate in the study on a voluntary basis. The student will take a pre-test, a learning task (in either a digital game form or a paper form), a face-to-face interview, an immediate post-test, and a delayed post-test about vocabulary learning in a classroom, which is quite similar to their common learning tasks/assigns in the classroom. The researcher will get access to and exit from participants in a normal circumstance. There is no potential risk for the students.

2. Power and participant relations

The relationship is teacher and student. There will be special precautions which will preserve the participants' rights to decline to join or withdraw from participation once the research has started.

3. Information given to participants

Participants in this study are students in universities. They will be informed that the purpose of this study to identify the best learning strategies to enhance word learning. The researcher will provide information about the detail procedures and instructions for taking learning tasks, tests and interviews. In addition, they will be informed that the project information including the format of study, data to be collected, the method of using data and their right of withdrawal.

4. Participant's right of withdrawal

Participants will be informed that they have every right to withdraw from the study before or during the measurement without penalty of any kind. If any student decides to withdraw, data generated from this participant will be eliminated by the researcher and the related data will not be included in this study.

5. Informed Consent

The researcher will prepare a consent form with an information sheet for the participant in this research. The researcher will introduce the project information and explain the rationale of this project to the participants. Participants will be informed that participation is voluntary. They will sign the consent form to express their consent.

6. Complaints procedure

If participants need to complain about any issues related to the research or have any concerns about the conduct of this study, they can directly contact the researcher or the related assistants. Furthermore, they can also contact the researchers' supervisor Prof. YU Guoxing, the Human Research Ethics Committee in the Education University of Hong Kong, or the SoE ethics committee at the University of Bristol.

7. Safety and well-being of participants/researchers

The research will be conducted in a classroom in the university campus. The settings of this study will be similar to common classroom teaching and learning activities. Participants will learn some new words through learning tasks without any loss. As the implementation of the study is similar to common activities in a campus classroom, there is no potential physical safety threat.

8. Anonymity/confidentiality

All information related to participants will remain confidential and will be identifiable by codes known only to the researcher or relevant assistants. Individual information will not be disclosed. The researcher and relevant assistants will comply with strict confidentiality.

9. Data collection

Data will be collected by using pre-tests, learning tasks, interviews, and post-tests. The detail procedures for data collection is included in section 3.3 experimental procedures in the research proposal.

10. Data analysis

The data collected from participants by using the pre-tests, learning tasks, interviews, and post-tests are then used for further analysis. The detail data analysis method is introduced in section 3.5 data analysis.

11. Data storage

The data collected from participants will be stored in the format of computer files and hard copies of transcripts. The computer files will be stored in researchers' office computer with password protection. Meanwhile, all test and interview scripts will be stored in a locked cabinet in the researcher's office.

12. Data protection (see: http://www.bristol.ac.uk/secretary/data-protection/)

The researcher and relevant assistants will comply with the following principles of Data Protection Act.

- Data may only be used for the specific purposes for which it was collected.
- Data must not be disclosed to other parties without the consent of the individual whom it is about unless there is legislation or other overriding legitimate reason to share the information (for example, the prevention or detection of crime). It is an offence for Other Parties to obtain this personal data without authorisation.
- Individuals have a right of access to the information held about them, subject to certain exceptions (for example, information held for the prevention or detection of crime).
- Personal information may be kept for no longer than is necessary and must be kept up to date.

- The departments of a company that are holding personal information are required to have adequate security measures in place. Those include technical measures (such as firewalls) and organizational measures (such as staff training).
- Subjects have the right to have factually incorrect information correct (note: this does not extend to matters of opinion).

13. Feedback

The results of this research will be published in the form of EdD thesis, journal articles, book chapters and/or conference proceedings. Individual information can be not identified and will be identifiable by codes known only to the researcher or relevant assistants. The results of the project will not be given to participants. The results of this project will not be given to any other persons or organizations other than the participants.

14. Responsibilities to colleagues/academic community

The research will strictly follow the code of academic integrity (http://www.bristol.ac.uk/students/study/teaching/integrity/). The results of this study will be accurately reported.

15. Reporting of research

The researcher will report the finding s of the research according to data obtained in an objective and critical manner to prevent any bias that may be infused.

If you feel you need to discuss any issue further or to highlight difficulties, please contact the GSoE's ethics co-ordinators who will suggest possible ways forward.

Signed: signature removed due to permissions issue. (Researcher) Signed: signature removed due to permissions issue. (Discussant)

Date: 24 June 2019

Appendix 9. Test Sample

The University of Bristol, Graduate School of Education,

& The Education University of Hong Kong

Department of Mathematics and Information Technology

Digital Game-based Vocabulary Learning for English Foreign Language Learners

****

Instructions:

1. If you cannot remember the meaning of the word, please tick (√) in Column [A].

2. If you can remember the meaning of the word, please write it in Column [B]. You can use either English or Chinese.

3. If you can write a sentence by using this word, please write one in Column [C]. Your sentence should include at least seven words.

		[A]	[B]	[C]
		I cannot remember the meaning of the word.□	I can remember the meaning of the word, it means:□	I can write the sentence by using this word, for example
1	trait			
2	apprehensive			
3	taunt			
4	divulge			
5	lassitude			
6	assiduous			
7	trait			
8	indispensable			
9	renege			
10	ostensible			

Appendix 10. Student Interview Sample Questions

The University of Bristol, Graduate School of Education
The Education University of Hong Kong
Department of Mathematics and Information Technology

Digital Game-based Vocabulary Learning for English Foreign Language Learners

Student Interview Sample Questions

****

1. To what extent do you agree with the following statements about the ADVANTAGES of this learning task?

2. To what extent do you agree with the following statements about the DISADVANTAGES of this learning task?

3. What kind of elements or contents in learning tasks are most useful for you?

Appendix 11. Student Consent Form

Consent Form and Information Sheet

THE UNIVERSITY OF BRISTOL

Graduate School of Education

THE EDUCATION UNIVERSITY OF HONG KONG

Department of Mathematics and Information Technology

STUDENTS' CONSENT TO PARTICIPATE IN RESEARCH

Digital Game-based Vocabulary Learning for English Foreign Language Learners

I _____ hereby consent to participate in the captioned research conducted by the Principal Investigator - Dr. XIE Haoran （謝浩然博士）.

I understand that information obtained from this research may be used in future research and may be published. However, my right to privacy will be retained, i.e., my personal details will not be revealed.

The procedure as set out in the **attached** information sheet has been fully explained. I understand the benefits and risks involved. My participation in the project is voluntary.

I acknowledge that I have the right to question any part of the procedure and can withdraw at any time without penalty of any kind.

Name of participant	_____
Signature of participant	_____
Name of Researcher	XIE Haoran
Signature of Researcher	_____
Date	_____

Appendix 12. Information Sheet

INFORMATION SHEET

Project Name: Digital Game-based Vocabulary Learning for English Foreign Language Learners

Your class is invited to participate in the captioned project conducted by the principal investigator - Dr. XIE Haoran, who is an EdD candidate of the Graduate School of Education in The University of Bristol and an academic staff of the Department of Mathematics and Information Technology in The Education University of Hong Kong.

This research project aims to identify the best learning strategies to enhance word learning for foreign language learners. The project is going to be carried out from June 2019 to Dec 2020. Both tests and interviews will be collected at the beginning and end of the study respectively. The test will last no more than 10 minutes; the interview will last 20 minutes or so. The questions will not infringe the participant's privacy and will not make the participant feel uncomfortable. About 100 participants will be recruited in this study. Your contact details are normally obtained by us through the following three ways:

The researcher of this study will have some posters in the university and promote it via social networks (e.g., Facebook, WeChat etc).

If participants in these activities are interested and agree to join the research study, their contact details will be obtained by research team members and used in academic purpose only. All the data collected will be kept strictly confidential.

The participant's name and a number will not be revealed in any research reports. Only the main researchers can get access to the data solely for data analysis and academic writing purposes.

If you are interested and agree to join the research study in above ways, your contact details will be obtained by research team members. All the data collected will be kept strictly confidential. The participant's name and number will not be revealed in any research reports. Only the Principal Investigator and researchers can get access to the data solely for data analysis and academic writing purposes.

There is no potential risks will be involved in this study. The results of this study may be published disseminated in conference papers or journal articles. All data will not bear individual names.

You have every right to withdraw from the study before or during the measurement without penalty of any kind. All information related to you and your students will remain confidential and will be identifiable by codes known only to the researcher.

If you have any concerns about the conduct of this research study, please do not hesitate to contact the researcher Dr. XIE Haoran by email email address removed due to permissions issue at email address removed due to permissions issue, his EdD supervisor Prof. YU Guoxing by email at email address removed due to permissions issue at the University of Bristol, the School of Education ethics committee in the

University of Bristol by email at gsoe-ethics@bristol.ac.uk, or the Human Research Ethics Committee by email at hrec@eduhk.hk or by mail to Research and Development Office, The Education University of Hong Kong.

If you would like to obtain more information about this study, please contact Dr. XIE Haoran at telephone number telephone removed due to permissions issue.

Thank you for your interest in participating in this study.

Yours sincerely,

Dr. XIE Haoran
Principal Investigator