Barry Lee Reynolds *Editor*

Vocabulary Learning in the Wild

Springer

Vocabulary Learning in the Wild

Barry Lee Reynolds
Editor

Vocabulary Learning in the Wild

Springer

Editor
Barry Lee Reynolds
University of Macau
Taipa, Macau SAR

ISBN 978-981-99-1489-0 ISBN 978-981-99-1490-6 (eBook)
https://doi.org/10.1007/978-981-99-1490-6

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

To Ian, my inspiration and rock.

Foreword

One of the biggest obstacles to learning in foreign language classes is the often unstated belief that learning relies on teaching. In an ideal world, the majority of learning in a foreign language course should occur without teaching but should occur through the engagement of the learners with use of the foreign language, receptively and productively, with a large part of this engagement occurring out-of-class. This engagement can occur through extensive reading, extensive listening and viewing, use of the internet, gaming, social networking, and study abroad.

Out-of-class learning can have various kinds of relationships with the classroom (Nation, 2022 Chapter 6). It can involve learning that has no direct relationship with the classroom and is not under the control of the teacher. This *learning in the wild* or *extramural English* is the subject of this very timely volume. Out-of-class English can also involve learning that is related to a course of study, *extra-curricular learning*, and is intended to supplement classroom work. The third kind of relationship occurs when out-of-class learning is the language course, *self-directed learning*.

Learning in the wild is strongly affected by the relationship between the learners' first language and the foreign language. Much of early work on learning in the wild involved the learning of English by learners whose first language was closely related to English and the presence of cognates had a strong effect on learning. However, the research showed that a lot of learning of words that were not cognates also occurred. Clearly, learning in the wild can be an important way of learning another language.

Research on learning vocabulary in the wild is in its infancy, but the large amount of research on vocabulary learning in general suggests useful directions for such research, and some of the papers in this volume have begun to head in these directions. A very fruitful direction for research is to examine the opportunities for such learning. This corpus-based research can be a useful first step for seeing if the means for languages in the wild, such as gaming, social networking, listening to music, and watching movies, provides enough opportunities for vocabulary and multiword units to be learned through quantity of language input and productive use. Some opportunities for learning in the wild involve interaction, and it is likely that certain kinds of interaction will support learning more than other kinds. It is also likely that involved observation of others using the language will be a fruitful source of learning, with the detachment of observation allowing more focus on language features (DeHaan, Reed & Kuwada, 2010).

Research on learning the vocabulary of English as a first language has been hampered by the use of tests, such as the Peabody Picture Vocabulary Test, which do not give an indication of total vocabulary size. Researching the learning of a foreign language in the wild would be greatly enhanced by the use of measures that give some indication of how far learners have progressed along the journey of learning the vocabulary of the language. This would give a very helpful indication of how much vocabulary learning occurs through learning in the wild beyond the words actually tested.

The pervasive learning principles of spaced repetition, quantity of attention, variation, elaboration, and analysis apply to learning in the wild in the same way that they apply to deliberate learning and incidental learning in the classroom. Research on learning in the wild will shed light on how these principles work in this very different kind of environment. The papers in this volume are a very welcome contribution to developing our understanding of this exciting new area of research.

Paul Nation
Victoria University of Wellington
Wellington, New Zealand

References

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), 79–94.

Nation, I. S. P. (2022). *Learning Vocabulary in Another Language.* 3rd edition. Cambridge University Press, Cambridge.

Preface

I can trace my interest in informal and incidental language learning back to when I was studying for my master's in TESOL at Murray State University in the USA. At the time, I was heavily influenced by my TESOL professors' emphasis on task-based language teaching, content-based language teaching, and English for academic purposes. Besides being a master's student, I was also an ESL instructor teaching academic reading skills to international students. Given the liberty to teach an academic reading course using materials I developed on American cuisine inspired me to incorporate tasked-based language teaching approaches into my language lessons. I was further granted permission to use a staff kitchen to engage students in cooking activities. This is probably when I first saw the power of informal language learning. Of course, my students were engaging in language learning activities because I had asked them to do so, but these activities also gave them a real purpose for using the language. I soon discovered that the language they learned through activity completion went well beyond those targets I set for the lessons. My students were genuinely engaged and enjoying the interaction with their classmates while preparing different American dishes. Through the negotiation of meaning, they were *picking up* new words, phrases, and grammar incidentally. Still etched in my memory is the day my students happily pieced together a recipe I cut up for a jigsaw reading task and then, using the reconstructed recipe, they first prepared and then baked a pizza. These ESL teaching experiences eventually led me to select *learning English through gastronomy* as a TESOL course project. At the time, I had yet to read the related research on incidental learning, but I knew from a language teacher's point of view that engaging students in activities they enjoy would result in nearly effortless language learning.

After those experiences in the USA, I relocated to Taiwan where I was engaged as a university lecturer teaching skill-based language courses and English for specific purposes. During that time, many students would approach me wanting advice on how they could improve their English. My answer never faltered: "Start a new hobby that requires you to use English to engage in the hobby." If they wanted to interact with other language learners, I shared resources such as online forums, chatrooms, and online multimedia sites. If they preferred more independent hobbies, I encouraged activities such as watching movies, watching TV, listening to music, playing video games, and reading comics, among many others. I gave those answers because I knew they were already spending class time learning linguistic structures but what students lacked was putting that learned language to use.

Fast forward a couple of years later to when I was pursuing a topic for my PhD study. For about a year, I dabbled in video games and other technologies for language learning before eventually settling on researching the potentials of reading for incidental vocabulary acquisition. I devoured the research literature, finding a lot of inspiration from the work of previous researchers, especially critical work by Tom Cobb and the early work of Paul Nation. Much of this synthesized research helped to reinforce the ideas and experiences I had as a language teacher. A pivotal moment occurred at the 2012 BAAL conference. After presenting part of my PhD research, Paul Meara approached me to offer several encouraging words of support. His acknowledgment of my potential for conducting vocabulary research solidified my research trajectory. Those words inspired me in ways I am sure he never anticipated.

Now at the University of Macau, I have been given the privilege of pursuing my own research agenda and have taken on a new role as an English teacher educator. When planning studies, I find myself drawing inspiration from those earlier experiences with my ESL students, revisiting ideas about game-based language learning, and considering research on learner engagement in hobbies as a route to proficiency in a target language. While all the topics covered in this book are relevant to the interests of contemporary vocabulary researchers, their inclusion also represents some of my current research interests as well. The editing of this book has reaffirmed my stance that language learners must take advantage of opportunities that the physical and digital "wild" provides to increase their vocabulary learning potential. My use of the term *vocabulary learning in the wild* in this edited volume attempts to bring together several related fields of informal and incidental vocabulary learning. A bridge between classroom-based learning and out-of-class learning must be built. The seed for this book came from the after-class discussions I had with in-service and pre-service teachers in Macau wondering how they could get their learners more exposure to the language outside of class time. It is my wish that language teachers can help students explore second language vocabulary outside the formal classroom.

The book contains an introduction and four parts. Conducting research and editing volumes like this one allows me to develop a better understanding of different areas of vocabulary research. I now know much more about informal language learning, online informal learning of English, informal digital learning of English, extramural English, digital game-based learning, incidental language learning, and nonformal language learning. The introduction provides readers a scoping review of second language vocabulary learning in the wild. Part 1 consists of three chapters on informal vocabulary learning in formal education contexts. Part 2 contains three chapters on vocabulary learning from digital gaming. Part 3 is comprised of four chapters on vocabulary learning from video viewing, and Part 4 ends the volume with two chapters on nonformal vocabulary learning. Beyond adding to our collective knowledge of vocabulary learning in the wild, the chapter authors also provide practical advice on how second language teachers can integrate "wild learning" into their formal classroom teaching. I hope readers find the chapters intellectually inspiring and rich with opportunities for learning.

September 2022

Barry Lee Reynolds
University of Macau
Taipa, Macau SAR

Acknowledgments

A sincere thank you goes to the reviewers of the chapters and the anonymous reviewers of the entire manuscript for their commitment to its publication. Their valuable advice and feedback ensured the authors were held to a high standard.

María Pilar Agustín-Llach, University of La Rioja
Judith Borràs, University of Lleida
Dale Brown, Kanazawa University
Andrés Canga Alonso, University of La Rioja
Yu-Hua Chen, Coventry University
Gilbert Dizon, Himeji Dokkyo University
Fan Fang, Shantou University
Clarence Green, Federation University
Rui (Tammy) Huang, University of Florida
Long He, Iowa State University
Nurul Aini Mohd Jelani, Victoria University of Wellington
Rui Li, Hunan University
Qing Ma, The Education University of Hong Kong
Charles M. Mueller, Fuji Women's University
Ulugbek Nurmukhamedov, Northeastern Illinois University
Rebekah Rast, The American University of Paris
Sophia Skoufaki, University of Essex
Nathan Thomas, University College London
Yukio Tono, Tokyo University of Foreign Studies
Wen-Ta Tseng, National Taiwan University of Science and Technology
Sharon H. Ulanoff, California State University, Los Angeles
Boris Vazquez-Calvo, University of Malaga
Duy Van Vu, KU Leuven

Matthew P. Wallace, University of Macau
Martin Willis, Tokyo Woman's Christian University
Kevin Wong, Pepperdine University
Hua (Flora) Zhong, Western Sydney University

Contents

Editor and Contributors

About the Editor

Barry Lee Reynolds is Associate Professor of English Language Education in the Faculty of Education at the University of Macau. His research interests include vocabulary learning, computer-assisted language learning, written corrective feedback, and language teacher education. More specifically, his vocabulary research has focused on incidental learning from reading novels, playing video games, text reconstruction, and listening to academic lectures. He has also conducted vocabulary studies on word card use, reader choice, lexical coverage, contextual richness, perception, extramural engagement, and metacognition, among others.

Contributors

Griet Boone Ghent University, Ghent, Belgium

Duygu Candarli University of Southampton, Southampton, UK

Emma Chen Western Washington University, Bellingham, WA, USA

June Eyckmans Ghent University, Ghent, Belgium

Ferran Gesa Department of Linguistic and Literary Education, Universitat de Barcelona, Barcelona, Spain

Julian Heidt Carleton University, Ottawa, ON, Canada

Mangeet Kaur Khera University of Nottingham Malaysia, Semenyih, Malaysia

Kriss Lange University of Shimane Matsue Campus, Matsue, Japan

Phoebe Lin Department of English and Communication, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

Joshua Matthews University of New England, Armidale, Australia

Jessica McLaughlin University of Nottingham Malaysia, Semenyih, Malaysia

Imma Miralpeix English Studies Department, Universitat de Barcelona, Barcelona, Spain

Chi-Duc Nguyen VNU University of Languages and International Studies, Vietnam National University, Hanoi, Vietnam

Geoffrey G. Pinchbeck Carleton University, Ottawa, ON, Canada

Debbie Pushor University of Saskatchewan, Saskatoon, SK, Canada

Amin Rasti-Behbahani Department of Language and Communication Studies, University of Jyvaskyla, Jyvaskyla, Finland;
College of Arts and Applied Sciences—Foundation Program, Dhofar University, Salalah, Oman

Arathi Jane Reddy University of Nottingham Malaysia, Semenyih, Malaysia

Barry Lee Reynolds University of Macau, Taipa, Macau SAR

Michael P. H. Rodgers Carleton University, Ottawa, ON, Canada

Maria-del-Mar Suárez Department of Linguistic and Literary Education, Universitat de Barcelona, Barcelona, Spain

Csaba Szabo University of Nottingham Malaysia, Semenyih, Malaysia

Ching-Wen Wang The Language Center, Chaoyang University of Technology, Taichung City, Taiwan

Gunther M. Wiest ADA University, Baku, Azerbaijan

Kevin Reay Wrobetz Department of Business Administration, Kobe Gakuin University, Kobe, Japan

Introduction: A Scoping Review of Second Language Vocabulary Learning in the Wild

Barry Lee Reynolds

Abstract This chapter reports a scoping review of research on vocabulary learning in the wild published between January 2010 and November 2021. The aims of the review were to examine the extent of research activity and identify gaps in the existing vocabulary learning in the wild literature. Seventy-three title keywords refined by three Web of Science categories in the SSCI and ESCI databases were searched. The search yielded 257 journal articles, of which 51 met a set of inclusion criteria and were coded. First, the numerical information was used to provide an overall summary of the extent, nature, and distribution of the articles. Then, the literature was organized thematically to synthesize and summarize the articles' aims and results. The reviewed studies investigated (incidental) vocabulary learning through engagement or the completion of activities; learner factor effects; language input factor effects; effects on variables other than vocabulary knowledge; technology and vocabulary learning; and bridging incidental and intentional vocabulary learning. However, the reviewed studies underrepresented languages other than English and learners at the pre-primary and primary levels of education. There was a paucity of studies that assessed the productive form, productive use, and receptive use aspects of vocabulary knowledge.

Keywords Incidental vocabulary learning · Informal vocabulary learning · Vocabulary activities · Vocabulary assessment · Scoping review

1 Introduction

Vocabulary learning[1] is a lifelong process. Words come in and out of fashion and are constantly being created to describe new and evolving phenomena (Algeo, 1980). Thus, regardless of whether or not learners have obtained mastery in a second/foreign/additional language (L2 from hereafter), regularly using that

[1] The terms learning and acquisition are used interchangeably to refer to the same construct.

B. L. Reynolds (✉)
University of Macau, Taipa, Macau SAR
e-mail: BarryReynolds@um.edu.mo

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_1

language will result in encountering unknown words *in the wild* needed for comprehension of spoken and written language input.[2] Using *in the wild* to describe this phenomenon emphasizes the spontaneous and informal nature of such incidental exposure to unknown words. Dressman's (2020) definition of informal language learning has been adapted to operationalize *vocabulary learning in the wild* for the present study. Specifically, it "includes [engagement in] all activities undertaken by learners outside a formally organized program of language instruction" that consciously or unconsciously results in the learning of vocabulary knowledge (Dressman, 2020, p. 4). This learning of unknown words are by-products of incidental exposure *in the wild* (Hulstijn, 2012).

Bringing attention to vocabulary learning in the wild stresses the public knowledge that all the words one understands in a first or second language has rarely been gained within the confines of a classroom and instead has occurred through incidental exposure outside the classroom (Nagy et al., 1985). Taking cues from Dressman (2020), who advocated an encompassing view of informal language learning, the current scoping review considers vocabulary learning in the wild to encompass vocabulary learning that has been referred to in the previous literature as incidental vocabulary learning, informal vocabulary learning, and nonformal vocabulary learning, among others. The aim of the present scoping review was not to make claims that these terms refer to the same construct. Instead, the term vocabulary learning in the wild provides an inclusive label of all the myriad ways learners engage in informal learning of vocabulary.

While some researchers have been extoling the benefits of incidental exposure to language for many years through academic discussions (e.g., Krashen, 1985), views given in these discussions were not always considered mainstream. As language learners take more control of what, when, and how they learn, informal, incidental, and extramural vocabulary learning is moving from the peripheral to the center stage of academic discussions (Dressman, 2020). In other words, it is learner agency that connects the concepts of incidental vocabulary learning, informal vocabulary learning, and nonformal vocabulary learning. For example, Peters and Muñoz (2020) recently edited a special issue for *Studies in Second Language Acquisition* on language learning from multimodal input. The academics that published papers within that issue discussed, among other topics, how input from TV viewing, reading, listening, and reading-while-listening led to the incidental acquisition of different aspects of vocabulary knowledge.

Reynolds and Teng (2021) also edited a special issue for *TESOL Journal* on incidental and informal vocabulary learning with most of the articles focusing on bridging the vocabulary learning in the classroom to activities outside the classroom. The articles discussed what vocabulary should be taught inside classrooms and what vocabulary should be incidentally acquired outside the classroom, vocabulary learning

[2] L2 is used for simplicity's sake and without the intention of suggesting a deficient or outsider perspective about language learning (Dewaele, 2017). Instead, the use of L2 in this chapter refers to any additional language other than one's first language (Anderson, 2022).

through video viewing, incidental acquisition of derivational knowledge, and incidental vocabulary acquisition from classroom language, among others. Likewise, Dressman and Sadler's (2020) *The Handbook of Informal Language Learning* is probably the most up-to-date volume covering relevant informal language learning issues spanning from theory to digital contexts, learning abroad, and the mixing of formal and informal language learning contexts. While due attention has been delayed, the research is steadily catching up to provide a clearer picture of the vocabulary learning potentials through engagement in varied activities in the wild.

The academic community's delay in legitimizing the field of vocabulary learning in the wild could be a result of much of the vocabulary research relying on a number of assumptions. Corpus studies, for example, have provided estimates for vocabulary sizes necessary to use a language to engage in particular activities. We now know that a 5,000 word vocabulary is recommended for English video game play (Rodgers & Heidt, 2020) and for comprehension, a 6,000–7,000 word vocabulary is needed for spoken texts and 8,000–9,000 word vocabulary is needed for written texts (Nation, 2006). However, these and other studies have often been conducted using field conventions (e.g., lexical coverage figures) that some may argue have not been replicated enough to be accepted as a rule (Song & Reynolds, 2022).

A similar situation happened decades ago when researchers endeavored to explain the large vocabulary sizes obtained by native English speakers (Nagy & Anderson, 1984). Their main argument was that English native speakers engage in a large amount of reading during school years, and this reading engagement presumably is the source of the vocabulary growth. However, indirect findings could only be used to marginally support this *default hypothesis* of vocabulary acquisition (Nagy, 1997). The default hypothesis assumes large gains in first language vocabulary knowledge are the result of incidental exposure through reading and other input sources outside the formal classroom (Nagy et al., 1985). In other words, the support of this hypothesis hinged on the observable phenomenon that teachers did not spend much time on direct teaching of vocabulary and students did not spend much time on direct learning of vocabulary—not on any proof of the actual source of the vocabulary learning (Nagy & Anderson, 1984). Therefore, the nonexistence of proof for intentional vocabulary learning was the purported proof for the existence of incidental vocabulary learning.

L2 researchers, working mostly with learners of English, have since replicated these results, finding incidental gains in vocabulary knowledge, mostly modest, from reading (e.g., Pitts et al., 1989) and other tasks (e.g., Soyoof et al., 2022). Still, there has been some criticism of the default hypothesis. The *default argument* is weakened when the paucity of target language use in foreign language contexts are considered (Laufer, 2005). Nevertheless, such criticism seems unsupported in modern societies where an abundance of language input is readily available and opportunities for face-to-face and digital interactions with others occurs daily in many foreign language contexts (Dressman, 2020). Likewise, at university it is increasingly becoming the norm to study abroad or to pursue a dual degree where part of one's studies is completed abroad (Boone, 2021). The affordances provided by modern travel and the Internet have made the world more connected and accessible, thereby increasing incidental exposure to target languages and incidental vocabulary learning through this exposure (Soyoof et al., 2021).

While empirical research is the most robust route to understanding how language learners acquire substantial target-language vocabulary sizes, there is a lot to be said about gathering an informal understanding of vocabulary learning processes through conversations with advanced language users (Dressman, 2020). Without ignoring possible exceptions, most advanced language users when prompted to provide an explanation for how they learned so many words in a L2 will be able to share vivid memories of picking up vocabulary from watching movies, listening to music, reading books and magazines, playing video games, and possibly interacting with native and non-native speakers locally, abroad, and online.

Highlighting these opportunities for informal and incidental learning of vocabulary through these interactions is not to deemphasize or lesson the role of direct learning or teaching of vocabulary. In contrast, there is ample evidence that direct learning and teaching of vocabulary is highly recommended, especially for the most frequent words of a language; direct teaching of high frequency words facilitates the development of learner autonomy because knowledge of these words provides learners the means to engage in independent learning outside the classroom (Nation, 2013, 2022). This process has been referred to as a virtuous circle of vocabulary growth (Pulido & Hambrick, 2008).

Considering the current acceptance of informal language learning as a new field of second language research and the interconnectedness of the global village, conducting a scoping review of vocabulary learning in the wild has relevance to the following stakeholders: informal language learning researchers, vocabulary acquisition researchers, language teachers, and language learners. Unlike the previous systematic reviews and meta-analyses that have been limited by a single theoretical stance, research design, or variable of interest, the current review provides a more complete picture of vocabulary learning in the wild.

The aims of the current scoping review were to examine the extent of vocabulary learning in the wild research activity and identify research gaps in the existing literature. Furthermore, the scoping review can underscore current informal vocabulary learning trends while highlighting particular areas that deserve more attention. Synthesizing and summarizing the literature can provide useful information to language teachers that can be used to help learners bridge the gap between in-class and out-of-class learning. This scoping review was guided by one overarching research question and seven specific research questions:

- What is known from the existing literature about vocabulary learning in the wild through engagement in different activities?
  - Who conducted the studies? When and where were they conducted?
  - What types of participants (i.e., educational level, L1, L2) were recruited?
  - What activities were investigated? What were the durations of the activities?
  - What methodologies and theoretical frameworks were used?
  - What type(s) of vocabulary knowledge were measured?
  - What were the aims of the studies?

– What were the important findings of the studies?

2 Previous Vocabulary Learning in the Wild Related Reviews

The sustained interest in vocabulary learning in the wild issues is clear by the large number of research syntheses and meta-analyses published in the last twelve years. While excluded from the analysis for the current scoping review, several research syntheses and theoretical articles were retrieved from the database searches, organized based on their foci (i.e., language activity/input type), and are summarized in this section. All of these studies fell within the scope of vocabulary learning in the wild that was defined earlier in this chapter. Their findings highlight issues pertinent to the current scoping review: task completion, reading, listening and reading with glosses, listening to spoken language, video viewing, nonformal learning with flashcard software, and technology-mediated learning.

2.1 Task Completion

A sizeable portion of vocabulary research has aimed to understand how the completion of different learning tasks lead to varying amounts of vocabulary learning. Researchers have most often used the *Involvement Load Hypothesis* to frame these studies. The Involvement Load Hypothesis is composed of one motivational (i.e., need) and two cognitive (i.e., evaluation, search) components. It provides researchers with a simple method of predicting vocabulary learning outcomes from completing different tasks (Laufer & Hulstijn, 2001). Researchers compute involvement load scores for different tasks based on whether each component is missing (0), weakly present (1), or strongly present (2) in these tasks. Calculating the involvement load for different tasks can predict whether completing one task will result in more vocabulary learning than another task.

To calculate trustworthy involvement load scores requires variables other than need, evaluation, and search (i.e., potential moderators) to be the same among the compared tasks. However, these moderating variables can be controlled and further investigated with advanced statistical techniques. Huang et al. (2012), for example, conducted a meta-analysis of 12 incidental vocabulary acquisition studies, finding learners who completed output tasks (i.e., writing) learned more vocabulary than those that completed input tasks (i.e., reading). However, learners who read more than a single text genre gained more vocabulary than those who only read a single text genre. In addition, the longer learners spent on task completion, the more vocabulary they learned. Huang et al. (2012) did not find a significant difference in vocabulary gains between participants who read texts with 95% lexical coverage and those who read texts with 98% lexical coverage. This result is important as incidental vocabulary acquisition researchers have suggested learners be given texts with a lexical coverage (i.e., percentage of known words in a text) of either 95% or 98% in order for contexts to allow learners to infer the meaning of unknown words encountered (Hu & Nation, 2000; Laufer, 1989). Overall, this meta-analysis supported the claims made by the Involvement Load Hypothesis.

More recently, Yanagisawa and Webb (2021) conducted a meta-analysis of 42 studies to further explore the predictive ability of the Involvement Load Hypothesis while also investigating the potential moderating effects of several variables. They found the evaluation component of the Involvement Load Hypothesis explained learners' vocabulary learning from task completion more than the need component; the search component did not appear to have contributed to vocabulary learning in the studies they reviewed. Furthermore, they found the type of vocabulary knowledge assessed (i.e., test format) and frequency of exposure to targeted words moderated vocabulary learning gains. Interestingly, involvement load scores still affected vocabulary learning more than time on task. As the Involvement Load Hypothesis continues to be the theoretical framework used for most of the newly published vocabulary research, Yanagisawa and Webb (2021) suggest continuing empirical research that specifically focuses on increasing our understanding of its predictive effects.

2.2 Reading

Reading leads to vocabulary growth (Nation, 2022). This vocabulary growth has been considered to have occurred incidentally as learners engaged in reading are focused on comprehending the texts and not learning vocabulary (Nagy, 1997). Restrepo Ramos (2015) conducted a literature review with the aim of defining incidental vocabulary acquisition, uncovering the strategies that lead to more robust incidental vocabulary acquisition outcomes, and highlighting potential future vocabulary research trajectories. Restrepo Ramos (2015) reported that much of the incidental vocabulary acquisition research literature discusses how inferencing from informative contexts can lead to lexical gains. Furthermore, these gains can be further enhanced through multimodal glosses due to their raising readers' awareness. Restrepo Ramos (2015) suggests researchers to push our understanding of incidental vocabulary acquisition forward by focusing future research on incidental exposure to vocabulary through listening and the use of technology. Moreover, as vocabulary is not always learned as individual words (Snoder & Reynolds, 2019), Restrepo Ramos (2015) suggested researchers to give more attention to the learning of multiword units such as collocations.

In a later review, Thomas (2020, p. 49) underscored that "reading is often the activity that researchers use to generate their findings" about incidental learning. While he does note that more recently listening and multimodal input have been receiving researchers' attention, these studies still indicate that participants focus more on the text than the aural or visual input. However, not all texts are the same and rich contexts along with multiple exposures to unknown words are more conducive to incidental learning of the unknown words. And while it appears that practitioners and researchers are increasingly becoming accepting of the notion that inferring unknown words through context while reading can lead to sizeable vocabulary growth (e.g., Chung, 2018), several mediating factors may reduce or enhance the likelihood of incidental acquisition, including "percentage of known words in the co-text and context, target word salience, dynamic exposure, depth of processing, and learning motivation," among others (Thomas, 2020, p. 56).

Published around the same time as Thomas (2020), Wang's (2020) literature review on the concerns, progresses, and future directions of incidental L2 vocabulary acquisition and reading also point out several relevant variables that moderate the amount of vocabulary incidentally acquired through reading. These included the quality of processing, quantity of processing, the input mode, contextual richness, lexical coverage, reading proficiency, text comprehension, input enhancement, and whether learners are required to only read or read before completing another task. Echoing several of the suggestions for future research offered by Thomas (2020) and Restrepo Ramos (2015), Wang (2020) expressed a desire for researchers to investigate the incidental acquisition of multiword units, especially from multi-modal e-learning technologies by using cognitive science data collection techniques such as eye-tracking, electroencephalography, and functional magnetic resonance imaging.

2.3 Listening and Reading with Glosses

Research has shown learners can pick up the meaning of previously unknown vocabulary they encounter while reading by using context to infer meaning (Nagy et al., 1985). However, this incidental learning is often small and fragile as it can sometimes be difficult to infer full or targeted meaning from context (Nation, 2022). Glosses have been offered as a route to boost the likelihood of incidental learning of unknown vocabulary encountered while reading. A gloss is usually a short definition or a translation that is added to a reading passage to assist with comprehension (Yanagisawa et al., 2020). A multimedia gloss goes one step further by providing information that could help to comprehend reading or listening material in the form of audio, visual, audiovisual, or other media (Mohsen & Balakumar, 2011). Glossing studies are relevant to the present scoping review as many learners, especially those from foreign language contexts, acquire a lot of their vocabulary through meaning-focused comprehensible input from reading and listening (Nation, 2013). The three studies reviewed below synthesized research on multimedia glosses, multimodal glosses, and all gloss types.

Mohsen and Balakumar (2011) conducted a narrative review on the effects of multimedia glosses for vocabulary learning by participants engaged in listening and reading tasks. They operationalized multimedia glosses as a type of gloss that "provide learners with different modalities (textual, visual, and auditory) and modes (video, picture, and text)" (Mohsen & Balakumar, 2011, p. 136). The 18 journal and conference papers published between 1993 and 2009 showed multiple annotations were more effective than a single or no annotation condition. As their study was conducted in order to offer suggestions for conducting future research, they pointed out a number of limitations that should be addressed. Multimedia glosses used for listening tasks were limited to aural glosses without the use of audiovisual material. Furthermore, most studies were conducted with tertiary-level participants thereby limiting generalizability to other participant groups. Lastly, few studies assessed productive vocabulary knowledge and even fewer assessed words that did not have a concrete meaning. They concluded that more research should be conducted to understand whether multimedia glossing is useful for abstract word learning for young learners to acquire productive vocabulary knowledge.

More recently, Ramezanali et al. (2020) conducted a meta-analysis of 22 multimodal gloss reading studies finding that dual glosses were more effective than a single mode for vocabulary learning. “Multimodal glossing is viewed as any triple gloss combination (e.g., definitions + audios + pictures, definition + pictures + videos)” (Ramezanali et al., 2020, p. 106). However, there was no noticeable benefit for a triple-mode gloss. Learners can benefit from two modes as the input can be processed in a dual mode; however, they suggested the third mode is either ignored due to dual mode processing limitation or unable to be processed due to cognitive overload.

Although Ramezanali et al. (2020) found a positive effect for dual glosses, this effect should be considered in connection with moderator variables. This is because moderating variables have the potential to strengthen, weaken, or negate the relationship between the independent and dependent variables under investigation. They focused their attention on a large number of moderator variables in their meta-analysis. Learners in secondary schools and language institutes seemed to have benefited more from the addition of a gloss mode than tertiary-level learners. Lower proficiency learners also seemed to benefit more from the addition of a gloss than more proficient learners. However, it should be noted that the additional gloss mode seemed to work better for expository texts compared to narrative texts. Furthermore, when an additional gloss was added, the effect was larger if the gloss was provided in the L2 rather than the L1. Lastly, lexical coverage (i.e., the percentage of known words in the read texts) seemed to reduce the effectiveness of adding an additional gloss. This meant the addition of a gloss could reduce the negative effect of a low lexical coverage of known words in the read texts.

The most recent meta-analysis on the effects of glossing was conducted by Yanagisawa et al. (2020). They defined glossing more generally as “providing L1 translations, L2 synonyms, or short explanations of unfamiliar words in a text” (Yanagisawa et al., 2020, p. 412). Examining 42 studies found glossing to be effective regardless of the type of vocabulary assessment used. Glosses that contained multiple choices were most effective. This could be due to learners’ engagement with the form and meaning of the glossed words. L1 glosses and L1 plus L2 glosses were similarly effective and L2 glosses were least effective. They did not find higher proficiency learners to benefit from glosses more than lower proficiency learners; instead, it appeared that the intermediate proficiency learners benefitted the most from glosses. There was no clear difference in the effect of different modes of glossing (e.g., textual, pictorial, auditory) and participants' level of schooling was not found to be a significant moderator of glossing effectiveness. Comprehension of the texts read by learners moderated the effectiveness of glosses while material type target (i.e., native speakers vs. learners), text type, and CALL use (i.e., digital text) were not significant moderators of glossing effectiveness.

2.4 Listening to Spoken Language

Language learners are exposed to an abundance of spoken language input. Spoken input includes listening to stories or audiobooks, watching videos, and conversing with a peer to solve a problem, among many other sources. This spoken language does not necessarily have to be focused on vocabulary in order for vocabulary to be incidentally acquired.

De Vos et al. (2018) conducted a meta-analysis to understand whether spoken language input in the form of audio, audiovisual, interaction tasks, and noninteractive tasks led to the greatest gains in vocabulary knowledge. Analysis of 32 primary studies published until August 2017 on meaning-focused word-learning spoken language input activities found a large effect of spoken input on vocabulary learning. Furthermore, adults were found to learn more vocabulary than children and tasks that required interaction were more effective than noninteractive tasks (i.e., listening to spoken language audio or watching and listening to audiovisual spoken language). Lastly, participants showed more receptive vocabulary knowledge than productive vocabulary knowledge gains. This meta-analysis provides clear evidence that word learning from meaning-focused spoken L2 input is effective. However, as with other types of activities and language input reported on above, several moderating factors affect the effectiveness of the spoken language input.

2.5 Video Viewing

The use of videos, especially captioned or subtitled video viewing, has been shown to be an effective means of vocabulary learning. This positive effect has mostly been attributed to the simultaneous aural and written modes presented to learners when watching the videos (Paivio, 1971). In a meta-analysis of 10 studies, Montero Perez et al. (2013) found a large effect from viewing captioned videos on vocabulary learning. They explained that captions increased learners' attention and encouraged focus on form, thereby leading to robust vocabulary learning. Furthermore, the participants' L2 language proficiency did not moderate this effect. While a large effect of captioned video viewing was shown for beginner, intermediate, and advanced learners, only one study examined advanced learners (Montero Perez et al., 2013), suggesting that beginners probably benefited the most from viewing videos with captions as they were still not proficient at word decoding. Given the recent interest in multimodal input (Pellicer-Sánchez, 2022) and the increasing number of studies published concerning the effects of viewing captioned and subtitled videos on vocabulary learning (Teng, 2020), it is surprising that more systematic reviews or meta-analyses were not retrieved by the database searches. It is likely that such studies have been reviewed as a part of more encompassing reviews, such as those conducted on multimedia input for more general second or foreign language learning (e.g., Zhang & Zou, 2021).

2.6 Nonformal Learning with Flashcard Software

While flashcards and word cards are most often associated with intentional learning, Nakata's (2011) synthesis on computer-assisted second language vocabulary learning with flashcard software is reviewed as these computer programs are often used for nonformal vocabulary learning. Nonformal learning may or may not be structured on a language-learning platform but is considered vocabulary learning in the wild as "how they are used remains almost completely within the control of the learner" (Dressman, 2020, p. 4). While some researchers have dismissed the decontextualized learning associated with flashcard use, the research is quite clear that flashcards and word cards, if used properly, are a robust method of fast form-meaning mapping (e.g., Reynolds & Shih, 2019; Reynolds et al., 2020)—the initial step necessary for learning new word knowledge (Nakata, 2011).

An ideal flashcard software should provide a number of creation, editing, and learning features. Nakata (2011) developed a quality flashcard application checklist based on the reviewed literature including the following affordances: flashcard creation, multilingual support, multi-word unit support, varied information support, support for data entry, flashcard sets, presentation and retrieval modes, different types of retrieval practice (i.e., receptive recall, receptive recognition, productive recall, and productive recognition), increasing retrieval effort, generative use, block size, adaptive sequencing, and expanded rehearsal. Nakata (2011) analyzed nine programs based on these 17 criteria, finding several programs met some but not all of these criteria; only one program was found that met most of the criteria. This result indicated that there is a lack of communication between researchers and flashcard application designers. Furthermore, there seems to not be any standard of what constitutes an effective flash card program.

2.7 Technology-Mediated Learning

The use of technology and specifically playing digital games has been shown to positively affect language learning due to their learner directedness and synergistic nature (Sykes & Reinhardt, 2013). Many games also go beyond the classic requirements of language input, output, and interaction for learning to occur (Sykes & Reinhardt, 2013). Knight et al., (2019, pp. 105–106) explained four benefits of informal gaming for language learning also relevant to vocabulary learning in the wild. Games provide: (1) a low-risk simulated language practice with meaningful consequences; (2) emotionally supportive environments; (3) complexity to prime learning; and (4) global competencies through online interactions.

In their theoretical narrative review, Lorenset and Piazza (2019) discuss the affordances of digital games specifically for vocabulary learning. Language learners that engage in digital gaming perceive such games as inherently motivational and pleasurable which results in long-term engagement in contextualized language use situations. Some digital game genres may also provide opportunities for communication and interactivity with other language users thereby providing real use for the language in context. Besides, the multimodal input provided by games require deeper processing of new vocabulary that encourages the development of vocabulary knowledge depth.

In a meta-analysis of 16 studies published in five databases between 2005 and 2011, Chiu (2013) found computer-assisted language learning (CALL) to have a medium effect on vocabulary learning. The following four factors were also investigated: treatment duration, educational level of participants, game-based learning, and teacher instruction. A large effect was shown for participants that used CALL technologies to learn vocabulary for less than one month and a medium effect was shown for participants that used CALL technologies to learn vocabulary for more than one month. She claimed that after a month, the novelty may wear off and learners may lose interest. She also found a difference in the effects for older and younger learners, with a large effect shown for older participants and a small effect for younger learners. The explanation given was questionable, claiming different learning styles or the difficulty of vocabulary assessed led to these differences. Instead, I tend to think agency was likely a moderator that affected the results. If learners were given the freedom to play the games they were interested in playing, then a mismatch between learner population and game was unlikely. It is also unlikely that the researchers of the reviewed primary studies would have given the participants games to play that did not match their age or proficiency levels.

In Chiu's (2013) meta-analysis non-game-based CALL was compared to game-based CALL. Game-based CALL had a small effect on vocabulary learning while non-game-based CALL had a large effect on vocabulary learning. She claimed that the majority of the games used in the reviewed studies were drill-and-practice games, and this type of game might be less efficient as it limits what learners can do in the games. Lastly, learning without the aid of a teacher was shown to have a large effect on vocabulary learning through CALL while learning with the aid of a teacher only had a small effect on vocabulary learning outcomes. This result points towards the autonomy that digital learning can provide and how less teacher involvement may encourage more exploration and vocabulary growth.

Building on the findings of Chiu (2013), Elgort (2017) conducted a systematic review of 82 primary studies published between 2010 and 2017 on technology-mediated vocabulary development. While the studies were found to be methodologically strong, few investigated fluency developments in vocabulary use, few adequately defined participants' proficiency levels, and far too many targeted comparisons between the presence or absence of technology. From a practical point of view, it is not surprising that comparing learning effects with and without the incorporation of technology results in a difference. Chiu's (2013) meta-analysis also supported this finding with the larger effect found for learners that were engaged in shorter periods of CALL. Likewise, discussing the different affordances provided by technology is much more informative to language educators than simply comparing technology-based instruction to traditional instruction (Tsai & Tsai, 2018).

Another interesting finding from Elgort (2017) was that most of the reviewed studies used researcher-designed assessments targeting the learning of specific words instead of measuring an overall construct of vocabulary knowledge using a standardized vocabulary test. Without longitudinal data, an instrument such as the Vocabulary Size Test that measures the breath of learners' L2 lexicons is less likely to be sensitive enough to show changes. This could be one reason that researchers opted to measured vocabulary learning gains with researcher-designed assessments targeting specific words. Lastly, a paucity of independent learning studies was found. Elgort (2017, p. 18) called for "researchers…to start collecting data about vocabulary development that takes place outside of the language classroom, using longitudinal and observation approaches that…are able to cover a wider range of learning strands and vocabulary knowledge aspects."

In response to Chiu's (2013) meta-analysis, Tsai and Tsai (2018) conducted a meta-analysis of digital gaming effects on second language vocabulary learning. A total of 26 studies published between 2001 and 2017 were categorized into four conditions based on their research designs: (1) experimental groups played games while control groups did something else; (2) experimental groups played games with a feature added or controlled while control groups played games without changes; (3) experimental groups played games while control groups received traditional instruction with additional content added to the game; and (4) all participants played games but compared by a non-game variable. Due to the constraints of meta-analysis, only experimental and quasi-experimental studies were included.

Unlike Chiu (2013), Tsai and Tsai (2018) found gaming superior to vocabulary instruction. They claimed the difference in their result and Chiu (2013) was due to the experience of the players and their learning goals, possibly hinting at freedom of the gamers to play the games they wanted to play. There are three takeaways from the analyses of the moderator variables: (1) university students with more advanced L2 proficiencies should play task-oriented digital games; (2) engagement in an activity in conjunction with gaming is more effective than gaming alone; and (3) custom-designed, web, or off-the-shelf games are equally suitable for vocabulary learning.

Most recently, a systematic review on the theoretical trends of 80 CALL and mobile-assisted language learning (MALL) research articles involving PreK-12 learners published between 2011 and 2020 found many of the reviewed primary studies did not explicitly use a theoretical framework (Yang et al., 2021). Largely, the field of vocabulary research is rather undertheorized (Meara, 1997) and appears to be a common thread in other areas of applied linguistics such as language learning strategies (Thomas et al., 2021). Still, Yang et al. (2021) reported that previous reviews also found a similar occurrence, suggesting that the use of theory to rationalize particular research interests has continued to be ignored. Among those studies that did use a theoretical framework, information processing/cognitive theories were used the most. The studies mostly measured knowledge of vocabulary form and meaning while paying less attention to vocabulary use. However, both productive and receptive vocabulary knowledge was assessed in these studies.

2.8 Gaps in the Vocabulary Learning in the Wild Literature

While there have been a number of research syntheses conducted with a narrow focus on vocabulary learning, a more inclusive review of vocabulary learning in the wild has not been conducted. The current scoping review addresses this issue by attempting to solidify newer emerging fields such as online informal learning of English (Toffoli & Sockett, 2015), informal digital learning of English (Lee & Dressman, 2018), extramural English (Sundqvist & Sylvén, 2016), nonformal learning (Benson, 2011), and incidental learning (Cervatiuc, 2018), among others. All of these fields focus on vocabulary encountered and learned in the wild.[3] Using the term vocabulary learning in the wild aims to underscore the similarity in all above-mentioned niche fields of vocabulary research by placing the focus on learner agency. A more inclusive view of vocabulary learning in the wild represents an ecologically valid view of the type of blurring between contexts that occurs when learning L2 vocabulary.

3 Methods

A scoping review is broad in coverage of a body of literature, does not address discrete research questions, and does not require included studies to be assessed on quality (Arksey & O'Malley, 2005). Scoping reviews "aim to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available" (Mays et al., 2001, p. 194). Furthermore, unlike meta-analyses that are limited to including research that provide particular types of statistical data for synthesists to calculate effect sizes (Cooper, 2015), a scoping review may include studies that collect different types of data. Thus, the aim of the current scoping review was to

---

[3] Interested readers are encouraged to read the cited texts for in-depth discussion of these subfields of informal language learning.

systematically examine the extent of vocabulary learning in the wild research activity and identify research gaps in the existing literature.

Conducting a scoping review or study requires completion of six stages (Levac et al., 2010): (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consultation (optional). All mandatory steps were completed for the present scoping review and are provided in detail in the following subsections.

3.1 Database Searches

"The decision on where to search for literature is contingent on a number of factors, namely time frame, language, types of publication" (Chong & Plonsky, 2021, p. 1028). The SSCI and ESCI databases in the Web of Science Core Collection under the categories Education & Educational Research, Linguistics, and Language & Linguistics were searched by title between January 2010 and November 2021. A title search under the stated categories was conducted to reduce the likelihood of irrelevant results. Moreover, the three categories are the most likely to include language learning research and can reduce the initial number of studies that must be subjected to manual filtering. Only journal articles published within the past 12 years were retrieved as literature published before this date was covered in the systematic reviews and meta-analyses retrieved and discussed in Sect. 2. Furthermore, this coverage was adequate for the present scoping review that can serve as a precursor to a systematic review (Munn et al., 2018). In other words, retrieving more recent literature allowed for extraction of data from publications in the emerging subfields of vocabulary learning in the wild. Only journal articles from the SSCI and ESCI databases were extracted as to ensure the quality of the research; this was important as no quality checks were to be performed on the retrieved research, as is the norm with a scoping review.

3.2 Search Terms

A three-part search string was constructed consisting of terms related to learning activities, second/foreign language, and vocabulary connected with the Boolean operator AND. Published reviews (e.g., Chiu, 2013; Elgort, 2017; Mohsen & Balakumar, 2011; Montero Perez et al., 2013; Ramezanali et al., 2020; Restrepo Ramos, 2015; Tsai & Tsai, 2018; Yanagisawa & Webb, 2021) and handbooks (e.g., Chapelle & Sauro, 2017; Dressman & Sadler, 2020) were consulted to generate the search string. The search string was refined by adding or deleting key terms that appeared to result in numerous irrelevant hits while also trying to be as comprehensive as possible. To ensure the reliability of the database searches, two research assistants independently performed the searches with a 100% alignment in the search results.

The first part of the search string contained a truncation of key terms related to informal, incidental, and extramural learning or language input mediums/sources (multimodal* OR multimedia OR multi-media OR informal OR virtual OR gam* OR self-paced OR fanfiction OR fan-fiction OR vlog* OR video* OR mobil* OR smartphone* OR song* OR rhyme* OR "linguistic landscapes" OR tourism OR digital OR context* OR "social network*" OR translanguaging OR technolog* OR translat* OR distance OR blended OR telecollaboration* OR "study abroad" OR author* OR "CALL" OR "CALT" OR "CALI" OR "CAI" OR computer-aided OR computer-assisted OR outside OR "out of class" OR "incidental" OR subtitle* OR caption* OR YouTube OR Ted-Ed OR TikTok OR WeChat OR Twitter OR Weibo OR WhatsApp OR Instagram OR Facebook OR Telegram OR Tencent OR read* OR wild OR extramural OR extracurricular OR "e learning" OR "internet" or "TELL" OR audiovisual).

The second part of the search string contained terms related to language learning (L2 OR SL OR "second language" OR FL OR "foreign language")**.**

The third part of the search string contained items related to vocabulary (word* OR vocab* OR collocation* OR "n gram" OR idiom* OR lex* OR chunk* OR phras* OR pattern* OR formulaic* OR figurative* OR fixed-frame* OR binomial*).

The search resulted in 257 potential articles (see Fig. 1).

3.3 Inclusion and Exclusion Criteria

The inclusion criteria were that the article was about (1) L2 vocabulary learning; and (2) informal, incidental, nonformal, or extramural engagement in an activity. Seven exclusion criteria were applied to the 257 potential articles. Fifty-one journal articles remained after excluding (1) duplicate publications ($n = 1$), (2) non-English publications ($n = 1$), (3) articles about non-mainstream learners ($n = 1$), (4) articles unrelated to informal, nonformal, incidental, or extramural L2 vocabulary learning ($n = 183$), (5) secondary research or theoretical articles ($n = 11$), (6) articles about engagement in nonformal learning activities that were not participant initiated ($n = 1$), and (7) articles that did not assess the learning of vocabulary knowledge ($n = 8$). Exclusion criteria one through four were applied by reading the title and abstract while exclusion criteria five through seven was applied after reading the full text of the articles (see Fig. 1). The articles were published between 2010 and 2021 with most articles published in 2019 (see Fig. 2).

3.4 Coding Publications

Articles were coded with the software NVivo 12 Plus by two research assistants with extensive experience in systematic reviews using a codebook that included publication characteristics (i.e., author, year of publication, study location), participant

Fig. 1 Search results and exclusion process

Fig. 2 Publication timeline

characteristics (i.e., educational level, L1, L2), activity characteristics (i.e., activity type, activity duration), methodology characteristics (i.e., methodology, theoretical framework, vocabulary outcome measures), primary study aims, and primary study findings (i.e., important results). The publication characteristics, participant characteristics, and the methodology characteristics were extracted for coding as they are the types of information that should be included in all syntheses regardless of the research questions being asked (Cooper, 2015).

As the present review aimed to provide an overview of the extant literature published on vocabulary learning in the wild, the following data were also extracted from the primary studies: vocabulary outcome measures, activity characteristics, primary study aims, and primary study findings. The codebook clearly defined each code (see Sects. 3.4.1 to 3.4.5 for summaries of these definitions) and was piloted by the research assistants who independently coded six articles (about 10% of the 51 articles). The interrater agreement for the pilot coding of the six articles was 99.23%. The incongruent coding was discussed among the two coders and the author to clarify misunderstandings. The incongruent coding and discussion centered entirely on vocabulary outcome measures as the terminology used by researchers to describe different aspects of vocabulary knowledge in published research has not always been consistent (see Schmitt [2010, ch. 5] for a discussion). After this, the two research assistants further completed two independent rounds of coding for all 51 articles. The interrater agreement for the first round of coding was 95.38%. After discussion of the incongruent coding, a second round of independent coding was conducted one week later with an interrater agreement of 99.62%. Those remaining incongruent codes were discussed among the two coders and the author to reach agreement.

3.4.1 Coding Publication Characteristics

The coded publication characteristics included author(s), year of publication, and study location (i.e., country or region).

3.4.2 Coding Participant Characteristics

The coded participant characteristics included educational level, L1, and L2. The educational level referred to whether participants were studying at the pre-primary, primary, secondary, tertiary, or mixed levels of education. The L1 referred to the first language or native language of the participants and L2 referred to the second, foreign, or additional language of the participants. Moreover, this L2 referred to the language that was the target for learning reported in the article. Bilinguals were also taken into consideration.

3.4.3 Coding Activity Characteristics

The coded activity characteristics included activity type and activity duration. Activity type referred to the activity that the participants engaged in that resulted in their exposure to vocabulary in the wild. The duration was the amount of time the participants spent engaged in the activity.

3.4.4 Coding Methodology Characteristics

The coded methodology characteristics included methodology, theoretical framework, and vocabulary outcome measures. The methodology was coded as either quantitative, qualitative, or mixed method. The theoretical framework was coded by the theory that was directly stated by the authors as being used to frame or guide the study. Not using a theoretical framework was also taken into consideration.

Vocabulary measurements were coded as either researcher-designed or standardized. A standardized measurement refers to a vocabulary assessment that has been previously analyzed and accepted as both a valid and reliable measurement of vocabulary knowledge through criterion validity, content validity, and reliability measures (Nation, 2013). Such measurements of vocabulary knowledge have often been used as global tests of vocabulary knowledge aimed at measuring a specific vocabulary construct (e.g., L2 vocabulary size). By contrast, a researcher-designed measurement is usually created to assess specific vocabulary words or items for specific research or classroom purposes (Nation, 2013). While it is possible for researchers to have reported on a researcher-designed measurement's validity and reliability, such measurements are not often designed with large participant populations in mind.

As vocabulary knowledge is considered a multidimensional construct, it was important for both researcher-designed and standardized vocabulary measurements to be further coded based on which aspects of vocabulary knowledge they assessed. The most widely adopted taxonomy of vocabulary knowledge was put forth by Nation (2022). He categorized vocabulary knowledge into three main aspects: *form*, *meaning*, and *use*. Form refers to spoken form, written form, and word parts. Meaning refers to the connection between form and meaning, concept and referents, and associations. Use refers to grammatical functions, collocations, and constraints on use.

Each of these types of knowledge can also be distinguished in terms of receptive and productive knowledge.

The learning of these different aspects of vocabulary knowledge does not occur in a prescribed sequence as it can be affected by individual differences of the learner or the context in which the learning takes place (Nation, 1990, 2013, 2022). Some vocabulary measurements were designed to simultaneously measure more than one aspect of vocabulary knowledge. In such cases, the vocabulary measurements were coded for all aspects measured. Thus, six different codes were used to code the aspects of vocabulary knowledge measured in the articles: receptive knowledge of form, receptive knowledge of meaning, receptive knowledge of use, productive knowledge of form, productive knowledge of meaning, and productive knowledge of use.

3.4.5 Coding Primary Study Aims and Primary Study Findings

The primary study aims and primary study findings included the aim of the primary study and the primary study's important results. The aim of the primary study was coded for each article by locating the sentence(s) that stated the aim of the study using words such as goal(s), objective(s), aim(s), aimed, purpose, investigate(s), investigated, investigation, research gap, examine(s), examined, assess, test, addressed, and designed.

The findings were coded for each article by reading the abstract, results, and conclusion sections of the articles and extracting up to three of the most important results.[4] Important results were selected most often by looking for the answer to the research question(s) or the confirmation of proposed hypotheses.

The primary study aims and primary study findings were synthesized using qualitative coding techniques. For the primary study aims and then again for the primary study findings, the author first used open attribute inductive coding to break the data into parts and then used pattern coding to draw connections between the codes used during open attribute inductive coding to construct themes (Miles et al., 2014, pp. 79, 86–89). This coding process resulted in the themes that were used to organize the discussion of the primary study aims and primary study findings of the reviewed articles. These themes are reported in Sect. 4.5 and 4.6.

[4] It should be noted that the study aims and important results were not reinterpreted by the coders or the author. Instead, the study aims and important result(s) as reported by the primary study authors were extracted verbatim for synthesizing.

4 Results

A systematic map of the coded 51 articles is provided in Appendix. These coding results are discussed below. Specifically, publication characteristics, participant characteristics, activity characteristics, methodology characteristics, synthesized primary study aims, and synthesized primary study findings are reported.

4.1 Publication Characteristics

4.1.1 Who Conducted the Studies? When and Where Were They Conducted?

As reported in Sect. 3.3 and shown in Fig. 1, more articles were published in 2019 ($k = 13$) than other years with the majority of articles having been published between 2018 and 2021 ($k = 30$). Most research was conducted in Taiwan ($k = 6$) and the US ($k = 6$) followed by Mainland China ($k = 5$) (see Table 1). The 51 articles were (co-) authored by 103 persons of which 98 contributed to the publication of only 1 article. Only five authors contributed to more than one publication: Barry Lee Reynolds ($k = 4$), Elke Peters ($k = 3$), Pia Sundqvist ($k = 2$), Gülcan Erçetin ($k = 2$), and Sarvenaz Hatami ($k = 2$). This indicates the reviewed literature represented a variety of author voices.

4.2 Participant Characteristics

4.2.1 What Types of Participants (i.e., Educational Level, L1, L2) Were Recruited?

Most studies ($k = 30$) recruited tertiary-level participants followed by participants studying at the level of secondary school ($k = 10$) (see Table 2). Most participants spoke Mandarin as their L1 ($k = 11$), followed by mixed L1 backgrounds ($k = 8$) and English ($k = 5$) (see Table 3). Only one article reported recruitment of bilinguals. The most frequently targeted L2 investigated in the articles was overwhelmingly English ($k = 42$) followed by German ($k = 2$) and Mandarin ($k = 2$) (Table 4).

Table 1 Study location

Location	*k*	%
Belgium	4	7.84
Canada	1	1.96
Mainland China	5	9.80
Finland	1	1.96
Germany	2	3.92
Hong Kong	1	1.96
Iran	4	7.84
Ireland	1	1.96
Japan	2	3.92
Netherlands	1	1.96
Saudi Arabia	2	3.92
Singapore	1	1.96
South Korea	2	3.92
Spain	3	5.88
Sweden	2	3.92
Taiwan	6	11.64
Turkey	4	7.84
UK	3	5.88
US	6	11.64
Total	51	100

Table 2 Education level

Education level	*k*	%
Pre-primary	2	3.92
Primary	2	3.92
Secondary	10	19.61
Tertiary	30	58.82
Mixed	7	13.72
Total	51	100

4.3 Activity Characteristics

4.3.1 Which Activities Were Investigated? What Were the Durations of the Activities?

Research participants engaged in 24 different activities (see Table 5); however, some activities were further classified under three categories: listening, reading, and watching. The articles reported participants mostly engaged in reading activities ($k = 18$), followed by watching ($k = 9$), digital gameplay ($k = 9$), and listening ($k =$

Table 3 L1

L1	k	%
Arabic	2	3.92
Bilingual*	1	1.96
Dutch	4	7.84
English	5	9.80
Farsi	3	5.88
Finnish	1	1.96
French	1	1.96
German	1	1.96
Japanese	2	3.92
Korean	2	3.92
Mandarin	11	21.57
Mixed	8	15.69
Persian	1	1.96
Spanish	3	5.88
Swedish	2	3.92
Turkish	4	7.84
Total	51	100

Note *Catalan and Spanish ($k = 1$)

Table 4 L2

L2	k	%
Arabic	1	1.96
Dutch	1	1.96
English	42	82.35
French	1	1.96
German	2	3.92
Mandarin	2	3.92
Spanish	1	1.96
Welsh	1	1.96
Total	51	100

5). Two studies combined reading with writing or listening. Participants read texts, stories, and novels; listened to passages, songs, and stories; and watched educational programs, video clips, movies and news programs with subtitles, and open courseware lectures. Nonformal learning through engagement in informal digital learning of English ($k = 3$), using mobile learning app(s) ($k = 3$), and using word cards ($k = 2$) were also reported. Participants also told stories, studied abroad, and were immersed in virtual reality, among others. Most studies ($k = 13$) reported participants having engaged in an activity that lasted less than an hour or between 1 and 3 h ($k = 8$) (see

Table 5 Activity type classification

Activity type	*k*	%
Digital gameplay	7	13.72
Engagement in informal digital learning of English	3	5.88
Extramural Language Engagement	1	1.96
Immersion in virtual reality	1	1.96
Listening		
Listening to passages	1	1.96
Listening to songs	2	3.92
Listening to songs and stories	1	1.96
Listening to stories	1	1.96
Reading		
Reading a text(s)	8	15.69
Reading and writing sentences	1	1.96
Self-choice reading	2	3.92
Reading a novel(s)	5	9.80
Reading or listening to a novel(s)	1	1.96
Reading stories	1	1.96
Studying abroad	1	1.96
Telling stories	1	1.96
Using mobile learning app(s)	3	5.88
Using word cards	2	3.92
Watching		
Watching educational programs	1	1.96
Watching movies and news programs with subtitles	1	1.96
Watching open courseware lectures	1	1.96
Watching TV programs	5	9.80
Watching video clips	1	1.96
Total	51	100

Table 6); however, there was also a sizeable number ($k = 7$) that reported participants engaging in an activity longer than 30 days but less than a year. Due to the design of certain studies, the time engaged in activities was participant determined ($k = 4$) or not reported ($k = 3$).

Table 6 Activity duration

Duration	k	%
< 1 h	13	25.49
≥ 1 h & $= 3$ h	8	15.69
> 3 h & $= 6$ h	1	1.96
> 6 h & $= 24$ h	2	3.92
> 24 h & $= 7$ days	4	12.90
> 7 days & $= 30$ days	5	9.80
> 30 days & $= 1$ year	7	13.73
> 1 year	4	12.90
Participant determined	4	12.90
N.A.	3	5.88
Total	51	100

Note For the durations that spanned two time blocks, the durations were grouped under the longer time block; N.A. = Not available

Table 7 Methodologies

Methodologies	k	%
Mixed-method	7	13.72
Qualitative	2	3.92
Quantitative	42	82.35
Total	51	100

4.4 Methodology Characteristics

4.4.1 What Methodologies and Theoretical Frameworks Were Used?

The majority of the reviewed articles were quantitative ($k = 42$) followed by mixed-method ($k = 7$) (see Table 7). Only two articles reported qualitative studies. Fifteen theoretical frameworks were used in the research reported in the articles, but most studies ($k = 31$) did not use a theoretical framework (see Table 8). Ten of the theoretical frameworks were applied in only one study. Other articles reported using Dual-coding Theory (Paivio, 1971) ($k = 5$), Levels of Processing Theory (Craik & Lockhart, 1972) ($k = 4$), the Generative Theory of Multimedia Learning (Mayer, 2001) ($k = 3$), and the Noticing Hypothesis (Schmidt, 1994) ($k = 2$).

4.4.2 What Type(s) of Vocabulary Knowledge Were Measured?

Most studies reported using researcher-designed vocabulary measurements ($k =$ 42) in lieu of standardized measurements ($k = 21$) (see Table 9); however, many studies used both types of measurements. Most standardized measurements ($k =$

Table 8 Theoretical frameworks

Theoretical framework	*k*	%
Additivity Hypothesis (Paivio, 1975)	1	1.79
Cognitive Load Theory (Sweller, 1994)	1	1.79
Complementary Learning Systems (CLS) Theories of Memory (McClelland et al., 1995)	1	1.79
Declarative/Procedural Model (Ullman, 2005)	1	1.79
Dual-coding Theory (Paivio, 1971)	5	8.93
Dual-processing Theory of Working Memory (Baddeley, 2007)	1	1.79
Involvement Load Hypothesis (Laufer & Hulstijn, 2001)	1	1.79
Levels of Processing Theory (Craik & Lockhart, 1972)	4	7.14
Monitor Theory for L2 Acquisition (Krashen, 1987)	1	1.79
Noticing Hypothesis (Schmidt, 1994)	2	3.57
Psycholinguistic Theories of Human Lexical Organization and Memory (Beckwith et al., 1991)	1	1.79
Skill Acquisition Theory (Anderson, 1982)	1	1.79
Grounded Cognition (Barsalou, 2008)	1	1.79
The Generative Theory of Multimedia Learning (Mayer, 2001)	3	5.36
Theory of Cognitive Dissonance (Festinger, 1957)	1	1.79
None	31	55.36

Note Some articles (Albaladejo et al., 2018; Barwasser et al., 2021; Lee & Pulido, 2017; Peters, 2019; Sok & Han, 2020) used more than one theoretical framework, resulting in a total higher than 51 ($n = 56$); % represent the proportion of studies out of 51 that applied the theory

17) and researcher-designed measurements ($k = 25$) assessed receptive knowledge of meaning (see Tables 10 and 11); however, productive knowledge of meaning ($k = 23$) was assessed by a large number of the researcher-designed measurements as well. None of the standardized measurements and only five of the researcher-designed vocabulary measurements assessed productive knowledge of use. Five studies used more than one standardized measurement and two studies used standardized measurements that assessed more than one aspect of vocabulary knowledge. Nineteen studies used researcher-designed measurements that measured more than one aspect of vocabulary knowledge. Overall, more studies ($k = 41$) measured receptive knowledge of meaning than any other aspect of vocabulary knowledge (see Table 12). In addition, more studies measured productive knowledge than receptive knowledge and meaning was measured in more studies than form or use (see Table 12).

Table 9 Vocabulary measurements

Vocabulary measurements	*k*	%
Researcher-designed measurements	42	82.35
Standardized measurements	21	41.18

Note Studies used both researcher-designed and standardized measurements, resulting in a total higher than 51; % represent the proportion of studies out of 51 that used researcher-designed and standardized measurements

Table 10 Vocabulary knowledge aspects measured by standardized measurements

Vocabulary knowledge aspect	*k*	%
Receptive knowledge of form (RF)	5	13.89
Receptive knowledge of meaning (RM)	17	47.22
Receptive knowledge of use (RU)	0	0
Productive knowledge of form (PF)	4	11.11
Productive knowledge of meaning (PM)	5	13.89
Productive knowledge of use (PU)	5	13.89

Note Five studies (Barwasser et al., 2021; Lee, 2019; Masrai & Milton, 2018; Sundqvist, 2019; Sundqvist & Wikström, 2015) used more than one standardized measurement and two studies (Lee, 2019; Sundqvist & Wikström, 2015) used standardized measurements that measured two aspects of vocabulary knowledge

Table 11 Vocabulary knowledge aspects measured by researcher-designed measurements

Vocabulary knowledge aspect	*k*	%
Receptive knowledge of form (RF)	12	13.48
Receptive knowledge of meaning (RM)	25	20.09
Receptive knowledge of use (RU)	14	15.73
Productive knowledge of form (PF)	10	11.24
Productive knowledge of meaning (PM)	23	25.84
Productive knowledge of use (PU)	5	5.62

Note Nineteen studies (Baills et al., 2021; Çakmak & Erçetin, 2017; Enayat & Haghighatpasand, 2019; Hatami, 2017, 2018; Lee & Pulido, 2017; Mohamed, 2018; Murphy et al., 2021; Ouyang et al., 2020; Pellicer-Sánchez, 2016; Puimège & Peters, 2019; Pujadas & Muñoz, 2019; Reynolds, 2014, 2015; Teng, 2018; Türk & Erçetin, 2012; van Zeeland & Schmitt, 2013; Wong et al., 2021; Zhao & Ren, 2019) used more than one researcher-designed measurement and eleven studies (Bahari, 2019; Borràs & Llanes, 2020; Çekiç & Bakla, 2019; deHaan et al., 2010; Enayat & Haghighatpasand, 2019; Hu et al., 2014; Peters, 2019; Reynolds, 2020; Teng, 2018; Urun et al., 2017; Wu, 2015) used researcher-designed measurements that measured more than one aspect of vocabulary knowledge

Table 12 Vocabulary knowledge aspects measured by standardized and researcher-designed measurements

Vocabulary knowledge types	*k*	%
Receptive knowledge of form (RF)	22	17.46
Receptive knowledge of meaning (RM)	41	32.54
Receptive knowledge of use (RU)	14	11.11
Productive knowledge of form (PF)	13	10.32
Productive knowledge of meaning (PM)	26	20.63
Productive knowledge of use (PU)	10	7.94

Note % represent the proportion of studies out of 51 that measured a particular aspect of vocabulary knowledge with a researcher-designed or standardized measurement

4.5 Synthesized Primary Study Aims

4.5.1 What Were the Aims of the Studies?

The aims of the studies reported in the articles were categorized under seven themes that were not mutually exclusive (see Table 13). These seven themes were (1) incidental vocabulary acquisition through engagement or the completion of an activity; (2) vocabulary acquisition through engagement or the completion of an activity; (3) bridging incidental and intentional vocabulary learning; (4) learner factor effects; (5) language input factor effects; (6) effects on variables other than vocabulary knowledge; and (7) technology and vocabulary learning.

Many of the reviewed studies aimed to investigate the incidental acquisition ($k = 28$) or the acquisition ($k = 23$) of vocabulary through engagement or the completion of an activity. These activities were reported in Sect. 4.3 and summarized in Table 5. There were also a few studies ($k = 5$) that aimed to bridge formal/informal or intentional/incidental learning of vocabulary. In addition to a focus on vocabulary acquisition, a number of the studies also aimed to investigate learner differences ($k = 10$) and language input effects ($k = 17$) on or in conjunction with vocabulary acquisition. Some articles ($k = 5$) also reported on the effects engagement in activities had on variables other than vocabulary knowledge. Three studies also investigated the effects of technology use on vocabulary learning.[5]

[5] These were studies with a main focus on technology for vocabulary learning; the use of technology was more than just a medium for the delivery of language input.

Table 13 Results of thematic coding of the primary study aims

Theme	Code(s)	Article(s)
Incidental vocabulary acquisition through engagement or the completion of an activity		Bahari (2019), Barwasser et al. (2021), Bisson et al. (2014), Bordag et al. (2015), Çakmak and Erçetin (2017), Çekiç and Bakla (2019), de Vos et al. (2019), Godfroid et al. (2013), Hatami (2017, 2018), Hu et al. (2014), Lee and Pulido (2017), Mohamed (2018), Murphy et al. (2021), Ouyang et al. (2020), Pellicer-Sánchez (2016), Peters and Webb (2018), Pichette et al. (2012), Puimège and Peters (2019), Reynolds (2014, 2015, 2020), Reynolds and Bai (2013), Sok and Han (2020), Teng (2018), Türk and Erçetin (2012), van Zeeland and Schmitt (2013), and Wong et al. (2021)
Vocabulary acquisition through engagement or the completion of an activity		Albaladejo et al. (2018), Alfadil (2020), Baills et al. (2021), Borràs and Llanes (2020), Calvo-Ferrer (2020), deHaan et al. (2010), Enayat and Haghighatpasand (2019), Lee (2019), Legault et al. (2019), Masrai and Milton (2018), Moufarrej and Salameh (2019), Niitemaa (2020), Peters (2019), Pujadas and Muñoz (2019), Sato et al. (2013), Sundqvist (2019), Sundqvist and Wikström (2015), Urun et al. (2017), Wang (2019), Wong et al. (2016), Wu (2015), Yang and Sun (2011), and Zhao and Ren (2019)
Bridging incidental and intentional vocabulary learning		Bahari (2019), Sok and Han (2020), Sundqvist and Wikström (2015), Wong et al. (2021), and Wong et al. (2016)
Learner factor effects	Prior vocabulary knowledge	Borràs and Llanes (2020), Peters and Webb (2018), and Puimège and Peters (2019)
	Target language proficiency	Borràs and Llanes (2020), Pujadas and Muñoz (2019), and Zhao and Ren (2019)
	Learning styles	Hatami (2018)

(continued)

Table 13 (continued)

Theme	Code(s)	Article(s)
	Reading behaviors	Ouyang et al. (2020)
	Memory	Murphy et al. (2021)
	Interest	Lee and Pulido (2017) and Reynolds and Bai (2013)
	Attention	Godfroid et al. (2013)
Language input factor effects	Concreteness	Puimège and Peters (2019)
	Word length	Puimège and Peters (2019)
	Part of speech	Puimège and Peters (2019)
	Collocate-node relationship	Puimège and Peters (2019)
	Association strength	Puimège and Peters (2019)
	Glosses	Hu et al. (2014), Ouyang et al. (2020), and Zhao and Ren (2019)
	Frequency	Hatami (2017), Mohamed (2018), Niitemaa (2020), Peters and Webb (2018), Puimège and Peters (2019), Reynolds (2015), and Teng (2018)
	Morphology	Reynolds (2015)
	Captions and subtitles	Peters (2019), Pujadas and Muñoz (2019), Teng (2018), and Wang (2019)
	Spacing	Çekiç and Bakla (2019)
	Cognateness	Peters and Webb (2018)
	Imagery	Peters (2019)
	Word relevance	Peters and Webb (2018)
	Text complexity	Bordag et al. (2015)
	Instruction	Puimège and Peters (2019)
Effects on variables other than vocabulary knowledge	Effects on reading comprehension	Borràs and Llanes (2020), Türk and Erçetin (2012), and Wang (2019)
	Effects on listening comprehension	Çakmak and Erçetin (2017)
	Effects on pronunciation	Baills et al. (2021)
Technology and vocabulary learning		Calvo-Ferrer (2020), Legault et al. (2019), and Wu (2015)

4.6 *Synthesized Primary Study Findings*

4.6.1 What Were the Important Findings of the Studies?

The findings reported in the articles were categorized under six themes that were not mutually exclusive. These six themes were (1) frequency and processing; (2) learning vocabulary from listening, reading, and writing; (3) multimodal input and gaming; (4) learner autonomy; (5) learner differences; and (6) bridging incidental and intentional learning of vocabulary.

4.6.2 Frequency and Processing

The more frequent learners encounter unknown words, the more likely they are to acquire knowledge of these words (Peters & Webb, 2018; Teng, 2018) and the less time they will spend fixated on the words (Mohamed, 2018). This is because learners begin to read unknown words encountered through text similarly to known words after eight exposures (Pellicer-Sánchez, 2016); however, as little as three exposures might be needed for vocabulary learning (Bordag et al., 2015), especially if learners are highly motivated and interested in the texts they read (Reynolds, 2014). Similarly, as few as three exposures could result in incidental learning of words encountered in multimodal input (Bisson et al., 2014; Teng, 2018) because presentation of verbal and visual information reduces cognitive load and results in better learning outcomes (Türk & Erçetin, 2012).

The research reported in the reviewed articles tended to indicate that quality of encounters with unknown words might be a better predictor of vocabulary acquisition than quantity of encounters with unknown words. One study found frequency of exposure was shown to be highly correlated with vocabulary learning outcomes (Hatami, 2017) while another study found frequency of exposure to only have a small effect on vocabulary learning outcomes (Yang & Sun, 2011). This indicates *raw* frequency may not be the most robust predictor of vocabulary learning outcomes. For example, if assessed targeted words vary derivationally in their forms, the frequency of encounters from reading may need to be increased for learners to be able to associate these different yet related forms and benefit from the frequency effect (Reynolds, 2015). Moreover, frequency seemed to have a stronger effect on form recognition compared with meaning recognition and meaning recall (Mohamed, 2018). Similarly, frequency was shown to influence vocabulary learning more than contextual richness (Reynolds, 2020). Data from several studies suggest that total time spent fixated on unknown words to be a better predictor of vocabulary learning than number of encounters with unknown words (Mohamed, 2018) and instead, longer reading times seem to be associated with vocabulary learning outcomes (Godfroid et al., 2013; Pellicer-Sánchez, 2016).

4.6.3 Learning Vocabulary from Listening, Reading, and Writing

Many of the reviewed studies reported how listening, reading, and writing differentially affected vocabulary learning. Specifically, writing results in more vocabulary gains than reading (Pichette et al., 2012) and reading results in more vocabulary gains than listening (Hatami, 2017); however, listening has a greater effect on retention than reading (Hatami, 2017). Concrete words are easier to acquire from both reading and writing than abstract words (Pichette et al., 2012). Frequency of exposure from reading was shown to have a larger effect on vocabulary learning compared to frequency of exposure from listening (Hatami, 2017) and up to 15 encounters may be needed for incidental acquisition of unknown words encountered from listening (van Zeeland & Schmitt, 2013).

Still, listening cannot be dismissed as a non-effective route to vocabulary learning. Well before a form-meaning link can be assessed, learners begin acquiring from listening basic aspects of vocabulary knowledge such as form recognition (van Zeeland & Schmitt, 2013). Listening to spoken song lyrics, though, was found to be less effective at encouraging vocabulary growth than listening to the lyrics sung (Moufarrej & Salameh, 2019); however, listening to songs or singing the songs seems to result in similar vocabulary learning outcomes (Baills et al., 2021). Moreover, listening to songs was not found to be as conducive for vocabulary learning compared with listening to stories (Albaladejo et al., 2018). When listening to academic lectures, learners were found to gain more technical rather than academic vocabulary knowledge (Yang & Sun, 2011).

A number of variables were shown to moderate the effectiveness of reading inducing vocabulary learning. Learners' possessing a higher proficiency in the target language tended to learn more vocabulary knowledge from reading (Hu et al., 2014; Lee & Pulido, 2017); however, vocabulary size seemed to have only a negligible effect (Reynolds, 2020). While interest did not moderate reading comprehension, it did positively affect incidental acquisition of vocabulary from articles read (Lee & Pulido, 2017; Reynolds & Bai, 2013). Glossing was further shown to enhance vocabulary acquisition from reading (Çakmak & Erçetin, 2017; Ouyang et al., 2020) but one study showed sustained effects of glossing only for vocabulary recognition (Zhao & Ren, 2019). Increasing the frequency of glossing is more helpful to lower proficiency learners as they require additional help with encoding and comprehending texts (Zhao & Ren, 2019). Spacing out the encounters with unknown words that are accompanied by glosses results in more effective vocabulary learning than when massed encounters with unknown words accompanied by glosses are provided (Çekiç & Bakla, 2019).

4.6.4 Multimodal Input and Gaming

Multimodal input reduces cognitive load and leads to better vocabulary learning outcomes (Türk & Erçetin, 2012); however, this result seems to be more robust for multimodal media consumption compared to digital game play (deHaan et al., 2010).

Still, game-based learning, self-directed computer-assisted language learning, and immersive virtual reality were all found to be conducive to vocabulary learning and often more so than traditional vocabulary instruction (Alfadil, 2020; deHaan et al., 2010; Enayat & Haghighatpasand, 2019; Legault et al., 2019; Sato et al., 2013; Sundqvist & Wikström, 2015; Urun et al., 2017; Wu, 2015).

The more time learners spend playing games, the more vocabulary they seem to learn (Sundqvist, 2019). When compared to informal exposure to language through music, film, websites, social networks, and games, the correlation between vocabulary outcomes and gaming was found to be the strongest (Niitemaa, 2020). Learners recruited in the reviewed research also seem to possess very positive views towards learning through game play (Enayat & Haghighatpasand, 2019). Specifically, learners that are comfortable with technology and respond to instant gratification are more prone to learning vocabulary through game play (Calvo-Ferrer, 2020).

In addition to game play, learners have been shown to acquire vocabulary through television viewing (Masrai & Milton, 2018; Peters & Webb, 2018; Puimège & Peters, 2019; Pujadas & Muñoz, 2019). Video viewing with captions (Peters, 2019; Teng, 2018; Wang, 2019) or pre-teaching (Pujadas & Muñoz, 2019) further increased the likelihood of incidental vocabulary learning. Recalling vocabulary form knowledge after video viewing is easier for learners than recalling vocabulary meaning knowledge (Pujadas & Muñoz, 2019). Viewing open courseware lectures also resulted in vocabulary growth (Yang & Sun, 2011).

4.6.5 Learner Autonomy

Learners that were left to their own devices were able to acquire vocabulary (Wong et al., 2016). The quality and not the quantity of the engagement in informal digital learning of English were associated with robust vocabulary learning outcomes (Lee, 2019). Niitemaa (2020) reported a group of 46 learners' informal exposure to language through music, film, websites, games, and social networking remained relatively stable over one year; half of the learners reported daily contact with the target language and the majority of the surveyed learners felt they learned the most vocabulary from self-initiated activities online and could give examples of newly learned vocabulary. Short study abroad experiences also resulted in marked increases in receptive vocabulary knowledge but not necessarily for productive vocabulary knowledge (Borràs & Llanes, 2020).

4.6.6 Learner Differences

Learners that noticed their gaps in vocabulary knowledge were shown to make greater learning gains than learners that did not notice these gaps (de Vos et al., 2019). This attention could be triggered if the contexts in which unknown words appeared were syntactically complex; this noticing or attention resulted in more incidental vocabulary learning (Bordag et al., 2015). Pre-exposure to targeted vocabulary could also increase the likelihood of learning (Bisson et al., 2014). Self-professed learning styles (Hatami, 2018) nor gender (Lee & Pulido, 2017) had a noticeable effect on vocabulary learning. Learners with higher proficiency in the target language (Pujadas & Muñoz, 2019) and better declarative memories (Murphy et al., 2021) made greater gains in vocabulary learning.

4.6.7 Bridging Incidental and Intentional Learning of Vocabulary

The learning of vocabulary knowledge through incidental or informal exposure occurs gradually. Combining intentional and incidental learning conditions could lead to even greater gains in vocabulary knowledge than either condition alone (Bahari, 2019; Barwasser et al., 2021; Sok & Han, 2020). However, some research reported intentional learning conditions led to more vocabulary learning than incidental learning conditions (Wong et al., 2021). This was most likely due to the amount of scaffolding or verbal elaboration provided by the input (Yang & Sun, 2011). As an example, Wong et al. (2016) found learners' engagement in an authentic out-of-class online social networking platform targeting second language learners fostered more robust vocabulary learning by students than formal in-class teaching. Comparing the language used by students outside the classroom on the social networking platform to that used inside the classroom, the researchers found students used significantly more lower frequency words on the social networking platform. The diversification of the vocabulary used on the social networking platform was attributed to the purposeful communication opportunities it provided.

5 Discussion and Implications

The pertinent findings are summarized and then discussed relative to their importance for future research practices. A summary of these findings and the implications are provided in Table 14. Highlighting these gaps in the literature provide avenues for researchers to follow. By addressing these gaps incrementally through a series of studies will lead to a greater depth of understanding about vocabulary learning in the wild.

5.1 English Language Dominance

Reviewing the publication and participant characteristics of the reviewed articles found several gaps that should be addressed with future research. While researchers from different regions around the world conducted these studies by recruiting participants from various L1 backgrounds, more than 80% of these articles focused on the learning of English vocabulary. The relationship between L2 orthography/phonology

Table 14 Main findings and implications for future research

Study findings	Research implications
Most of the reviewed studies focused on learning English vocabulary	Conduct research on learning of vocabulary for languages other than English
Most of the reviewed studies recruited tertiary-level participants	Conduct research that recruits early years language learners
Few of the reviewed studies involved output tasks	Conduct research requiring language production or interaction in both classroom and non-classroom contexts
More than half of the reviewed studies were not theory driven	Conduct research that aims to answer questions or address hypotheses that are theory driven
Only two of the reviewed studies used qualitative research methods	Conduct qualitative research studies that can illuminate how contextual factors affect vocabulary learning
Productive knowledge of form, receptive knowledge of use, and productive knowledge of use received less attention in the reviewed studies	Design assessment instruments that can allow measurement of productive knowledge of form and conduct longitudinal studies that would allow for the assessment of receptive and productive knowledge of use
Multimodal input and digital gaming are prominent foci of vocabulary research	Conduct more authentic multimodal reading and gaming research
Desirable difficulties in learning vocabulary received little attention by the reviewed studies	Conduct research that aims to understand why difficulty makes particular tasks more or less favorable for vocabulary learning

and L1 orthography/phonology can have positive or negative transfer effects on language learning, especially from and while reading (Birch & Fulop, 2020). Thus, there is a lot of room for future researchers to contribute to the existing vocabulary learning in the wild literature by conducting research on the learning of vocabulary for languages other than English, especially those that use non-Latin scripts. Considering the dominant role that English plays in the internationalization of education (Reynolds & Yu, 2017), it is understandable that the majority of the reviewed articles focused on the learning of English vocabulary. Another potential reason may have been due to exclusion of non-English publications and limited database searches (see Sect. 6). There is a need to continue the line of research on learning of languages other than English (Guo et al., 2021).

5.2 Marginalization of Young Learners

Most of the reviewed research on vocabulary learning in the wild was carried out by recruiting tertiary-level participants. There are obvious difficulties associated with obtaining approval from ethics review boards to gather data from young learners. Ensuring informed consent has been granted from parents and potentially getting permission from teachers and principals are other likely barriers. Still, having such a small number of studies targeting pre-primary and primary level participants is concerning.

Second language learning at the pre-primary and early primary level is mostly concerned with building up vocabulary knowledge (especially the spoken forms) that will serve as a foundation for grammar and skill learning during primary years (Cameron, 2010). Thus, if second language learning begins earlier in life, this time will constitute robust vocabulary growth (Cameron, 2010; Nagy, 1997; Nation, 2013, 2022). Although the most frequent words in the target language are likely to have been learned as a result of direct instruction (Nation, 2013), this is not the case for all vocabulary learning during early years.

Investigations that focus on these early years are sorely needed. For example, several lines of inquiry found young learners are spending an increasing amount of time using digital devices with and without the supervision of guardians (Billington, 2016; Virkus et al., 2017). Future research can reveal how use of these digital devices in the wild lead to vocabulary growth in young learners. In addition, it is likely that older learners might engage in particular activities specifically for the purpose of enhancing their language skills, while children may be motivated for other reasons. There are potentials for an abundance of by-product or incidental vocabulary learning during early years.

5.3 Lack of Theory-Driven Research

Examining the methodologies of the reviewed studies highlighted the need for theory-driven research designs. More than 55% of the reviewed articles did not overtly mention a theory that helped to guide the reported investigation. This result was highlighted in previous related syntheses (e.g., Yang et al., 2021). In general, vocabulary research is largely undertheorized and previous vocabulary research syntheses have also reported the problems associated with research lacking clear theoretical underpinning (e.g., Yang et al., 2021). This is unfortunate as a theory helps point researchers in a direction that can ensure a constant building up of our understanding of particular phenomenon. Theory can also help researchers in stating their claims for significance of a study and giving the rationale for their methods at tackling a problem (Belcher, 2019). While in recent years vocabulary research has really taken off, the attention to theory is something that should be addressed in future research. Theory should be driving the research problems and research questions and not be an afterthought.

5.4 Absence of Qualitative Investigations

Only two qualitative studies were identified in the reviewed articles, indicating an imbalance between qualitative and quantitative research designs. Presumably, researchers have opted for quantitative research designs to provide more robust generalizations to a large, targeted population. However, qualitative research designs, especially those using a case study data collection method, can provide thick description of critical incidents that may illuminate particular dimensions of language learning that might normally go unnoticed (Ortega & Iberri-Shea, 2005). As vocabulary learning in the wild is a contextualized activity, trying to understand how or why vocabulary learning occurs (or does not occur) in different contexts under different environmental constraints is well-suited to qualitative research designs (Ortega & Iberri-Shea, 2005).

All of the reviewed studies dealt with the measurement of some aspect of vocabulary knowledge. If this is one of the main aims of a study, there might be less incentive for a researcher to assess a single or a few participants' learning of vocabulary. Instead, researchers' implicit aims may have been to report vocabulary learning outcomes obtained from using reliable and valid vocabulary learning measurements whether standardized or researcher constructed. Conducting an authentic and ecologically valid experiment-driven in the wild vocabulary learning study is possible but it will be difficult for researchers to eliminate the looming presence of artificialization. For example, investigating the incidental acquisition of vocabulary from academic lectures in a natural context is difficult because a researcher cannot predetermine which words unknown to the students a lecturer might use when lecturing. The researcher could circumvent this by recording a series of lectures and then select potential target unknown words from these lectures and then confirming this by administering a pretest. Then the students could be asked to watch the recordings before administering a posttest and then a delayed posttest. However, this makes the lectures somewhat artificial and more like video viewing as the potential of interactions among the lecturer and the students is eliminated.

The suggestion to increase qualitative vocabulary learning in the wild research is not made to downplay the contributions of quantitative research but instead to encourage a balance in research designs. Doing so should provide an opportunity for a fuller picture to be drawn of vocabulary learning in the wild. Future researchers may consider research methods such as case study, ethnography, narrative research, and Q methodology, among others, that also collect and analyze data longitudinally (e.g., Huang, 2010). While a few of the studies touched upon learner induced vocabulary learning, more autonomous vocabulary learning in the wild studies are still needed. In other words, research designs that aim to observe and interpret and not intervene and measure are necessary (Elgort, 2017). Researchers should aim to understand what activities and why language learners are intrinsically motivated to engage in particular activities that leads to either intentional or incidental learning of vocabulary.

5.5 Need for Instruments Measuring Vocabulary Form and Use

Three aspects of word knowledge seem to have been given markedly less attention in the reviewed studies: productive knowledge of form, receptive knowledge of use, and productive knowledge of use.

Productive knowledge of form was assessed in only about 10% of the reviewed articles. Productive knowledge of form refers to the spelling and pronunciation of words (Nation, 2022). Considering how vocabulary teaching emphasizes the connection between form and meaning (Brown, 2010), it is surprising to see so few studies reporting assessment on assessing productive knowledge of form. This could be due to the difficulties in creating and validating relevant assessments.

To assess the written productive knowledge of form, researchers often opt to provide learners a translation and a blank with an initial letter as prompts on test sheets (e.g., Snoder & Reynolds, 2019). While possible for assessing some L2s, this is not possible with non-Latin alphabet L2s. Furthermore, this technique presupposes learners have already associated the L1 and L2 written forms. Alternatively, an aural prompt can be provided but researchers would need to ensure that the learners had mastered the receptive knowledge of form before such a route could be taken. Unlike the written form, the spoken form does not have a universal standard. Researchers would need to rationalize that the spoken forms used in their assessments were the more suitable exemplars. Likewise, assessing the spoken productive knowledge of form would require selecting a single standard and using multiple assessors to ensure reliability of the measurement. Thus, it is understandable why researchers have avoided, even if unintendedly, this thorny issue. Still, since written and spoken productive knowledge of form are aspects of vocabulary knowledge that are expected to be readily available to learners for meaning-focused use of the language, researchers are advised to prioritize assessment of productive form knowledge in future studies.

Fewer than 12% of the studies assessed receptive knowledge of use and fewer than 8% assessed productive knowledge of use. Measuring knowledge of use has the potential of being more complicated than form and meaning because it requires measuring grammatical functions, patterns, collocations, associations, constraints on use, and so on (Nation, 2022). This also indirectly indicates that researchers have placed less emphasis on how individual words are related to one another through collocation and multiword patterning. Most of the researchers that measured use did so through with the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993). As a relatively longer period of time is needed for learners to develop vocabulary knowledge of use (Nation, 2022), this may have made researchers hesitant to assess use aspects of vocabulary knowledge in studies with a shorter intervention, treatment, or data collection period. Studies with longitudinal designs that collect both spontaneous and controlled speech and writing would be useful in aiding our understanding in how knowledge of use can be acquired through engagement in different activities.

5.6 Disregard for Language Output Tasks

Researchers investigated how a range of activities were conducive to vocabulary learning in the wild. While the majority focused on activities involving language input such as listening, reading, and viewing, a fewer number of studies required participants to be involved in output tasks requiring speaking or writing. Two areas for future investigation would be how language production in non-classroom contexts could lead to vocabulary learning.

Some of the reviewed studies examined language use on social media, hinting at potential future research trajectories. Looking at the substantial body of computer-mediated communication literature might be a good starting point to inspire the design of studies that examine how vocabulary learning occurs naturally through online interactions (Smith, 2005; Yanguas, 2012). Likewise, as language learners in foreign language settings may not have a need for language use outside the classroom (Nation, 2013), it would be interesting to see more studies that highlight how classroom interactions could also be a potential source of incidental vocabulary learning. Similarly, some studies have started to examine how content lectures, for instance, may be an untapped source of incidental vocabulary learning.

5.7 Multimodal Input

Many of the reviewed articles were concerned with how engagement in digital gaming or exposure to multimodal input led to vocabulary growth. The dynamic nature of digital gaming increases the difficulty for researchers to analyze how interactions in games leads to vocabulary learning. While there has been some research that has circumvented this difficulty by using recordings of playthroughs completed by others or the use of offline measures (e.g., Kongmee et al., 2011), playing a video game and watching another person play a video game can result in very different vocabulary learning outcomes (deHaan et al., 2010). New techniques for data collection using headset eye trackers have already been used to understand eye movement behavior during action game play (Azizi et al., 2017). Using such techniques in future studies can provide more ecologically valid data on how vocabulary learning occurs through digital gameplay.

Pellicer-Sánchez (2022) highlighted in a recent review article that the published second language research has been giving more attention to multimodal reading. This represents a closing in the gap between the types of reading materials that learners are most exposed to and the types of investigations that are being given attention in the published literature. As language learners, especially those in foreign language settings, are exposed to most language input through reading (Nation, 2013), it is important for researchers to consider how the multimodal aspect of the reading materials leads to processing and thereby acquisition of new vocabulary encountered in these multimodal texts. Pellicer-Sánchez (2022) points out several areas of interest that vocabulary learning in the wild researchers should consider as potential avenues for investigation. These include investigations on reading purpose, relationship between attentional allocation and learning, and image-text relationship.

5.8 Desirable Difficulties

One last area of vocabulary learning in the wild that has not been covered in the reviewed literature is desirable difficulties. Examples of such desirable difficulties "include varying the conditions of learning rather than keeping conditions constant and predictable, distributing or spacing study or practice sessions rather than massing or blocking such sessions, using tests as learning events, reducing feedback to the learner, and providing contextual interference during learning" (Bjork & Kroll, 2015, p. 242). The increased language processing is assumed to lead to better retention of vocabulary knowledge. For example, in Bordag et al. (2015), the researchers found learners' attention was triggered when the contexts in which unknown words appeared were syntactically complex and this additional attention resulted in better incidental learning outcomes. This area of research is very promising as it possesses the potential of pushing the theoretical understanding of what makes particular tasks more or less favorable for vocabulary learning.

6 Limitations and Conclusions

While this scoping review was able to provide an overall summary of the extent, nature, and distribution of journal articles published on vocabulary learning in the wild, it is not without limitations. Being limited to categories and title, the database search was not exhaustive. In addition, as only journal articles were reviewed, the exclusion of unpublished research (e.g., PhD theses) could not account for a potential publication bias. Another limitation was only reviewing publications written in English. The lack of reviewing non-English publications could have resulted in a small number of the reviewed studies having investigated the learning of vocabulary in second languages other than English. While the systematic reviews and meta-analyses about vocabulary learning in the wild published between 2000 and 2021 were retrieved and critically evaluated to give an overview of relevant studies before these dates, it is possible some relevant studies published before the year 2000 were not touched upon in this scoping review.

The vocabulary learning in the wild literature published between 2000 and 2021 has increased the understanding of the effects of frequency and processing; learning vocabulary from listening, reading, and writing; multimodal input and gaming; learner differences; and how incidental and intentional learning of vocabulary can complement one another. Most of the 51 reviewed studies were conducted through the recruitment of tertiary-level participants under a quantitative research paradigm not underpinned by theory. Researchers most often opted to investigate the learning of vocabulary through engagement in or completion of a number of different activities after which participants were assessed in their receptive knowledge of English word meanings using researcher-designed measurements. This scoping review uncovered potential research trajectories for future researchers to examine vocabulary learning in the wild, including investigating the learning of vocabulary other than English, pre-primary and primary level learners' vocabulary learning, advancing theory, longitudinal research designs, qualitative investigations, productive form and use and receptive use assessment, digital gaming and multimodal reading, and the relationship between vocabulary task difficulty and long-term knowledge retention. In addition, the systematic map provided in the Appendix could also help to guide future research synthesists that may want to consider examination of the actual learning outcomes from the tasks that was not the focus on the current scoping review.

Acknowledgements The research reported in this chapter was supported by the Education Fund of the Macau SAR (HSS-UMAC-2021-02). I would also like to thank the reviewers for providing extensive comments and feedback on earlier versions of the chapter.

Appendix: Systematic Map of Reviewed Studies

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
1	Jonathan deHaan W. Michael Reed Katsuko Kuwada	2010	Japan	Tertiary	Japanese	English	Digital gameplay	20 min	Quantitative	None	RD PF + RU Form productive vocabulary knowledge test and use receptive vocabulary knowledge test
2	Hui-Chi Yang Yu-Chih Sun	2011	Taiwan	Tertiary	Mandarin	English	Watching open courseware lectures	20 to 26 min	Quantitative	None	RD PM Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993)
3	Emine Türk Gülcan Erçetin	2012	Turkey	Secondary	Turkish	English	Reading a text(s)	68 min	Quantitative	The Generative Theory of Multimedia Learning (Mayer, 2001)	1. RD RF Form receptive vocabulary knowledge test 2. RD PM Meaning productive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test 4. RD RM Meaning receptive vocabulary knowledge test
4	François Pichette Linda de Serres Marc Lafontaine	2012	Canada	Tertiary	French	English	Reading and writing sentences	Participant determined	Quantitative	Levels of Processing Theory (Craik & Lockhart, 1972)	RD PF Form productive vocabulary knowledge test
5	Aline Godfroid Frank Boers Alex Housen	2013	Belgium	Tertiary	Dutch	English	Reading a text(s)	N.A.	Quantitative	Noticing Hypothesis (Schmidt, 1994)	RD RU Use receptive vocabulary knowledge test
6	Barry Lee Reynolds Yi Ling Bai	2013	Taiwan	Tertiary	Mandarin	English	Self-choice reading	Participant determined	Quantitative	None	RD RM Meaning receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
7	Hilde van Zeeland Norbert Schmitt	2013	UK	Tertiary	Mixed	English	Listening to passages	50 min	Quantitative	None	1. RD RF Form receptive vocabulary knowledge test 2. RD RU Use receptive vocabulary knowledge test 3. RD PM Meaning productive vocabulary knowledge test
8	Takeshi Sato Mitsuyasu Matsunuma Akio Suzuki	2013	Japan	Tertiary	Japanese	English	Using mobile learning app(s)	30 min	Quantitative	Cognitive Load Theory (Sweller, 1994)	1. RD PM Meaning productive vocabulary knowledge test
9	Barry Lee Reynolds	2014	US & Taiwan	Tertiary	Mandarin	English	Reading a novel(s)	2 weeks	Mixed-method	None	1. RD RM Meaning receptive vocabulary knowledge test 2. RD PM Meaning productive vocabulary knowledge test 3. SM RM Vocabulary Size Test (VST) (Nation & Beglar, 2007)
10	Marie-Josée Bisson Walter J. B. van Heuven Kathy Conklin Richard J. Tunney	2014	UK	Tertiary	English	Welsh	Using word cards	N.A.	Quantitative	None	RD RM Meaning receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
11	Si-min Hu Viphavee Vongpumivitch Jason S. Chang	2014	Taiwan	Secondary	Mandarin	English	Reading a text(s)	2 weeks	Quantitative	None	RD PF + RU Form productive vocabulary knowledge test and use receptive vocabulary knowledge test
12	Barry Lee Reynolds	2015	Taiwan	Tertiary	Mandarin	English	Reading a novel(s)	2 weeks	Quantitative	Declarative/Procedural Model (Ullman, 2005)	1. RD PM Meaning productive vocabulary knowledge test 2. RD RM Meaning receptive vocabulary knowledge test
13	Denisa Bordag Amit Kirschenbaum Erwin Tschirner Andreas Opitz	2015	Germany	Tertiary	Mixed	German	Self-paced reading	40 min	Quantitative	Complementary Learning Systems (CLS) Theories of Memory (McClelland et al., 1995)	SM PM + PU Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993)
14	Jing Wu	2015	Mainland China	Tertiary	Mandarin	English	Engagement in informal digital learning of English	5 weeks	Quantitative	Psycholinguistic Theories of Human Lexical Organization and Memory (Beckwith et al., 1991)	RD PF + RF + RM + RU + PU Form meaning use receptive vocabulary knowledge test(s) and form use productive vocabulary knowledge test(s)
15	Pia Sundqvist Peter Wikström	2015	Sweden	Secondary	Swedish	English	Digital gameplay	1 year	Quantitative	None	1. SM PF + PU Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999) 2. SM RM Vocabulary Levels Test (VLT) (Nation, 2022)

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
16	Ana Pellicer-Sánchez	2016	UK	Mixed	Mixed	English	Engagement in informal digital learning of English	45 min	Quantitative	None	1. RD RF Form receptive vocabulary knowledge test 2. RD PM Meaning productive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test
17	Lung-Hsiang Wong Ronnel B. King Ching Sing Chai May Liu	2016	Singapore	Primary	English	Mandarin	Using mobile learning app(s)	13 months	Qualitative	None	RD PU Use productive vocabulary knowledge test
18	Fidel Çakmak Gülcan Erçetin	2017	Turkey	Tertiary	Turkish	English	Listening to stories	80 min	Quantitative	The Generative Theory of Multimedia Learning (Mayer, 2001)	1. RD RF Form receptive vocabulary knowledge test 2. RD PM Meaning productive vocabulary knowledge test
19	Mehmet Fatih Urun Hasan Aksoy Rasim Comez	2017	Turkey	Tertiary	Turkish**	English	Digital gameplay	1 year	Mixed-method	None	1. RD RM + RU Meaning receptive vocabulary knowledge test and use receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
20	Sarvenaz Hatami	2017	Iran	Tertiary	Farsi	English	Reading or listening to a novel(s)	75 min	Quantitative	None	1. SM RM Vocabulary Levels Test (VLT) (Schmitt et al., 2001) 2. RD RF Form receptive vocabulary knowledge test 3. RD RF Form receptive vocabulary knowledge test 4. RD PM Meaning productive vocabulary knowledge test 5. RD RU Use receptive vocabulary knowledge test 6. RD RU Use receptive vocabulary knowledge test 7. RD RM Meaning receptive vocabulary knowledge test***
21	Sunjung Lee Diana Pulido	2017	South Korea	Secondary	Korean	English	Reading a text(s)	8 weeks	Quantitative	1. Involvement Load Hypothesis (Laufer & Hulstijn, 2001) 2. Levels of Processing Theory (Craik & Lockhart, 1972)	1. RD RF Form receptive vocabulary knowledge test 2. RD RM Meaning receptive vocabulary knowledge test 3. RD RU Use receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
22	Ahmed Masrai James Milton	2018	Saudi Arabia	Tertiary	Arabic	English	Watching movies and news programs with subtitles	Between 24 and 54 h 20 min	Qualitative	None	1. SM RF Aural Lex (A-Lex) (Milton & Hopkins, 2005) 2. SM RF XK-Lex (Al-Masrai & Milton, 2012)
23	Ayman A. Mohamed	2018	US**	Tertiary	Mixed	English	Reading a novel(s)	45 to 70 min	Quantitative	None	1. RD RF Form receptive vocabulary knowledge test 2. RD PM Meaning productive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test 4. SM RM EFL Vocabulary Test (Meara, 1992)
24	Elke Peters Stuart Webb	2018	Belgium	Tertiary	Dutch	English	Watching TV programs	105 min	Quantitative	None	RD RF + PM Form receptive vocabulary knowledge test and meaning productive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
25	Feng Teng	2018	Hong Kong	Primary	Mixed	English	Watching video clips	60 min	Quantitative	Dual-coding Theory (Paivio, 1971)	1. SM RM Vocabulary Levels Test (VLT) (Schmitt et al., 2001) 2. RD RF + PM Form receptive vocabulary knowledge test and meaning productive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test
26	Sara Albaladejo Albaladejo Yvette Coyle Julio Roca de Larios	2018	Spain	Pre-primary	Spanish	English	Listening to songs and stories	6 weeks	Quantitative	1. Dual-coding Theory (Paivio, 1971) 2. Additivity Hypothesis (Paivio, 1975)	RD RM Meaning receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
27	Sarvenaz Hatami	2018	Iran	Tertiary	Farsi	English	Reading a novel(s)	3 weeks	Quantitative	None	1. RD RF Form receptive vocabulary knowledge test 2. RD RF Form receptive vocabulary knowledge test 3. RD PM Meaning productive vocabulary knowledge test 4. RD RU Use receptive vocabulary knowledge test 5. RD RU Use receptive vocabulary knowledge test 6. RD RM Meaning receptive vocabulary knowledge test
28	Ahmet Çekiç Arif Bakla	2019	Turkey**	Mixed	Turkish	English	Reading a text(s)	3 to 9 weeks	Quantitative	None	1. SM PM + PU Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993) 2. RD RM + RU Meaning receptive vocabulary knowledge test and Use receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
29	Akbar Bahari	2019	Iran	Mixed	Farsi	English	Using mobile learning app(s)	12 weeks	Mixed-method	None	RD PM + PU Meaning productive vocabulary knowledge test and use productive vocabulary knowledge test
30	Elke Peters	2019	Belgium	Secondary	Dutch	English	Watching TV programs	100 min	Quantitative	1. Dual-processing Theory of Working Memory (Baddeley, 2007) 2. The Generative Theory of Multimedia Learning (Mayer, 2001)	1. RD RF + PM Form receptive vocabulary knowledge test(s) and meaning productive vocabulary knowledge test(s) 2. SM RM Vocabulary Size Test (VST) (Nation & Beglar, 2007)
31	Eva Puimège Elke Peters	2019	Belgium	Tertiary	Dutch	English	Watching TV programs	30 min	Quantitative	None	1. RD RF + RU Form and use receptive vocabulary knowledge test 2. RD PF Form productive vocabulary knowledge test 3. RD PM Meaning productive vocabulary knowledge test 4. SM RM Vocabulary Size Test (Nation & Beglar, 2007)

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
32	Geòrgia Pujadas Carmen Muñoz	2019	Spain	Secondary	Catalan and Spanish (Bilingual)	English	Watching TV programs	1 year	Quantitative	None	1. SM RF XK-Lex (Al-Masrai & Milton, 2012) 2. RD PF Form productive vocabulary knowledge test 3. RD PM Meaning productive vocabulary knowledge test
33	Guilnard Moufarrej Charbel Salameh	2019	US	Tertiary	English	Arabic	Listening to songs	300 min	Quantitative	None	RD RM Meaning receptive vocabulary knowledge test
34	Jennifer Legault Jiayan Zhao Ying-An Chi Weitao Chen Alexander Klippel Ping Li	2019	US	Tertiary	English	Mandarin	Immersion in virtual reality	20 min	Quantitative	Embodied Cognition (Barsalou, 2008)	RD RM Meaning receptive vocabulary knowledge test
35	Johanna F. de Vos Herbert Schriefers Kristin Lemhöfer	2019	Netherlands	Mixed	German	Dutch	Using word cards	60 to 75 min	Quantitative	Theory of Cognitive Dissonance (Festinger, 1957)	1. RD PF Form productive vocabulary knowledge test 2. SM RF Dutch LexTALE (Lemhöfer & Broersma, 2012)

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
36	Ju Seong Lee	2019	South Korea	Tertiary	Korean	English	Engagement in informal digital learning of English	30 min	Mixed-method	None	1. SM PF + PU Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999) 2. SM RM Receptive Vocabulary Levels Test (RVLT) (Nation, 1990)
37	Mostafa Janebi Enayat Mohsen Haghighatpasand	2019	Iran	Tertiary	Persian	English	Digital gameplay	600 min	Mixed-method	None	1. RD RU Use receptive vocabulary knowledge test 2. RD PF + PU From productive vocabulary knowledge test and use productive test
38	Pia Sundqvist	2019	Sweden	Secondary	Swedish*	English	Digital gameplay	Participant determined	Mixed-method	None	1. SM PF Productive Levels Test (PLT) (Laufer & Nation, 1999) 2. SM RM Vocabulary Levels Test (VLT) (Nation, 2022)
39	Ting Zhao Juan Ren	2019	Mainland China	Tertiary	Mandarin	English	Reading a text(s)	5 weeks	Quantitative	Levels of Processing Theory (Craik & Lockhart, 1972)	1. RD PM Meaning productive vocabulary knowledge test 2. RD RM Meaning receptive vocabulary knowledge test
40	Yangting (Tina) Wang	2019	Mainland China	Tertiary	Mandarin	English	Watching TV programs	12 to 20 min	Quantitative	Dual-coding Theory (Paivio, 1971)	RD PM Meaning productive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
41	Barry Lee Reynolds	2020	Taiwan	Tertiary	Mandarin	English	Reading stories	4 weeks	Quantitative	None	1. SM RM Vocabulary Size Test (VST) (Nation & Beglar, 2007) 2. RD RF + RM Form receptive vocabulary knowledge test and meaning receptive test
42	Jinghui Ouyang Lingshan Huang Jingyang Jiang	2020	Mainland China	Tertiary	Mandarin	English	Reading a text(s)	N.A.	Quantitative	Noticing Hypothesis (Schmidt, 1994)	1. RD PM Meaning productive vocabulary knowledge test 2. RD RM Meaning receptive vocabulary knowledge test
43	José Ramón Calvo-Ferrer	2020	Spain	Tertiary	Spainsh	English	Digital game play	5 days	Quantitative	None	RD PM Meaning productive vocabulary knowledge test
44	Judith Borràs Àngels Llanes	2020	Ireland	Mixed	Spanish	English	Studying abroad	3 weeks	Quantitative	None	1. SM RM the Updated Vocabulary Levels Test (Updated VLT) (Webb et al., 2017) 2. RD PF + PM + PU Form meaning use productive vocabulary knowledge test
45	Marja-Leena Niitemaa	2020	Finland	Secondary	Finnish	English	Extramural Language Engagement	Participant determined	Mixed-method	None	SM RM Vocabulary Levels Test (VLT) (Schmitt et al., 2001)
46	Mohammed Alfadil	2020	Saudi Arabia	Secondary	Arabic	English	Digital gameplay	35 to 45 min (12 days)	Quantitative	None	RD PM Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993)

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
47	Sarah Sok ZhaoHong Han	2020	US	Mixed	Mixed	English	Reading a novel(s)	3 days	Quantitative	1. Levels of Processing Theory (Craik & Lockhart, 1972) 2. Monitor Theory for L2 Acquisition (Krashen, 1987) 3. Skill Acquisition Theory (Anderson, 1982)	SM PM + PU Vocabulary Knowledge Scale (VKS) (Paribakht & Wesche, 1993)
48	Anne Barwasser Karolina Urton Turid Knaak Matthias Grünke	2021	Germany	Primary	Mixed	German	Telling stories	480 min	Quantitative	Dual-coding Theory (Paivio, 1971)	1. SM RM the Culture Fair Intelligence Test (CFT) (Weiß, 2006) 2. SM RM The Peabody Picture Vocabulary Test (PPVT- IV) (Dunn & Dunn, 2007) 3. SM PF Expressive Vocabulary Test (EVT) (Williams, 2007)
49	Florence Baills Yuan Zhang Yuhui Cheng Yuran Bu Pilar Prieto	2021	Mainland China	Mixed	Mandarin	French	Listening to songs	4 min	Quantitative	None	1. RD PF Form productive vocabulary knowledge test 2. RD RM Meaning receptive vocabulary knowledge test

(continued)

(continued)

No	Author(s)	Year	Location	Educational level	L1	L2	Activity type	Duration	Methodology	Theoretical framework(s)	Vocabulary outcome measures
50	Josiah Murphy Ryan T. Miller Phillip Hamrick	2021	US	Tertiary	English	Spanish	Reading a text(s)	3 days	Quantitative	None	1. SM RM Spanish Vocabulary Levels Test (SVLT) (Chandler, 2017) 2. RD RF Form receptive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test
51	Kevin M. Wong Rachel M. Flynn Susan B. Neuman	2021	US	Pre-primary	Mixed	English	Watching educational programs	35 to 40 min	Quantitative	Dual-coding Theory (Paivio, 1971)	1. SM RM The Peabody Picture Vocabulary Test (PPVT- IV) (Dunn & Dunn, 2007) 2. RD RM Meaning receptive vocabulary knowledge test 3. RD RM Meaning receptive vocabulary knowledge test***

Note *The participants were mainly Swedish speakers, **Provided by the author(s), *** = Two aspects of receptive knowledge of meaning was assessed, None = no theoretical framework was used, N.A. = the data was unavailable in the article and unable to be obtained from the authors, + = the same aspect of vocabulary knowledge was assessed using the same measurement, RD = researcher-designed measurement, SM = standardized measurement, RF = receptive knowledge of form, RM = receptive knowledge of meaning, RU = receptive knowledge of use, PF = productive knowledge of form, PM = productive knowledge of meaning, PU = productive knowledge of use, VKS = Vocabulary Knowledge Scale

References

Albaladejo, S. A., Coyle, Y., & de Larios, J. R. (2018). Songs, stories, and vocabulary acquisition in preschool learners of English as a foreign language. *System, 76*, 116–128. https://doi.org/10.1016/j.system.2018.05.002

Alfadil, M. (2020). Effectiveness of virtual reality game in foreign language vocabulary acquisition. *Computers & Education, 153*, 1–13. https://doi.org/10.1016/j.compedu.2020.103893

Algeo, J. (1980). Where do all the new words come from? *American Speech, 55*(4), 264–277. https://doi.org/10.2307/454567

Al-Masrai, A., & Milton, J. (2012). The vocabulary knowledge of university students in Saudi Arabia. *Perspectives (TESOL Arabia), 19*(3), 13–19. https://uksacb.org/wp-content/uploads/2013/06/Research-Paper-1.pdf

Anderson, J. (2022). What's in a name? Why 'SLA' is no longer fit for purpose and the emerging, more equitable alternatives. *Language Teaching*, 1–7. https://doi.org/10.1017/s0261444822000192

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Azizi, E., Abel, L. A., & Stainer, M. J. (2017). The influence of action video game playing on eye movement behaviour during visual search in abstract, in-game and natural scenes. *Attention, Perception, & Psychophysics, 79*(2), 484–497. https://doi.org/10.3758/s13414-016-1256-7

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press.

Bahari, A. (2019). The impact of applying the FonF practice model on developing L2 listening and speaking with a focus on intentional and incidental vocabulary acquisition in call context. *Revista de Lingüística y Lenguas Aplicadas, 14*, 45–57. https://doi.org/10.4995/rlyla.2019.10785

Baills, F., Zhang, Y., Cheng, Y. H., Bu, Y. R., & Prieto, P. (2021). Listening to songs and singing benefitted initial stages of second language pronunciation but not recall of word meaning. *Language Learning, 71*(2), 369–413. https://doi.org/10.1111/lang.12442

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barwasser, A., Urton, K., Knaak, T., & Grünke, M. (2021). Intentional and incidental vocabulary acquisition through multi-component storytelling: The case of German L2 primary school students. *Language Teaching Research*, 1–34. https://doi.org/10.1177/13621688211047962

Beckwith, R., Fellbaum, C., Gross, D., & Miller, G. A. (1991). Wordnet: A lexical database organized on psycholinguistic principles. In U. Zernik (Ed.), *Using on-line resources to build a lexicon* (pp. 211–231). Psychology Press. https://doi.org/10.4324/9781315785387

Belcher, W. L. (2019). *Writing your journal article in twelve weeks: A guide to academic publishing success*. University of Chicago Press.

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan. https://doi.org/10.1057/9780230306790_2

Billington, C. (2016). *How digital technology can support early language and literacy outcomes in early years settings: A review of the literature*. National Literacy Trust. https://cdn.literacytrust.org.uk/media/documents/2016_06_16_free_research_-_digital_technology_early_literacy_review_2016_oWCanmt.pdf

Birch, B. M., & Fulop, S. (2020). *English L2 reading: Getting to the bottom*. Routledge. https://doi.org/10.4324/9780429397783

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning, 64*(4), 855–877. https://doi.org/10.1111/lang.12085

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology, 128*(2), 241–252. https://doi.org/10.5406/amerjpsyc.128.2.0241

Boone, G. (2021). How social interaction affects students' formulaic development in L2 German in a multilingual SA context. In R. Mitchell & H. Tyne (Eds.), *Language, mobility and study abroad in the contemporary European context* (pp. 159–170). Routledge. https://doi.org/10.4324/9781003087953

Bordag, D., Kirschenbaum, A., Opitz, A., & Tschirner, E. (2015). Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon. *Bilingualism-Language and Cognition, 18*(3), 372–390. https://doi.org/10.1017/S1366728914000078

Borràs, J., & Llanes, À. (2020). L2 reading and vocabulary development after a short study abroad experience. *VIAL-Vigo International Journal of Applied Linguistics, 17*, 35–55. https://doi.org/10.35869/vial.v0i17.1464

Brown, D. (2010). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research, 15*(1), 83–97. https://doi.org/10.1177/1362168810383345

Çakmak, F., & Erçetin, G. (2017). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24–47. https://doi.org/10.1017/s0958344017000155

Calvo-Ferrer, J. R. (2020). Exploring digital nativeness as a predictor of digital game-based L2 vocabulary acquisition. *Interactive Learning Environments, 28*(7), 902–914. https://doi.org/10.1080/10494820.2018.1548489

Cameron, L. (2010). Teaching languages to young learners. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511733109

Çekiç, A., & Bakla, A. (2019). The effects of spacing patterns on incidental L2 vocabulary learning through reading with electronic glosses. *Instructional Science, 47*(3), 353–371. https://doi.org/10.1007/s11251-019-09483-4

Cervatiuc, A. (2018). Incidental learning of vocabulary. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). Wiley. https://doi.org/10.1002/9781118784235.eelt0819

Chandler, P. M. (2017, July). *A Spanish vocabulary levels test* [conference presentation]. The Annual Conference of the American Association of Teachers of Spanish and Portuguese, Chicago, IL, United States.

Chapelle, C. A., & Sauro, S. (2017). *The handbook of technology and second language teaching and learning*. Wiley Blackwell.

Chiu, Y.-H. (2013). Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology, 44*(2), E52–E56. https://doi.org/10.1111/j.1467-8535.2012.01342.x

Chong, S. W., & Plonsky, L. (2021). A primer on qualitative research synthesis in TESOL. *TESOL Quarterly, 55*(3), 1024–1034. https://doi.org/10.1002/tesq.3030

Chung, E. (2018). Revisiting second language vocabulary teaching: Insights from Hong Kong in-service teachers. *The Asia-Pacific Education Researcher, 27*(6), 499–508. https://doi.org/10.1007/s40299-018-0412-3

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach*. Sage.

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

de Vos, J. F., Schriefers, H., & Lemhöfer, K. (2019). Noticing vocabulary holes aids incidental second language word learning: An experimental study. *Bilingualism-Language and Cognition, 22*(3), 500–515. https://doi.org/10.1017/s1366728918000019

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning, 68*(4), 906–941. https://doi.org/10.1111/lang.12296 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. http://llt.msu.edu/vol14num2/dehaanreedkuwada.pdf

Dewaele, J.-M. (2017). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker. *Applied Linguistics, 39*(2), 236–240. https://doi.org/10.1093/applin/amw055

Dressman, M. (2020). Introduction. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 1–12). Wiley. https://doi.org/10.1002/9781119472384.ch0

Dressman, M., & Sadler, R. W. (2020). *The handbook of informal language learning*. Wiley. https://doi.org/10.1002/9781119472384

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Elgort, I. (2017). Technology-mediated second language vocabulary development: A review of trends in research methodology. *CALICO Journal, 35*(1), 1–29. https://doi.org/10.1558/cj.34554

Enayat, M. J., & Haghighatpasand, M. (2019). Exploiting adventure video games for second language vocabulary recall: A mixed-methods study. *Innovation in Language Learning and Teaching, 13*(1), 61–75. https://doi.org/10.1080/17501229.2017.1359276

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. https://doi.org/10.1515/9781503620766

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 35*(3), 483–517. https://doi.org/10.1017/s0272263113000119

Guo, Q., Zhou, X. E., & Gao, X. A. (2021). Research on learning and teaching of languages other than English in system. *System, 100*, 1–9. https://doi.org/10.1016/j.system.2021.102541

Hatami, S. (2017). The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge. *Reading in a Foreign Language, 29*(1), 61–85. https://doi.org/10125/66728

Hatami, S. (2018). Does perceptual learning style matching affect L2 incidental vocabulary acquisition through reading? *Canadian Journal of Applied Linguistics, 21*(2), 102–125. https://doi.org/10.7202/1058463ar

Hu, H.-C. M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430. https://doi.org/10125/66973

Hu, S. M., Vongpumivitch, V., Chang, J. S., & Liou, H. C. (2014). The effects of L1 and L2 e-glosses on incidental vocabulary learning of junior high-school English students. *ReCALL, 26*(1), 80–99. https://doi.org/10.1017/s0958344013000244

Huang, H.-T. (2010). *How does second language vocabulary grow over time? A multi-methodological study of incremental vocabulary knowledge development* (Publication Number 3415908) [Doctoral dissertation, University of Hawai'i at Manoa]. ProQuest Dissertations and Theses Global.

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. https://doi.org/10.1111/j.1540-4781.2012.01394.x

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–5). Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0530

Knight, S. W., Marean, L., & Sykes, J. M. (2019). Gaming and informal language learning. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 101–115). Wiley Blackwell. https://doi.org/10.1002/9781119472384.ch7

Kongmee, I., Strachan, R., Pickard, A., & Montgomery, C. (2011). Moving between virtual and real worlds: Second language learning through massively multiplayer online role playing games (MMORPGs). In *2011 3rd computer science and electronic engineering conference (CEEC)* (pp. 13–18). IEEE. https://doi.org/10.1109/CEEC.2011.5995817

Krashen, S. (1987). *Principles and practice in second language acquisition*. Prentice Hall.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* Addison-Wesley Longman Ltd.

Laufer, B. (1989). What percentage of lexis is essential for comprehension. In C. Laurén & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters Ltd. https://www.lextutor.ca/cover/papers/laufer_1989.pdf

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 311–329). De Gruyter Mouton. https://doi.org/10.1515/9783110197372.2.311

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

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Lee, J. S. (2019). Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality? *British Journal of Educational Technology, 50*(2), 767–778. https://doi.org/10.1111/bjet.12599

Lee, J. S., & Dressman, M. (2018). When IDLE hands make an English workshop: Informal digital learning of English and language proficiency. *TESOL Quarterly, 52*(2), 435–445. https://doi.org/10.1002/tesq.422

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research, 21*(1), 118–135. https://doi.org/10.1177/1362168816637381

Legault, J., Zhao, J. Y., Chi, Y. A., Chen, W. T., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages, 4*(1), 1–32. https://doi.org/10.3390/languages4010013

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(69), 1–9. https://doi.org/10.1186/1748-5908-5-69

Lorenset, C. C., & Piazza, P. T. (2019). Digital games, foreign language and vocabulary development. *Revista EducaOnline, 13*(3), 1–19. https://revistaeducaonline.eba.ufrj.br/edições-anteriores/2019-3/digital-games-foreign-language-and-vocabulary-development

Masrai, A., & Milton, J. (2018). The role of informal learning activities in improving L2 lexical access and acquisition L1 Arabic speakers learning EFL. *Language Learning Journal, 46*(5), 594–604. https://doi.org/10.1080/09571736.2018.1520655

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164603

Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In N. Fulop, P. Allen., A. Clarke, & N. Black (Eds.), *Studying the organisation and delivery of health services: Research methods* (pp. 200–232). Routledge. https://doi.org/10.4324/9780203481981

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Meara, P. (1992). *EFL vocabulary test*. Swansea University.

Meara, P. (1997). Towards a new approach to modeling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109–121). Cambridge University Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage.

Milton, J., & Hopkins, N. (2005). *Aural Lex*. Swansea University.

Mohamed, A. A. (2018). Exposure frequency in L2 reading an eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition, 40*(2), 269–293. https://doi.org/10.1017/s0272263117000092
Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in call literature. *ReCALL, 23*(2), 135–159. https://doi.org/10.1017/s0958344011000005x
Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013
Moufarrej, G., & Salameh, C. (2019). The effects of songs on vocabulary retention in foreign language acquisition the case of Arabic. *Al-Arabiyya-Journal of the American Association of Teachers of Arabic, 52*, 101–123. https://www.jstor.org/stable/48600457
Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018, November 19). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x
Murphy, J., Miller, R. T., & Hamrick, P. (2021). Contributions of declarative memory and prior knowledge to incidental L2 vocabulary learning. *Mental Lexicon, 16*(1), 49–68. https://doi.org/10.1075/ml.20012.mur
Nagy, W. E. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64–83). Cambridge University Press.
Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*(3), 304–330. https://doi.org/10.2307/747823
Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253. https://doi.org/10.2307/747758
Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning, 24*(1), 17–38. https://doi.org/10.1080/09588221.2010.520675
Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle and Heinle.
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59
Nation, I. S. P. (2013). *What should every EFL teacher know?* Compass Publishing.
Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13. http://www.jalt-publications.org/archive/tlt/2007/07_2007TLT.pdf
Niitemaa, M.-L. (2020). Informal acquisition of L2 English vocabulary: Exploring the relationship between online out-of-school exposure and words at different frequency levels. *Nordic Journal of Digital Literacy, 15*(2), 86–105. https://doi.org/10.18261/issn.1891-943x-2020-02-02
Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26–45. https://doi.org/10.1017/S0267190505000024
Ouyang, J. H., Huang, L. S., & Jiang, J. Y. (2020, Nov). The effects of glossing on incidental vocabulary learning during second language reading: Based on an eye-tracking study. *Journal of Research in Reading, 43*(4), 496–515. https://doi.org/10.1111/1467-9817.12326
Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery* (pp. 7–32). Elsevier. https://doi.org/10.1016/B978-0-12-635450-8.50008-X
Paivio, A. (1975). Coding distinctions and repetition effects in memory. *Psychology of Learning and Motivation, 9*, 179–214. https://doi.org/10.1016/S0079-7421(08)60271-6
Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9–29. https://doi.org/10.18806/tesl.v11i1.623

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97–130. https://doi.org/10.1017/s0272263115000224

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. *ITL-International Journal of Applied Linguistics, 173*(1), 2–17. https://doi.org/10.1075/itl.21039.pel

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Muñoz, C. (2020). Introduction to the special issue. *Studies in Second Language Acquisition, 42*(3), 489–497. https://doi.org/10.1017/s0272263120000212

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/s0272263117000407

Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics, 33*(1), 66–82. https://doi.org/10.1093/applin/amr037

Pitts, M., White, H., & Krashen, S. D. (1989). Acquiring second language vocabulary through reading: A replication of the *clockwork orange* study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271–275. https://doi.org/10125/67017

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *Language Learning Journal, 47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning adolescents. *Language Learning Journal, 47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Pulido, D., & Hambrick, D. Z. (2008). The virtuous circle: Modeling individual differences in L2 reading and vocabulary development. *Reading in a Foreign Language, 20*(2), 164–190. http://hdl.handle.net/10125/66821

Ramezanali, N., Uchihara, T., & Faez, F. (2020). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly, 55*(1), 105–133. https://doi.org/10.1002/tesq.579

Restrepo Ramos, F. D. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *PROFILE: Issues in Teachers' Professional Development, 17*(1), 157–166. https://doi.org/10.15446/profile.v17n1.43957

Reynolds, B. L. (2014). A mixed-methods approach to investigating first- and second-language incidental vocabulary acquisition through the reading of fiction. *Reading Research Quarterly, 50*(1), 111–127. https://doi.org/10.1002/rrq.88

Reynolds, B. L. (2015). The effects of word form variation and frequency on second language incidental vocabulary acquisition through reading. *Applied Linguistics Review, 6*(4), 467–497. https://doi.org/10.1515/applirev-2015-0021

Reynolds, B. L. (2020). The effects of nonce words, frequency, contextual richness, and L2 vocabulary knowledge on the incidental acquisition of vocabulary through reading: More than a replication of Zahar et al. (2001) & Tekmen and Daloğlu (2006). *IRAL-International Review of Applied Linguistics in Language Teaching, 58*(1), 75–102. https://doi.org/10.1515/iral-2015-0115

Reynolds, B. L., & Bai, Y. L. (2013). Does the freedom of reader choice affect second language incidental vocabulary acquisition? *British Journal of Educational Technology, 44*(2), E42–E44. https://doi.org/10.1111/j.1467-8535.2012.01322.x

Reynolds, B. L., & Shih, Y.-C. (2019). The learning effects of student-constructed word cards as homework for the adolescent English language classroom. *System, 81*, 146–162. https://doi.org/10.1016/j.system.2019.01.005

Reynolds, B. L., & Teng, M. F. (2021). Incidental and informal vocabulary learning: Introduction to the special issue. *TESOL Journal, 12*(4). https://doi.org/10.1002/tesj.642

Reynolds, B. L., Wu, W.-H., & Shih, Y.-C. (2020). Which elements matter? Constructing word cards for English vocabulary growth. *SAGE Open, 10*(2), 1–12. https://doi.org/10.1177/2158244020919512

Reynolds, B. L., & Yu, M. H. (2017). Addressing the language needs of administrative staff in Taiwan's internationalised higher education: Call for an English as a lingua franca curriculum to increase communicative competence and willingness to communicate. *Language and Education, 32*(2), 147–166. https://doi.org/10.1080/09500782.2017.1405017

Rodgers, M., & Heidt, J. (2020). Levelling up comprehensible input and vocabulary learning: The lexical profile of videogames. In W. Werner & F. Tegge (Eds.), *Pop culture in language: Theory, research, practice* (pp. 213–225). Routledge. https://doi.org/10.4324/9780367808334

Sato, T., Matsunuma, M., & Suzuki, A. (2013). Enhancement of automatization through vocabulary learning using call: Can prompt language processing lead to better comprehension in L2 reading? *ReCALL, 25*, 143–158. https://doi.org/10.1017/s0958344012000328

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). Academic Press.

Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. *Palgrave Macmillan.* https://doi.org/10.1057/9780230293977

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly, 39*(1), 33–58. https://doi.org/10.2307/3588451

Snoder, P., & Reynolds, B. L. (2019). How dictogloss can facilitate collocation learning in ELT. *ELT Journal, 73*(1), 41–50. https://doi.org/10.1093/elt/ccy024

Sok, S., & Han, Z. H. (2020). A study of L2 vocabulary acquisition under incidental and intentional conditions. *VIAL-Vigo International Journal of Applied Linguistics, 17*, 113–140. https://doi.org/10.35869/vial.v0i17.1467

Song, T., & Reynolds, B. L. (2022). The effects of lexical coverage and topic familiarity on the comprehension of L2 expository texts. *TESOL Quarterly, 56*(2), 763–774. https://doi.org/10.1002/tesq.3100

Soyoof, A., Reynolds, B. L., Shadiev, R., & Vazquez-Calvo, B. (2022). A mixed-methods study of the incidental acquisition of foreign language vocabulary and healthcare knowledge through serious game play. *Computer Assisted Language Learning*, 1–34. https://doi.org/10.1080/09588221.2021.2021242

Soyoof, A., Reynolds, B. L., Vazquez-Calvo, B., & McLay, K. (2021). Informal digital learning of English (IDLE): A scoping review of what has been done and a look towards what is to come. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2021.1936562

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology, 23*(1), 87–113. https://doi.org/10.10125/44674

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice.* Springer. https://doi.org/10.1057/978-1-137-46048-6

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System, 51*, 65–76. https://doi.org/10.1016/j.system.2015.04.001

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sykes, J. M., & Reinhardt, J. (2013). *Language at play: Digital games in second and foreign language teaching and learning*. Pearson.

Teng, F. (2018). Incidental vocabulary learning for primary school students: The effects of L2 caption type and word exposure frequency. *Australian Educational Researcher, 46*(1), 113–136. https://doi.org/10.1007/s13384-018-0279-6

Teng, M. F. (2020). *Language learning through captioned videos: Incidental vocabulary acquisition.* Routledge. https://doi.org/10.4324/9780429264740

Thomas, N. (2020). Incidental L2 vocabulary learning: Recent developments and implications for future research. *Reading in a Foreign Language, 32*(1), 49–60. https://doi.org/10125/66576

Thomas, N., Bowen, N. E. J. A., Reynolds, B. L., Osment, C., Pun, J. K. H., & Mikolajewska, A. (2021). A systematic review of the core components of language learning strategy research in Taiwan. *English Teaching & Learning, 45*(3), 355–374. https://doi.org/10.1007/s42321-021-00095-1

Toffoli, D., & Sockett, G. (2015). University teachers' perceptions of online informal learning of English (OILE). *Computer Assisted Language Learning, 28*(1), 7–21. https://doi.org/10.1080/09588221.2013.776970

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. https://doi.org/10.1016/j.compedu.2018.06.020

Türk, E., & Erçetin, G. (2012). 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. https://doi.org/10.1080/09588221.2012.692384

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141–178). Georgetown University Press. https://doi.org/10.1038/35094573

Urun, M. F., Aksoy, H., & Comez, R. (2017). Supporting foreign language vocabulary learning through kinect-based gaming. *International Journal of Game-Based Learning, 7*(1), 20–35. https://doi.org/10.4018/ijgbl.2017010102

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Virkus, S., Koutsomiha, D., & Garoufallou, E. (2017). Digital literacy in the early years: Practices in formal settings, teacher education, and the role of informal learning spaces: A review of the literature. In S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi, & L. Roy (Eds.), *ECIL 2017: Information literacy in the workplace* (pp. 332–339). Springer. https://doi.org/10.1007/978-3-319-74334-9_35

Wang, Q. (2020). Incidental L2 vocabulary acquisition and reading: Concerns, progresses and future directions. *Chinese Journal of Applied Linguistics, 43*(4), 469–488. https://doi.org/10.1515/cjal-2020-0031

Wang, Y. T. (2019). Effects of L1/L2 captioned TV programs on students' vocabulary learning and comprehension. *CALICO Journal, 36*(3), 204–224. https://doi.org/10.1558/cj.36268

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Weiß, R. H. (2006). *Grundintelligenztest skala 2: Revision (cft 20-r) mit wortschatztest und zahlenfolgentest [basic intelligence test scale 2: Revision (cft 20-r) with vocabulary test and number sequence test]*. Hogrefe.

Williams, K. T. (2007). *Expressive vocabulary test manual* (2nd ed.). AGS Publishing.

Wong, K. M., Flynn, R. M., & Neuman, S. B. (2021). L2 vocabulary learning from educational media: The influence of screen-based scaffolds on the incidental-intentional continuum. *TESOL Journal, 12*(4), 1–20. https://doi.org/10.1002/tesj.641

Wong, L. H., King, R. B., Chai, C. S., & Liu, M. (2016). Seamlessly learning Chinese: Contextual meaning making and vocabulary growth in a seamless Chinese as a second language learning environment. *Instructional Science, 44*(5), 399–422. https://doi.org/10.1007/s11251-016-9383-z

Wu, J. (2015). Effects of call on self-directed FL vocabulary learning. *Studies in Self-Access Learning Journal, 6*(2), 191–215. http://sisaljournal.org/archives/jun15/wu

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A Meta-Analysis. *Language Learning, 71*(2), 487–536. https://doi.org/10.1111/lang.12444

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? *Studies in Second Language Acquisition, 42*(2), 411–438. https://doi.org/10.1017/s0272263119000688

Yang, H.-C., & Sun, Y.-C. (2011). It is more than knowledge seeking: Examining the effects of opencourseware lectures on vocabulary acquisition in English as a foreign language (EFL) context. *Computer Assisted Language Learning, 26*(1), 1–20. https://doi.org/10.1080/09588221.2011.624523

Yang, X., Kuo, L.-J., Eslami, Z. R., & Moody, S. M. (2021). Theoretical trends of research on technology and L2 vocabulary learning: A systematic review. *Journal of Computers in Education, 8*(4), 465–483. https://doi.org/10.1007/s40692-021-00187-8

Yanguas, I. (2012). Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO Journal, 29*(3), 507–531. https://doi.org/10.11139/cj.29.3.507-531

Zhang, R., & Zou, D. (2021). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2021.1896555

Zhao, T., & Ren, J. (2019). Incidental L2 lexical acquisition in reading: The role of L2-gloss frequency and learner proficiency. *Language Learning Journal, 47*(5), 608–624. https://doi.org/10.1080/09571736.2017.1349168

Barry Lee Reynolds is Associate Professor of English Language Education in the Faculty of Education at the University of Macau. His research interests include vocabulary learning, computer-assisted language learning, written corrective feedback, and language teacher education. More specifically, his vocabulary research has focused on incidental learning from reading novels, playing video games, text reconstruction, and listening to academic lectures. He has also conducted vocabulary studies on word card use, reader choice, lexical coverage, contextual richness, perception, extramural engagement, and metacognition, among others.

Informal Vocabulary Learning in Formal Contexts

Productive Collocation Knowledge in L2 German: Study Abroad and L1 Congruency

Griet Boone and June Eyckmans

Abstract This study explores the effect of a 5-month study abroad (SA) experience and first language (L1) congruency on the acquisition of second language (L2) German collocations using a pretest-posttest two group design. Participants were 45 Belgian students (L1 = Dutch) majoring in German and another foreign language. One group ($n = 26$) spent the semester abroad in a German-speaking country, while the other group ($n = 19$) went to a country with a different vernacular. To gauge participants' productive collocation knowledge in German, a gap-fill collocation task with 50 collocations, containing 20 congruent and 30 incongruent collocations, was administered pre- and post-SA. Although a large proportion of the variance was due to variation among participants and items rather than to the effects of SA and L1 congruency, the results point to a slight advantage in acquiring incongruent collocations for students spending a semester in the target language country.

Keywords Formulaic language · Collocations · Study abroad · L1 congruency · German

1 Introduction

Collocations—one type of formulaic language—are usually described as recurring word combinations of two or more words (e.g., *heavy rain, catch a cold*). They are considered crucial in both learning and using a second or foreign language. In the words of Wray (2002): "To know a language you must know not only its individual words, but also how they fit together" (p. 143). Research has shown that collocations can be learned successfully through classroom instruction (for reviews see, Boers & Lindstromberg, 2012; Szudarski, 2017), but due to the vast number of collocations and limited classroom time, most collocations will need to be acquired

G. Boone (✉) · J. Eyckmans
Ghent University, Ghent, Belgium
e-mail: Griet.Boone@UGent.be

J. Eyckmans
e-mail: June.Eyckmans@UGent.be

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_2

outside the classroom. According to usage-based theories, extensive naturalistic L2 exposure is needed for language learning (Ellis, 2002), and in classroom contexts this exposure is often rather limited (DeKeyser, 2007; Durrant & Schmitt, 2010; Gyllstad, 2007). Thus, it would be reasonable to expect that studying abroad in the L2 environment promotes learners' collocation development, not only because a SA experience provides the necessary L2 exposure, but also because it offers plenty of opportunities for output practice, which is crucial in L2 learning (Swain, 1985).

A number of studies in the broad area of formulaic language (e.g., speech formulas, multi-word expressions, lexical bundles and collocations) have addressed the effect of SA on the development of L2 learners' formulaic competence (Adolphs & Durow, 2004; Arvidsson, 2019; Bardovi-Harlig & Bastos, 2011; Boone, 2021; Groom, 2009; Nesselhauf, 2005; Taguchi et al., 2013). Nevertheless, only a few studies, mostly with a strong focus on the English language, have explored the effect of SA on students' development of collocation knowledge (Groom, 2009; Nesselhauf, 2005). The aim of this study is therefore to extend this line of research, by focusing on students' productive collocation knowledge in L2 German.

Previous research in SA contexts has typically examined the linguistic development of students staying in the target language (TL) country. This focus, of course, is very relevant, as it is commonly believed that the combination of immersion in the TL country, together with formal classroom learning, creates an excellent environment for L2 learning (Freed, 1995). However, it has been shown that the linguistic gains of SA participants are often limited and that there is considerable individual variation between students (DeKeyser, 2007; Kinginger, 2009). DeKeyser (2007, p. 212) argues that during SA "many students do not receive the amounts of practice that one could naively expect". One reason for this lack of practice is that SA students regularly live and socialize with other international students rather than with locals (Mitchell, 2015). On the one hand, this might have a negative impact on the learning of the TL; on the other hand, this international environment may offer opportunities for the acquisition of other languages. Recently, the complex social context of SA (with a strong focus on the multilingualism of the European setting) and its implications on L2 students' linguistic development has attracted research attention (Borràs & Llanes, 2019; Peréz-Vidal & Llanes, 2021). Especially in the case of English, which is often used as a medium of instruction or a lingua franca, it has been shown that a SA experience in countries where English is not the country's official language (e.g., France, Germany, Finland) may contribute to L2 learners' English development (Kölyü, 2016; Llanes, 2019; Llanes et al., 2016). Further, it has been suggested that there might be linguistic development in the case of other languages too after a SA in a multilingual environment (Boone, 2021), although more research is needed.

An important factor affecting collocation development is L1 congruency (Ding & Reynolds, 2019; Laufer & Waldman, 2011; Nesselhauf, 2005; Peters, 2016; Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). First, it has been found that many collocation errors produced by L2 learners are L1 based, because learners often rely on a word-for-word translation from their L1 (Laufer & Waldman, 2011). Second, congruent collocations (i.e., collocations with a word-for-word-translation L1 equivalent in the L2) have been shown to offer a processing advantage for the L2 learner compared to incongruent collocations (i.e., L2-specific collocations with no direct equivalent in learners' L1) (Ding & Reynolds, 2019; Wolter & Gyllstad, 2011, 2013). However, to the best of our knowledge, only the study of Yamashita and Jiang (2010) has examined the effect of both L1 congruency and L2 exposure on the acquisition of L2 collocations by using a phrase-acceptability judgement task. The current study aims to add to these findings by exploring students' acquisition of L2 German collocations, using a productive collocation task and taking into account the effect of a 5-month SA experience and L1 congruency. By adopting a pretest-posttest two group design, the study compares two groups of SA exchange students in their collocation development in L2 German: one group ($n = 26$) staying in the TL country (Germany or Austria), the other group ($n = 19$) spending the same period abroad, but outside the TL country (i.e., in France, Spain, the United Kingdom, Ireland, Turkey or Russia). Collocation knowledge was assessed with a productive collocation task, including 20 congruent and 30 incongruent collocations, to explore both the role of L1 congruency and the role of the specific SA context in the development of L2 productive collocation knowledge.

2 Background

2.1 Collocations in Second Language Acquisition

Over recent decades, it has become clear that vocabulary consists not only of single words which are held together by syntax, but that it also includes a considerable number of formulaic sequences (FSs) (Schmitt, 2012). These FSs, fixed combinations of words, can be divided into various categories, for example, idioms (e.g., *it's a piece of cake*), pragmatic expressions (e.g., *have a nice day*), and collocations (e.g., *pay attention, strong coffee*), to name just a few. Collocations, defined as "frequently recurring two-to-three-word syntagmatic units which can include both lexical and grammatical words" (Henriksen, 2013, p. 29), are the focus of the present study. It has been shown that FSs are widespread in authentic written and spoken language (Erman & Warren, 2000), and that they are extremely important for L2 learners if they aim to increase their L2 fluency and proficiency (Pawley & Syder, 1983). In general, they are seen as "important building blocks for language acquisition and processing" (Christiansen & Arnon, 2017, p. 542).

Despite the importance of collocations, the development of L2 collocation competence has been shown to be a complex, slow and challenging process for L2 learners, with even advanced learners using atypical word combinations in the L2 (Boers et al., 2014; Conklin & Schmitt, 2012; Laufer & Girsai, 2008; Laufer & Waldman, 2011; Nesselhauf, 2005). Alongside the traditional focus on single words in L2 classroom vocabulary learning (Szudarski, 2017), the lack of semantic salience of collocations is another issue that makes the acquisition of collocations a slow process. For example, in verb-noun collocations containing a high-frequency verb (e.g., *have a nightmare*, *make a mistake*), the noun tends to be more salient because it carries most of the meaning and is therefore given more attention (Boers et al., 2014). The fact that the semantically more informative noun constituent may overshadow the verb component in the input seems to diminish L2 learners' ability to notice important differences in word choice between the L2 and the L1 (Ellis, 2006). As a result, learners tend to produce collocations using L1 translation equivalents (Nesselhauf, 2005). In the case of congruent collocations, the result will be an acceptable collocation in the L2. However, for incongruent collocations, this way of processing will often result in non-idiomatic combinations in the L2. Indeed, different studies which have examined the effect of L1 congruency on students' collocation processing and use have shown that especially incongruent collocations constitute a learning burden for L2 learners (Ding & Reynolds, 2019; Nesselhauf, 2005; Peters, 2016; Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). Another major issue for L2 learners in their (classroom) acquisition of collocations is said to be the lack of repeated and varied L2 exposure to word combinations which sound natural to speakers of the TL (Durrant & Schmitt, 2010; Gyllstad, 2007). Moreover, a classroom context does not offer sufficient opportunities for learners to reinforce their collocation knowledge by repeatedly using the TL (Henriksen, 2013). For example, Gyllstad (2007) found that a period of 4–6 months of exposure to English in a university setting in Sweden is not enough for Swedish students to develop their receptive English collocation knowledge significantly.

2.2 Collocation Development After a Stay in the TL Country

To help reduce two of the aforementioned barriers (L1 influence and lack of repeated L2 exposure), a long stay in an L2-speaking country seems promising, as the immersion in the language and culture of the L2 has the potential to offer students a huge amount of input, output, and interaction opportunities (Pérez-Vidal, 2014). Of course, students' language proficiency gains after SA depend on a range of individual differences, including learners' active engagement in the L2-speaking community (Coleman, 2013; DeKeyser, 2007; Kinginger, 2015). But it has also been shown that the acquisition of FSs can indeed be enhanced by SA in an L2 environment (Adolphs & Durow, 2004; Arvidsson et al., 2019; Bardovi-Harlig & Bastos, 2011).

However, the few studies that have been conducted on collocation development in SA contexts have produced mixed findings. Siyanova and Schmitt (2008), for instance, explored the judgements of native speakers (NSs) and non-native speakers (NNSs) of native-like and atypical collocations. They found that exposure during a stay in an English-speaking country led to significantly better intuitions in NNSs about which English collocations are native-like. However, to develop intuitions about which collocations were atypical, a stay of more than 12 months in an English-speaking country was needed. Nesselhauf (2005) examined the written production of English collocations by advanced German English as a Foreign Language (EFL) students. In her large-scale corpus-study, it was established that there was only a slight improvement in the number of accurate collocations produced after an extended stay in an English-speaking country. She also found that the length of stay did not seem to lead to a higher use of L2 collocations; on the contrary, immersion-students seemed to produce fewer collocations than non-immersion students. Groom (2009) also investigated the effects of L2 immersion on L2 collocation development, analysing students' writing in two groups of Swedish learners of English. One group spent less than a month in an English-speaking environment and one group stayed 12 months or more in the TL environment. He used a lexical bundle analysis (searching for exact repetitions of a multiword sequence in a corpus such as *you know what*, *on the other hand*) and a node and collocates analysis (identifying single word forms that occur significantly frequently within a short span of a pre-specified 'node' word, such as *agree on*, *knowledge of*) on corpus data from the *USE—Uppsala Student English Corpus* (Axelsson, 2000). In accordance with the findings in Nesselhauf (2005), Groom (2009) also found that collocation accuracy in student essays correlated positively with L2 immersion. However, he stated that the gap between L2-immersed and non-immersed students might be larger than Nesselhauf (2005) suggests. The lexical bundle analysis showed that the number of produced collocations correlated negatively with the amount of time spent abroad. However, the node and collocate analysis showed a positive correlation between the number of collocations produced and L2 immersion. Apparently, the different findings on collocation development depend on the research method used (e.g., lexical bundle analysis versus node and collocates analysis) (Groom, 2009).

In two longitudinal studies, Li and Schmitt (2009, 2010) provided a description of individual students' development of collocations in written production. Li and Schmitt (2009) followed an advanced Chinese master's student over the course of one academic year in Nottingham. In this case study, it was found that lexical phrases were learned incrementally. In particular, the student gradually increased her productive phrasal knowledge in English in terms of appropriacy, meaning that partially known phrases became more appropriately used over time. However, the absolute number of phrases she used was still rather low. From interviews, it became clear that phrases were learned from both explicit and implicit sources, and that she became more confident in using them. In their second study, Li and Schmitt (2010) analysed the essays and theses of four Chinese students of English studying in Nottingham for one year, focusing on the production of adjective-noun collocations. Li and Schmitt (2010) found little to no improvement in students' production of collocations, although there was considerable variation between the individual students.

What should be noted is that the aforementioned studies do not take into account the effect of L1 congruency on collocation development. To the best of our knowledge, only Yamashita and Jiang (2010) investigated both L1 congruency and L2 exposure on the acquisition of L2 collocations. Using a phrase—acceptability judgement task, they compared different groups: 20 NSs of English, 24 Japanese users of English as a second language (ESL) with about 5 years of residence in English-speaking-countries, and 23 Japanese EFL learners who had never lived in English-speaking-countries. They found that the EFL learners made more errors with incongruent collocations, and that when learners had to respond to incongruent collocations their reaction time was longer. For ESL users, there was no statistically significant difference in reaction time between congruent and incongruent collocations, although they too made more mistakes on incongruent collocations. The authors concluded that "the long-lasting congruency effect on the ESL users' error rate suggests that incongruent collocations are difficult to accept in the L2 mental lexicon" (Yamashita & Jiang, 2010, pp. 660–661). In short, their study showed that the acquisition of L2 collocations takes a long time, with both L1 congruency and the amount of L2 input influencing the acquisition process, and that especially the acquisition of incongruent collocations requires a massive amount of exposure to the L2.

2.3 L2 Development After SA in a Multilingual Environment

Interestingly, it is often assumed that students who spend a semester or longer in the TL country will automatically get the necessary exposure, and that opportunities for language development are limited to a stay in the TL country. However, in the case of L2 English, it has been shown that successful L2 development is also possible after SA in countries where English is not the official language of the country (Kölyü, 2016; Llanes, 2019; Llanes et al., 2016). This might not be surprising, as English is used as a lingua franca in many countries, especially in academic settings. However, similar language contact contexts might arise for other languages: L2 learners might interact with international students who are L1 users of the language(s) they want to learn. For example, a German student majoring in Italian and interacting regularly with Italian students during SA in France. In an exploratory study with four cases, Boone (2021) focused on this multilingual context during SA, and investigated the effect of social interaction on students' formulaic development in German with the focus on pragmatic expressions and collocations. The four cases illustrated that language learners who engage in meaningful L2 social interaction, even in a non-TL country, can benefit from SA in terms of formulaic development. As for collocation development, the results of the study suggest that incongruent collocations are the most difficult to acquire, even with substantial L2 exposure. However, because this study was exploratory in nature and limited to only 4 participants, additional quantitative studies are needed to confirm these qualitative findings on the effect of L1 congruency and a multilingual SA experience on L2 learners' collocation development.

3 The Study

The study presented here is part of a larger, longitudinal research project tracking students' phraseological competence over time. Its aim is to evaluate learners' productive collocation knowledge in L2 German after SA. The focus is on productive knowledge because it has been shown that L2 learners encounter problems with the accurate and idiomatic production of collocations (Laufer & Waldman, 2011). Additionally, productive tasks are often more demanding than receptive ones (Webb, 2008), which is something to take into account when investigating a language pair with a high lexical similarity like Dutch and German. Finally, as far as we know, no study to date has investigated the effect of L2 exposure and L1 congruency on students' L2 production of collocations. Only the study of Yamashita and Jiang (2010) examined the interaction between L1 influence and L2 exposure on learners' receptive knowledge of L2 collocations. In this study, the interaction between L1 congruency and the time spent abroad in a German-speaking versus a non-German-speaking country will be investigated.

Accordingly, the research question is as follows: What is the effect of destination country (students staying in the TL country vs. students staying in a non-TL country) and type of collocation (congruent vs. incongruent collocations) on students' productive collocation development during study abroad?

3.1 Participants

The participants in this study were 45 Dutch-speaking Belgian students, of whom 36 were female and 9 male. They were 20 or 21 years old at the time of the study, with a mean age of 20.29 years (SD = 0.5). They were undergraduate students, majoring in German and another foreign language (i.e., French, English, Spanish, Italian, Russian or Turkish). Their proficiency level in German was B1 (intermediate) for speaking and writing and B2 (upper-intermediate) for listening and reading according to the Common European Framework of Reference (CEFR) (Council of Europe, 2001). At the beginning of our study, they all had had the same amount of formal instruction in German at university (405 contact hours) and had earned 36 credit points for German according to the European Credit Transfer System (ECTS). As a curriculum requirement, students are expected to participate in the ERASMUS Programme, the student exchange programme of the European Union with a minimum stay of 3 months and a maximum stay of 12 months at a host university abroad. The participants in our study are obliged to go on a 5-month exchange to a country in which one of their L2 languages is spoken during the third academic year of their degree. Twenty-six of the 45 participants went to a German-speaking country (i.e., Germany and Austria) while 19 spent the semester abroad in a non-German-speaking country (i.e., France, Spain, the United Kingdom, Ireland, Turkey and Russia). The participants continued to study German at their respective host university and they all completed their SA programme successfully. All participants took part in the present study on a voluntary basis and in accordance with the ethical standards set out by the home university.

3.2 Materials and Procedure

The instrument used for the data collection was a productive gap-fill collocation test targeting 20 congruent and 30 incongruent collocations. The participants took this test before SA and after they came back. In this test, students were asked to fill in the gaps in sentences. The gaps had to be filled with a German collocation, and a L1 (Dutch) translation was provided between brackets (see Appendix).

To design this task, the official German B1 word list (Glaboniat et al., 2013) was taken as a starting point. Collocations appearing in the sample sentences accompanying items in this list were selected (for example, *rein – Der Pullover ist aus **reiner Wolle***, ‘The sweater is made of **pure wool**’). For German, only two collocation dictionaries are available: Das *Wörterbuch der Kollokationen im Deutschen* (Quasthoff, 2011) and *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag* (Häcki Buhofer et al., 2014). Both were consulted to check that all target collocations appear in at least one of the two. If this was the case, the collocations were selected as candidate items, resulting in a pool of 55 items. The collocations were also cross-checked with the German news corpus of the Leipzig Corpora Collection (2011). A gap-fill task was developed and piloted with a group of 22 students of German who completed the gap-fill task with the 55 items. The aim of this pilot test was to check item reliability and identify ambiguous items. As a result, five items were discarded because they could be translated with a single verb. For example, the Dutch collocation *verslag uitbrengen* can be translated in German with *Bericht erstatten* (‘give a report’), but also with the verb *berichten* (‘report’), which, of course, is not a collocation. Then, two similar gap-fill tasks were developed, containing the same 50 collocations. The items were put in randomized order on the pre- and posttest. The reliability of the collocation task showed a Cronbach’s alpha of 0.67 for the pretest and 0.73 for the posttest, indicating acceptable internal consistency (George & Mallery, 2010).

To examine the influence of L1 congruency, the 50 collocations were divided into two categories: congruent and incongruent. Since the level of congruency between L1 and L2 collocations can be open to debate (Nesselhauf, 2005), the following definition was applied: Collocations were considered congruent in L1 and L2 when they have a word-for-word-translation equivalent in the L1, and incongruent when there is no such literal L1 translation equivalent (following Ding & Reynolds, 2019; Nesselhauf, 2005; Peters, 2016; Wolter & Gyllstad, 2013). For example, for Dutch-speaking learners of German, the German collocation *Erfarhungen sammeln* (‘gain experience’ or, literally, ‘collect experience’) is incongruent with Dutch, as the Dutch equivalent is *ervaring opdoen* (not ‘collect’, but ‘acquire’ experience). An example of a congruent collocation for Dutch-speaking learners of German is *eine Rolle spielen* (*een rol spelen* in Dutch = ‘play a role’). Although these examples are quite transparent, congruency is not always that straightforward. Therefore, the target collocations were presented to 11 independent raters. The raters were sent an Excel-file with the 50 L2 collocations and their L1 translation. They were asked to read each collocation and each translation carefully and to indicate whether the provided L1 translations were literal equivalents for the L2 collocations. In the case of literal L1 equivalency, they labelled the collocation as 'congruent'; when this was not the case, they labelled it as 'incongruent'.

All raters were experienced university lecturers of German. Interrater reliability was estimated with Fleiss's kappa, which indicates agreement of $\kappa = 0.654$, 95% CI [0.616, 0.691] between the raters' judgements. According to Landis and Koch (1977), a kappa coefficient between 0.61 and 0.80 establishes substantial agreement. Thus, it was decided to give the items the dichotomous labels 1 (for congruent) and 0 (for incongruent) based on the raters' majority votes. This resulted in 20 congruent and 30 incongruent collocations.

The first data collection took place before the students' departure, the second after the students' return. Participants completed the gap-fill collocation test in about 25 min. Scoring was dichotomous: one point for a correct translation into German, zero points for an incorrect or incomplete translation. For each participant, a total score and a subtotal score for both congruent and incongruent collocations was calculated.

4 Results

4.1 Descriptive Statistics

Key descriptive statistics can be found in Table 1. Participants' mean values obtained during the pre- and posttest for congruent and incongruent collocations are presented together with standard deviations (in parentheses) as percentages for ease of comparison. Table 1 also contains the students' total score.

In the case of congruent collocations, it can be seen that the mean test score of the entire group ($N = 45$) before SA amounted to 69% and remained relatively stable after SA (68.89%). In the case of incongruent collocations, overall scores were considerably lower than the scores for congruent collocations, and there was a small mean gain between pre and post SA (49.11% vs. 51.56%). The total mean test score was hardly affected between pre- and posttest (57.07% vs. 58.49%).

When the scores are compared pre- and post-SA in both groups and for both collocation types, no significant change in learners' knowledge of congruent collocations after SA was attested: A small gain (+1.58%) was observed in the non-TL-group and a small attrition (−1.35%) was seen in the TL-group. However, for incongruent collocations the opposite was true: In the non-TL-group attrition (−1.76%) was observed, whereas the TL-group gained 5.51%. Since mean scores are susceptible to random variation arising from both participants and test items, inferential analysis is needed.

Table 1 Descriptive statistics for both groups pre- and post-SA

Group	$N_{Learners}$	Congruent collocations Mean test scores (%)		Incongruent collocations Mean test scores (%)		Total collocations Mean test scores (%)	
		Pre-SA	Post-SA	Pre-SA	Post-SA	Pre-SA	Post-SA
SA non-TLC	19	69.21 (11.70)	70.79 (13.46)	52.46 (9.55)	50.70 (13.22)	59.16 (9.09)	58.74 (11.91)
SA TLC	26	68.85 (13.44)	67.50 (13.87)	46.67 (11.43)	52.18 (11.70)	55.54 (10.81)	58.31 (11.45)
Total	45	69.00 (12.60)	68.89 (13.65)	49.11 (10.95)	51.56 (12.24)	57.07 (10.18)	58.49 (11.51)

Note SA (non)-TLC = Study Abroad in (non)-target language country

4.2 *Inferential Statistical Analysis*

Prior to the analysis process, all relevant assumptions were tested. Each participant belonged only to one group, ensuring independence of observations. Using boxplot methods, no outliers were detected. Results of the Shapiro-Wilk test and QQ plots computed for each group indicated that it was reasonable to assume distributional normality for our data. Homogeneity of variances was examined using Levene's test, which was non-significant for both groups.

An independent sample *t*-test was conducted to make sure that both groups were comparable with regard to their pre-departure productive collocation knowledge. No significant pre-SA differences were found between the 19 participants who were going to a non-TL country compared to the 26 who were going to a TL country for congruent collocations ($t(43) = -0.095$, $p = 0.925$), incongruent collocations ($t(43) = -1{,}1796$, $p = 0.08$), and total collocation score ($t(43) = -1{,}184$, $p = 0.243$).

To be able to account for individual variance among participants and items, it was decided to analyze the data of this study with generalized linear mixed modeling (GLMM) instead of with a three-way mixed ANOVA. Since GLMM can handle more than one random factor (Brysbaert, 2020), this statistical analysis was preferable for the purposes of this study. Effects of random factors are assumed to be study-specific (i.e., specific to the participants and items tested in the study), whereas effects of fixed factors are thought to stay constant and can be replicated (Brysbaert, 2020). The results of a mixed-effects analysis also afford a superior basis for generalizing to linguistic items and learners that were not involved in the study (e.g., Baayen, 2008). Moreover, GLMM is remarkably robust when assumptions are violated or when data are unbalanced (Schielzeth et al., 2020). The R (R Core Team, 2019) package lme4, version 1.1–21 (Bates et al., 2019) was used to construct the model using the glmer-function because responses for the dependent variable (collocation score) were binary: 0 for an incorrect collocation, 1 for a correct collocation.

Our model included random effects for participants and items. As the main research question was to examine the interactions between group (TL vs. non-TL country), time (pre- and post-SA) and congruency (congruent vs. incongruent collocations), these three variables were entered to predict the outcome variable (collocation score). Models were fitted using a maximum likelihood technique.

Marginal R^2 was calculated, which measures the variance explained by the fixed effects only, and conditional R^2, which measures the variance explained by both the fixed effects (in this case, group, time, and congruency) and the random effects (in this case, participants and items), using the performance package (Lüdecke et al., 2020) in R. Interactions were plotted using the sjPlots package (Lüdecke, 2021).

For our model, including all fixed and random effects, the results show a significant positive two-way interaction between time and group ($p = 0.045$), and a statistically significant negative three-way-interaction between time, congruency, and group ($p = 0.045$). This indicates that students spending time abroad in a German-speaking country scored better on the incongruent collocations compared to their peers staying in another country. Table 2 provides the parameter estimates and their standard errors

Table 2 Generalized linear mixed effects model predicting right or wrong answers on the test

	Random effects	
Parameter	Variance	*SD*
Participant	0.28	0.53
Item	2.15	1.47

	Fixed effects			
Parameter	Estimate	*SE*	*Z*	*p*-value
(Intercept)	0.28	0.37	0.76	0.447
Time	−0.10	0.16	−0.63	0.528
Group	−0.76	0.34	−2.26	0.024*
Congruency	0.71	0.56	1.28	0.201
Time × Group	0.42	0.21	2.00	0.045*
Time × Congruency	0.21	0.23	0.95	0.345
Group × Congruency	0.91	0.47	1.93	0.053
Time × Group × Congruency	−0.59	0.30	−2.00*	0.045*

Note *p < 0.05

and *z* values. Figure 1 provides a visual representation of the interaction effects. The model reported a marginal R^2 of 0.05 and a conditional R^2 of 0.45, which showed that an extra 40% of the variance was explained by the random effects.

5 Discussion

The aim of this study was to evaluate changes in the productive collocation knowledge of Dutch-speaking Belgian students of German before and after SA, thereby focusing on the effect of destination and congruency. The overall gain was small, a finding which is in line with previous studies and which corroborates previous findings that collocation development is slow, even for advanced language students in an academic context (Boers et al., 2014; Gyllstad, 2007). Students' productive collocation knowledge developed little during the 5 months of the study, despite exposure to German at the respective host universities abroad. This could be explained by the fact that 5 months in a university setting in a non-TL country may not be enough to make good progress in L2 productive collocation knowledge. As Gyllstad (2007) found, a period of 4–6 months of university exposure to English in Sweden did not lead to significant gains in students' receptive English collocation knowledge. The group of students that spent their SA in a German-speaking country did not make much progress, either. This finding confirms the view that miraculous language gains after studying in a TL country cannot be expected across the board (DeKeyser, 2007), and that quantity of TL contact alone does not determine formulaic development during SA (Arvidsson, 2019).

Fig. 1 Prediction of learners' development of score for incongruent (CONGR = 0) and congruent (CONGR = 1) collocations according to SA destination (non-TL destination: SA_TARGET = 0; TL destination: SA_TARGET = 1)

Regarding congruency, a clear difference was attested in students' test scores: Both pre- and posttest scores for congruent collocations were about 20% higher than those for incongruent collocations. This might be explained by the fact that German and Dutch are highly related languages and that collocations are often produced through a process of word-for-word translation (Laufer & Waldman, 2011). Using a literal L1 equivalent works perfectly for congruent collocations, but in case of incongruent collocations this results in non-idiomatic L2 collocations. The higher rate of correct congruent collocations in our study thus seems to support the evidence in the literature on L1–L2 congruency (Ding & Reynolds, 2019; Nesselhauf, 2005; Peters, 2016; Wolter & Gyllstad, 2011).

Another observation is that the group spending a semester outside the TL country started out with higher pre-test scores, compared to the TL country group. Although the difference was not significant, the lower pre-test scores for the TL-group could account for those students' choice to go to a German-speaking country. This is in line with the findings of the qualitative study of Boone (2021), in which students with lower formulaic language scores said they picked a German-speaking country as SA destination to improve their German language skills. It also shows that students expect their L2 to improve when they spend time in a TL country.

The central research question of this study concerned the interaction of destination, time, and congruency. We wanted to find out whether the development in students' productive collocation knowledge varies by group (TL vs. non-TL destination) and by type of collocation (congruent vs. incongruent). The significant interaction effects which were found in the GLMM (i.e., 2-way between time and group, and 3-way between time, group, and congruency), both visualized in Fig. 1, point to an advantage in the acquisition of incongruent collocations for students spending a semester in the TL country. The choice of those students who decided to go to a German-speaking country appeared to have had an impact on their post-test scores for incongruent collocations. Although students' overall gain in collocation knowledge was small after spending 5 months abroad, studying in the TL country did seem beneficial for acquiring incongruent collocations. This finding adds to the results of Yamashita and Jiang (2010), who found that both L1 congruency and L2 exposure affect students' receptive knowledge of L2 collocations. In this study, it seems that L1 congruency and L2 exposure also affect students' productive knowledge.

What should be remarked, however, is the fact that in our GLMM the fixed effects explain relatively little of the overall variance in the post-test scores. On the other hand, the random factors explain a substantial part of that variance, which certainly highlights the extent to which different learners respond differently to similar input and the extent to which different collocations are differently responded to. This finding suggests that both individual differences and item-related factors play a role in L2 acquisition and should be considered in future studies.

6 Conclusion and Pedagogical Implications

The present study set out to examine the effect of SA and L1 congruency on students' productive knowledge of L2 German collocations. It has expanded the scope of existing research on collocation development by looking at another language than English. It was found that although only little progress in productive collocation knowledge was made, the results point to an advantage in the acquisition of incongruent collocations for students spending a semester in the TL country. Our findings are generally consistent with previous research, indicating an effect of L1 congruency and suggesting that SA in a TL country is helpful when learning a foreign language.

This study has a few limitations, some of which point to avenues for future research. First, despite five months being a common duration within the ERASMUS programme, a 5-month period could be considered rather brief to measure gains in collocation knowledge. Further studies over a longer time span might reveal additional insights. Second, since the scoring used in this study was binary, the exact source of error remains unidentified. Therefore, it would be interesting to engage in error analysis in future studies to identify sources of error. Third, we do not have evidence for the fact that our target collocations were not taught in students' classes at the different universities abroad, although this seems very unlikely. In any case, different students will reach different outcomes depending on who they interact with and what they do when abroad. Therefore, as DeKeyser (2007) suggests, it is important to combine quantitative and qualitative methodologies when examining the influence of SA on students' linguistic development. Finally, because it was found in this study, as in other previous studies, that collocation learning is a highly complex and dynamic process, more factors affecting the acquisition of collocations should be considered in future studies: Both item-related factors (e.g., frequency, collocate-node relationship) and learner-related factors (e.g., prior vocabulary knowledge, phonological memory, engagement in social interaction).

As for pedagogical implications, the findings suggest that a L2 immersion experience in the TL country might be useful for acquiring the more difficult incongruent collocations, and that teachers should therefore emphasize the importance of using the opportunities of a SA experience. Based on this study, in which (incidental) gains were rather small, it seems that explicit instruction on collocations is needed to boost the learning process. We recommend that teachers start with emphasizing the importance of collocations for fluent and appropriate communication. We propose that they direct learners' attention to the differences or similarities between L1 and L2 collocations, and work with contrastive analysis and translation, as suggested by Laufer and Girsai (2008); or that they use other techniques that have been shown to be effective, like studying collocations as holistic units (Boers et al., 2017), dictogloss exercises (Lindstromberg et al., 2016; Snoder & Reynolds, 2019) or retrieval practice activities (Boone & Eyckmans, 2020). Finally, it is important to present learners with sufficient L2 exposure and give them the opportunity to produce the collocations in class. This will lead students to call up idiomatic word combinations, instead of resorting to word-for-word translation strategies.

Acknowledgements We thank two anonymous reviewers for their constructive comments that improved this paper. We also thank Seth Lindstromberg for his extraordinary statistical support and helpful feedback on an earlier draft of this manuscript. We are grateful to all participants for their time and effort.

Appendix

Part of the L2 (German) gap-fill translation task with the target items in L1 Dutch between brackets

1. Wir haben eine Safari gemacht und hatten da kein (stromend water) _______________. Wir konnten sogar kein (naar de radio luisteren) _________________, aber wir konnten interessante (gesprekken voeren) _______________________________.
2. Sie müssen (in gedachten houden) ______________________________, dass Sie im Notfall andere und sich selbst zuerst (in veiligheid brengen) ___________________ sollten oder den Notarzt rufen sollten, bevor Sie (eerste hulp) ______________ leisten.
3. Eine (nauwe samenwerking) ______________________________ zwischen den Abteilungen ist etwas, worauf wir großen (belang hechten) ____________________. Wir sind aber auf dem (juiste weg) __________________.
4. Es ist wichtig, dass Sie, als (toekomstige moeder) _______________________ nicht ständig (een sigaret opsteken) _______________________________ oder (blond bier) ________________________ trinken.
5. Die Wettervorhersage für heute: leider kein (blauwe hemel) ___________________.

References

Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 107–126). John Benjamins. https://doi.org/10.1075/lllt.9

Arvidsson, K. (2019). Quantity of target language contact in study abroad and knowledge of multi-word expressions: A Usage-Based approach to L2 development. *Study Abroad Research in Second Language Acquisition and International Education, 4*(2), 145–167. https://doi.org/10.1075/sar.18001.arv

Arvidsson, K., Forsberg Lundell, F., & Bartning, I. (2019). Réseaux sociaux et développement linguistique: Une étude de cas en français L2. *Language, Interaction and Acquisition, 10*(2), 255–288. https://doi.org/10.1075/lia.19004.arv

Axelsson, M. (2000). USE—The Uppsala Student English corpus: An instrument for needs analysis. *ICAME Journal, 24*, 155–157.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Bardovi-Harlig, K., & Bastos, M.-T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics, 8*(3), 347–384. https://doi.org/10.1515/iprg.2011.017

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2019). Linear mixed-effects models using 'Eigen' and S4. *R Package lme4, Version 1*.1–21. https://github.com/lme4/lme4/

Boers, F., Dang, T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research, 21*(3), 362–380. https://doi.org/10.1177/1362168816651464

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83–110. https://doi.org/10.1017/S0267190512000050

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *VIAL-Vigo International Journal of Applied Linguistics, 11*, 41–62.

Boone, G. (2021). How social interaction affects students' formulaic development in L2 German in a multilingual SA context: Four case studies. In R. Mitchell & H. Tyne (Eds.), *Language and mobility: Study abroad in the contemporary European context* (pp. 154–170). Taylor & Francis.

Boone, G., & Eyckmans, J. (2020). Attention-direction versus retrieval practice: Which fosters the productive recall of German formulaic sequences best? *Vigo International Journal of Applied Linguistics, 17*, 9–34. https://doi.org/10.35869/vial.v0i17.1463

Borràs, J., & Llanes, À. (2019). Re-examining the impact of study abroad on L2 development: A critical overview. *The Language Learning Journal*, 1–14. https://doi.org/10.1080/09571736.2019.1642941

Brysbaert, M. (2020). *Basic statistics for psychologists* (2nd ed.). Red Globe Press.

Christiansen, M., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science, 9*(3), 542–551. https://doi.org/10.1111/tops.12274

Coleman, J. (2013). Researching whole persons and whole lives. In C. Kinginger (Ed.), *Social and cultural aspects of language learning in study abroad* (pp. 17–44). John Benjamins.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61. https://doi.org/10.1017/S0267190512000074

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge University Press.

DeKeyser, R. (2007). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge University Press. https://doi.org/10.1017/CBO9780511667275

Ding, C., & Reynolds, B. L. (2019). The effects of L1 congruency, L2 proficiency, and the collocate-node relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics, 17*(2), 331–357. https://doi.org/10.1075/rcl.00038.din

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text—Interdisciplinary Journal for the Study of Discourse, 20*(1). https://doi.org/10.1515/text.1.2000.20.1.29

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. (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. https://doi.org/10.1093/applin/aml015

Freed, B. (Ed.). (1995). *Second language acquisition in a study abroad context*. Benjamins.

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 17.0 update* (10th ed.). Allyn & Bacon.

Glaboniat, M., Perlmann-Balme, M., & Studer, T. (2013). *Zertifikat B1: Deutschprüfung für Jugendliche und Erwachsene: Prüfungsziele, Testbeschreibung* (1. Auflage). Hueber Verlag.

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 21–33). Palgrave Macmillan UK. https://doi.org/10.1057/9780230245327_2

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund University.

Häcki Buhofer, A., Dräger, M., Meier, S., & Roth, T. (2014). *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag*. Francke Verlag.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29–56). Eurosla.

Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. Palgrave Macmillan.

Kinginger, C. (2015). Student mobility and identity-related language learning. *Intercultural Education, 26*(1), 6–15. https://doi.org/10.1080/14675986.2015.992199

Köylü, Z. (2016). *The influence of context on L2 development: The case of Turkish undergraduates at home and abroad* (Unpublished doctoral dissertation). University of South Florida.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. https://doi.org/10.2307/2529310

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. https://doi.org/10.1093/applin/amn018

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English: Verb-noun collocations in L2 writing. *Language Learning, 61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Leipzig Corpora Collection. (2011). *German news corpus based on material crawled in 2011*. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=deu_newscrawl_2011

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85–102. https://doi.org/10.1016/j.jslw.2009.02.001

Li, J., & Schmitt, N. (2010). The development of collocations use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23–46). Continuum.

Lindstromberg, S., Eyckmans, J., & Connabeer, R. (2016). A modified dictogloss for helping learners remember L2 academic English formulaic sequences for use in later writing. *English for Specific Purposes, 41*, 12–21. https://doi.org/10.1016/j.esp.2015.08.002

Llanes, À. (2019). Study abroad as a context for learning English as an international language: An exploratory study. In M. Howard (Ed.), *Study abroad, second language acquisition and interculturality* (pp. 136–154). Multilingual Matters. https://doi.org/10.21832/9781788924153

Llanes, À., Arnó, E., & Mancho-Barés, G. (2016). Erasmus students using English as a *lingua franca*: Does study abroad in a non-English-speaking country improve L2 English? *The Language Learning Journal, 44*(3), 292–303. https://doi.org/10.1080/09571736.2016.1198099

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science. *R package*. https://CRAN.R-project.org/package=sjPlot

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Performance: Assessment of regression models performance. *CRAN*. https://doi.org/10.5281/zenodo.3952174

Mitchell, R. (2015). The development of social relations during residence abroad. *Innovation in Language Learning and Teaching, 9*(1), 22–33. https://doi.org/10.1080/17501229.2014.995762

Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing Company. https://doi.org/10.1075/scl.14

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman.

Pérez-Vidal, C. (Ed.). (2014). *Language acquisition in study Abroad and formal instruction contexts* (Vol. 13). John Benjamins Publishing Company. https://doi.org/10.1075/aals.13

Pérez-Vidal, C., & Llanes, À. (2021). Linguistic effects of international student mobility in European perspective. In R. Mitchell & H. Tyne (Eds.), *Language and mobility: Study abroad in the contemporary European context* (pp. 22–33). Taylor & Francis.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113–138. https://doi.org/10.1177/1362168814568131

Quasthoff, U. (2011). *Wörterbuch der Kollokationen im Deutschen.* De Gruyter.

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Schielzeth, H., Dingemanse, N., Nakagawa, S., Westneat, D., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N., Garamszegi, L., & Araya-Ajoy, Y. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schmitt, N. (2012). Formulaic language and collocation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics.* Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0433

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429

Snoder, P., & Reynolds, B. L. (2019). How dictogloss can facilitate collocation learning in ELT. *ELT Journal, 73*(1), 41–50. https://doi.org/10.1093/elt/ccy024

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development'. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Szudarski, P. (2017). Learning and teaching L2 collocations: Insights from research. *TESL Canada Journal, 34*(3). https://doi.org/10.18806/tesl.v34i3.1280

Taguchi, N., Li, S., & Xiao, F. (2013). Production of formulaic expressions in L2 Chinese: A developmental investigation in a study abroad context. *Chinese as a Second Language Research, 2*(1), 23–58. https://doi.org/10.1515/caslar-2013-0021

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1). https://doi.org/10.1017/S0272263108080042

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 Intralexical knowledge. *Applied Linguistics, 32*(4), 430–449. https://doi.org/10.1093/applin/amr011

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition, 35*(3), 451–482. https://doi.org/10.1017/S0272263113000107

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge University Press.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Griet Boone is Assistant Professor and Lecturer of German at the University of Antwerp and at Ghent University (Belgium). She currently teaches German for specific purposes, German

as a foreign language, interpreting and translation. Her primary research areas include phraseology, foreign language learning and teaching, intercultural professional communication, and study abroad.

June Eyckmans is Associate Professor of Applied Linguistics at Ghent University (Belgium), where she teaches courses on second language acquisition and teaching English as a foreign language. Her research is focused on second language acquisition and learning, cognitive linguistic approaches to vocabulary learning, study abroad, individual differences in SLA and contextual language learning.

Parents in the Classroom: Translanguaging and Informal English Vocabulary Learning Among Newcomer Prekindergarten Students

Emma Chen and Debbie Pushor

Abstract This study examined benefits from the pedagogical practice of translanguaging in a prekindergarten classroom at Edgeview School (To protect their anonymity, pseudonyms have been used for the name(s) of the school, teacher, parents, and children in this study.) in Western Canada. Translanguaging considers bilinguals' language practices as the norm and centers newcomer children's and their families' linguistic knowledge. Newcomer parents were invited to the classroom and on the broader school landscape, bringing their home language alongside the English instructional language. Parent engagement in language learning provided a unique opportunity because parents possess the capacity to navigate both the school and home language. Further, extensive research demonstrates that when parents are engaged in their children's teaching and learning, children are more successful, academically and socially. By bringing in home language in an informal way throughout daily classroom activities, English language vocabulary development of the young newcomer children was enhanced. This narrative inquiry shed light on the importance of parent engagement to newcomer children's language development. By offering parents a place in children's learning, currently untapped teaching and learning resources were opened up to maximize young children's language learning outcomes.

Keywords Translanguaging · Parent engagement · Newcomer children · Vocabulary learning · Narrative inquiry

---

E. Chen (✉)
Western Washington University, Bellingham, WA, USA
e-mail: Emma.Chen@wwu.edu

D. Pushor
University of Saskatchewan, Saskatoon, SK, Canada
e-mail: debbie.pushor@usask.ca

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_3

1 An Inquiry into Translanguaging

Our narrative inquiry examined benefits that arise from the pedagogical practice of "translanguaging" (Baker, 2001, p. 281) in a prekindergarten classroom at Edgeview School, a diverse elementary school in an urban center in Western Canada with a student population approximately one third majority Canadian, one third Indigenous, and one third newcomer. Interested in how English language vocabulary development was supported through the informal use of home language throughout daily classroom activities, we inquired: How is the English language vocabulary development of young newcomer children enhanced?

1.1 Theoretical Approach

1.1.1 Supporting Newcomer Families Through a Translanguaging Lens

Translanguaging has offered researchers and educators a way to recognize and value language-minoritized students' linguistic expertise. As described by García and Wei (2014), translanguaging considers bilinguals' language practices as the norm and, therefore, taking up a translanguaging lens with newcomer families helps (re)center the linguistic knowledge of the children and parents in these families. We agree with Lin (2019) that translanguaging provides a "dynamic, distributed view of language, seeing language as embodied, emplaced, and ensembled in its physical and social environments" (p. 8). Building on this social view of language practice, in our inquiry, we layered young children's language learning with their parents' linguistic knowledge and experiences and took the translanguaging lens to an expanded and extended level of support. Newcomer parents were invited to join their children in the classroom and on the broader school landscape, bringing their home language alongside the English instructional language. Parent engagement in language learning provided a unique opportunity because parents possess the capacity to navigate both the school and home languages on certain levels.

1.1.2 Applying a Pedagogy of Walking Alongside to Parent Engagement

Parent engagement is founded in a philosophy and a pedagogy of "walking alongside" (Pushor, 2015b). Within such a pedagogical approach, both parents and teachers are seen to be holders and constructors of knowledge of children, teaching, and learning. Parent knowledge, "the particular knowledge held and used by someone who nurtures children in the complex act of childrearing and in the complex context of a home and family" (Pushor, 2015a, p. 15), is valued by the teacher and used alongside the teacher's own knowledge in ways that are complementary. While teachers have specialized knowledge about such things as curriculum, instructional strategies, and assessment, parents have specialized knowledge about their child, their heritage language, and their home learning and literacy context. Given that children move, continuously, in and out of their home and school contexts, it follows that when the influential people in children's lives work together, there is a stronger contribution to their learning and development (Feasey, 2017), in both dominant and heritage languages.

2 Methodological Approach

Our research methodology was narrative inquiry, the focus of which is "lived experience – that is, lives and how they are lived" (Clandinin & Connelly, 2000, p. xxii). Based in a Deweyian (1938) view of experience, narrative inquiry as methodology attends to criteria of experience that include continuity, interaction, and situation. Similar to our pedagogy of "walking alongside," narrative inquiry is "done in a situation of mutual trust, listening, and caring for the experience described by the other" (Clandinin & Connelly, 2000, p. 109). In our inquiry, then, as our participants told stories of their experiences with heritage language and second language learning, we were interested in the detailings of their past experiences, their present living, and their future imaginings. We wanted to capture their outward social experiences in both home and school contexts as well as their inward personal thoughts and reflections about their place and voice in their children's informal language learning. We were also attuned to how the particularity of their contexts—the prekindergarten classroom, Edgeview School, the apartment building in which the families live which happens to be home only to Bangladeshi families—have shaped their knowledge. We were curious to know how offering parents a place in their children's learning on the school landscape would open up currently untapped teaching and learning resources to the teacher and school and potentially enhance young children's language learning outcomes. It is our intention, as we share findings from our inquiry, that our narrative will have an "explanatory, invitational quality" (Clandinin & Connelly, 2000, p. 185) that will evoke the reader's own stories, connections, and processes of rethinking.

2.1 Participants

Narrative inquiry aims to present and reflect on individual experience in a deep and fine-grained way. In this research, we intended to provide detailed narrative accounts from which we would gain insights and learn about translanguaging from participants' particular experiences. To that end, we worked with the one teacher in prekindergarten—the children's and families point of entry into the formal school system, and three parents who were engaged in a Parent Mentor[1] program at the school. Parent Mentor is a program that enabled parents to support children's learning in classrooms alongside teachers, while simultaneously building relationships with each other and with the school. Three parents were selected for this inquiry as they all have at least one child who attends or attended the prekindergarten program. All parents and their children in the participating families are bilingual; all parents were available during the day to engage in their children's schooling; and all three parents expressed interest in participating in the study.

Bridget, the Prekindergarten teacher, had 10 years of teaching experience and had taught at Edgeview School for two years at the time of our research. She spoke to us about who she is as a teacher by stating:

> So, my overall philosophy is … play-based learning but it's all following the lead of the children into what their interests are. … What do they know when they come to me? They know their family … right? …They know their mom and dad, some extended family, and maybe some neighbours, right? But those are where their stories all start, is all with families. So, for me, a lot of my families – not all of them, but a lot of my families – are non-English speakers and come from a lot of diverse backgrounds. When I came to Edgeview, the way that I started teaching had to look a little bit different – … in a really good way – to meet the stories that those kids were bringing to school. Their stories of families looked really different from other communities that I've taught in. (Recorded conversation, July 2020)

Bridget noted that in her morning class of 16 students, four students were English speaking while in her afternoon class of 16 children, just one child spoke English while the rest spoke other languages as their first language.

The three parents with whom we engaged in this research were mothers who immigrated from Bangladesh and who spoke Bangla as their first language. They were all parents who were highly engaged with their children and with Edgeview School. Takshvi is the mother of two children. When they arrived at Edgeview School, her elder son was in Grade 2 and her younger son was on a waiting list for the Prekindergarten program. Mishita was the mother of one son when she arrived in Canada in 2012. She had a second son, just months after settling, and then her daughter was born in 2015. Currently, all her children are attending Edgeview School. When Aarvi settled in the Edgeview community, she really wanted to strengthen her English language and she was lonely at home. As a result, she began attending school with her only child, joining in her daughter's classroom activities, as well as taking part in adult Conversation Circles, working as a Parent Mentor in the school, and participating in the School Community Council meetings. Aarvi recently had a second child, and was home with her son during the time of our research.

We gathered field text for our narrative inquiry through recorded and transcribed conversations with each of our participants, through participant observation in the

[1] The Parent Mentor program, patterned after a program created by the Logan Square Neighborhood Association in Chicago (https://www.lsna.net), provided parents the opportunity to pursue a personal goal they each set for themselves while at the same time supporting a teacher and children in a classroom. A parent may choose to be a parent mentor to enhance their English language, to learn about the Canadian school system, to gain strategies to support their child's learning at home, for example. At the same time, children and teachers gain from the parents' linguistic and cultural knowledge, as well as the direct assistance they provide in classrooms.

prekindergarten classroom and the school, through home visits with the mothers, and through the collection of photographs and artifacts that were provided by the teacher or parents or created by the children and families. Our findings reflect our analysis and interpretation of this comprehensive and interwoven compilation of field text.

3 Findings

As a means of analysis, we engaged in multiple, individual readings of our field texts. We each pulled forward resonant threads and patterns that we then interpreted together as we discussed them at a deeper and more in-depth level. As we collaborated to compare and combine our analysis and interpretations, key findings clearly emerged. We took these findings back to our participants to ensure that we had captured their thoughts and feelings accurately and that they saw themselves in our representations.

3.1 Key Finding 1: A Welcoming Environment Is Essential for Translanguaging

3.1.1 Teacher as Learner

Often, as a result of their teacher education programs and understanding of what it means to be a professional educator, teachers take up a stance as knowers (Pushor, 2009, 2019). They feel expected to be, and to present to parents and others as, experts in all aspects of teaching and learning—curriculum, pedagogy, instructional strategies, and assessment (Pushor, 2009, 2019). Bridget, as an early years teacher with 10 years experience, had a solid repertoire of language and literacy practices and strong pedagogical beliefs underpinning her prekindergarten program. As she settled into Edgeview School, a community with greater diversity than her previous experiences and many more English language learners, she enthusiastically opened herself to a process of "dis/positioning" (Pushor, 2011). Learning from and with families and children, Bridget came to "un-know" (Vinz, 1997) the dominance of monolingual language use in her facilitation of children's language acquisition and to re-know it as a process of translanguaging, in an act of standing together with multilingual parents as co-constructors of their children's vocabulary and language development. Below, we make visible Bridget's actions to "un-know" and "re-know" as we detail her responses to home visits, to hearing parents' greetings each day, and through the way in which she facilitated and supported connections between children and between families.

Home Visits

As a teacher new to Edgeview and to such a multifacted community, Bridget embraced the diversity within her classroom and assumed a stance as a curious and enthused learner with the children and families. Immersing herself in learning from parents and children about their language, religion, and culture, she authentically bridged teaching and learning between home and school. One of Bridget's key practices to get to know and to build relationships of trust with families was to make home visits.

> I was always at their house, like I was over at people's houses almost every Friday.…
>
> I would always start my year off by setting up invitations[2] [in the classroom] about families because that's what [the children] know, and because I do home visits in Prek, [my educational assistant] and I started thinking about what we were seeing in homes. We would take pictures of our home visits and what we were learning about families in homes and then we would base our invitations off of that. (Recorded conversation, July 2020)

Bridget shared a fascinating story with us of her awakening to the way in which she and her educational assistant, Joanie, had set up invitations in the classroom focused on families. As white, Canadian women, they created an invitation reflective of how they live in their homes with their families. Amidst kitchen artifacts, they artfully set a small table with a tablecloth and dishes, with chairs surrounding the table. Given how central family was in the lives of all of the three- and four-year-olds with whom they worked, they were surprised there was little to no interaction by the children with this family-like setting.

After making a number of home visits, in which they were warmly welcomed into homes, served beautiful Arabic coffees and pastries, or hosted to shared meals with families, Bridget and Joanie returned to their classroom and reviewed their photos. What they suddenly realized was that the setting they had created in their classroom did not reflect at all the settings in the homes of the families they had visited. With excitement, they removed the table and chairs, put a large tablecloth on the floor and put the children's play dishes and artifacts on it. They were thrilled to see that immediately the children began to play in this setting, even taking off their shoes and lining them up beside the cloth, just as they did at home. Suddenly, the play emerged, as did the children's language. In a context familiar to them, the children were able to bring forward their lived experience, and to learn and grow together in a space that began with their knowing.

In just this same way, observing that parents were not lingering in the classroom, sitting at the larger table and chairs provided for them, Bridget and Joanie reflected on adult seating in the homes they had visited. They realized that parents sat on carpets on the floor with pillows, not because they did not have or could not afford furniture, but because it was a cultural choice. In response, they removed the legs from the large table in their classroom, and placed the tabletop on the floor with beautiful cushions surrounding it. Immediately, parents began to sit in the classroom, to stay with their children, in ways they had not previously. Once they were comfortable, parents brought their cultural, linguistic, and "parent knowledge" (Pushor, 2015a)

[2] Invitations, a term used in early learning, is deeply rooted in the Reggio Emilia philosophy of encouraging children to direct their own play by offering open-ended, meaningful resources for them to explore. An invitation is a set up of play materials in an aesthetic way that provokes children's exploration, discovery, and wonder.

into the classroom, engaging with the children and with the teachers, facilitating learning alongside them.

Greetings

With invitations created in the classroom reflective of diverse cultures and ways of being, and with parents and grandparents present who were encouraged to speak in their home languages as well as in English, the children's motivation to learn words in multiple languages created rich language-learning opportunities. The way in which individuals greeted each other became of great interest to Bridget and to the children.

> Even though I had families that were from Syria, Bangladesh, from different countries in Africa, from different areas, they were all Muslim so they had this connecting feature about their cultures, right? They all greeted each other the same way … . I thought, "Did Habib's dad just say that? Baber says that to me [too]: As-salamu alaykum, alaykum as-salamu."
>
> So, Baber was usually one that I went to a lot if I was unsure about things because he was really good at explaining things to me. They were very comfortable with us and laughed with us and they were really easy going. …Baber greeted a parent in the hallway when he was bringing his daughter to Prek one day. I was like, "Baber, can I ask you something? What are you saying? Is it okay for me to ask? …And it's, "May Peace be upon you," and then "And Peace be upon you too," when you say it back. And so I said, "How do you say that?" [So I practiced saying it], "As-salamu alaykum, alaykum as-salamu," and I was like, "Oh! I love this!" And Habib's family spoke no English at all, Mom and Dad had none. So I was like, "Oh, my gosh! I can say something to them!" (Recorded conversation, July 2020)

When Habib's mom joined the adult English language class at Edgeview School, with such excitement Bridget discussed and practiced the Arabic greeting with all of her young students. Each day, they walked down to the adult classroom and greeted Habib's mom with, "As-salamu alaykum," words she understood and that enabled her to respond in return.

As the children engaged across languages and cultural practices, in ways such as in the above example, they came to value the linguistic repertoire available to them and they felt confident in drawing upon the vocabulary they could access, whether it be in English or their home language. In this way, Bridget was encouraging languaging, "a process of using language to gain knowledge, to make sense, to articulate one's thoughts and to communicate about using language" (Li, 2018, p. 1224). She was also supporting translanguaging, as she "consider[ed] the language practices of bilinguals not as two autonomous language systems as has been traditionally the case, but as one linguistic repertoire with features that have been societally constructed as belonging to two separate languages" (García & Wei, 2014, p. 2). By engaging in diverse languages with them, Bridget was creating language-learning opportunities for children and families in which their linguistic repertoires were valued and celebrated.

3.1.2 Teacher as Bridge Builder

While Bridget valued her relationship with each of the parents and their families, and made space for their translanguaging in her classroom and on the school landscape, she also worked consciously to build relationships between and among children and families. Through creating a larger web of relationships and an expanded pool of linguistic knowledge, Bridget built and contributed to support structures that enhanced language and vocabulary development, for children and parents alike. In embracing their language, languaging, and translanguaging, Bridget honored their identity, simultaneously embracing their culture, religion, and ethnicity (Hurst & Mona, 2017).

Connecting Children Beyond the Classroom

As well as supporting one another as co-teachers in children's language development, both Bridget and the parents in our study continued to seek other ways to support children's language learning on the school landscape. The adults looked to other children in the school to play a role at times. At other times, they drew on their own linguistic repertoire.

In Bridget's instance, she drew on the language resources of older children at Edgeview through the establishment of a Care Partners program.

> I knew that they had siblings throughout the school. So, then I talked to the Grade 8 teacher and I was like, "Okay, I feel like your kids can help my kids. Can we make a connection here and have some time together?" So, what we started doing was we had our gym class together. I had the gym on Wednesday mornings for an hour and we booked our care partner time during that time. Our first day with the Grade 8s, I taught them how to talk to a three-year-old. [For example,] you have to get down and talk to them and play beside them and that kind of stuff. [The Grade 8s] got really good at it actually. … They would teach them games or teach them how to kick a ball, but we put them with the same language if we could. So, our Bangla kids we put with the older kids who spoke Bangla so that they could speak to them in their home language and explain things to them when they were with them. (Recorded conversation, July 2020)

Bridget continually taught from a stance in which she gave voice to children's heritage language and used it as a means from which to scaffold their ongoing language development. Operating within a broad language community, in which everyone was seen by her to be a knower and teacher, she drew on the rich community resources available to her. We see such a stance as one of "walking alongside" (Pushor, 2015b) parents, as noted earlier and as developed further in the following quote, but also one of walking alongside other family and community members as well.

> We engage parents in their children's learning because the parents' presence matters; it makes a difference to their children's achievement and attainment of other educational outcomes. We know we cannot do alone what we can do when we work with parents. Because we value parents' place in their children's lives, we seek to make them visible in their children's learning. (Pushor & the Parent Engagement Collaborative, 2013, p. 240)

Dis/positioning herself from a typical "schoolcentric" (Lawson, 2003) stance, where the school's agenda is central and served by both teachers and parents, we see in

Bridget's practice instead how she has taken up a "familycentric" (Pushor, 2015b) approach that "places families at the centre of [her] curriculum and pedagogy" (p. 248).

The three parents in our study, who all served as Parent Mentors at Edgeview School, worked from this familycentric stance too. While they had facility in their home language and in English, they often were called to draw on their linguistic repertoire, or that of other students in the school, when working with children from language groups other than their own. Mishita recounted a moment during a reading class when a Grade 6 student was called to support her sister in Grade 4. She also spoke of how teachers or parent mentors would frequently call on siblings or cousins when they were engaged in a conversation with a child in which they could not grasp the student's message or understand their need.

Working with a student from Africa, who spoke a language that was not prevalent in the school community, Mishita recounted how she strived to communicate with her using concrete objects and an oral, play-based approach.

> I am teach her the colour, and the shape, and they're matching. … Play with learn, that time. Then the shape – heart shape. And then I say, "Okay, this is a heart." And most the time, a heart inside, then she just thinking – and so, this is pink colour. The pink, say pink, that type, I'm just teaching her. And the number, also. One, two – like there's a two, Then she say, "Okay, this is the two." We're just talking to say everything. Just we are say everything, practice every day, yes. (Recorded conversation, August 2020)

As we talked with Mishita about why she did what she did, we learned that her approach drew on her parent knowledge and the strategies she used to support language learning in her home with her own children, as well as on the strategies she was observing, learning, and using in her work as a Parent Mentor. Like Bridget, she used the range of her linguistic knowledge and repertoire of strategies, in both heritage languages and in English, to scaffold children's language development.

Connecting Families Within the School Community

The parents in our study were very conscious, and appreciative, of the familycentric stance taken up by Bridget and felt that it was a stance taken up by other members of the school community too. The parents spoke of a school-wide experience called *Cooking for Cohesion*, in which families of one specific cultural group led an interactive cooking session on the school landscape but in an out of school time. Over a series of weeks and through several distinct sessions, parents, students, and family members from different ethnic backgrounds at Edgeview School were invited to prepare traditional ethnic dishes at the school. Those who watched learned cooking techniques and were introduced to spices, ingredients, and foods different from the ones they may have known and prepared in their homes. During and after the cooking, the families leading *Cooking for Cohesion* shared their personal experiences, stories, photos, and music about their homeland, culture, and religion, and stories of their journey to and settlement in Canada. Mishita shared her experience of *Cooking for Cohesion.*

> So we are coming from Bangladesh, so we three or four families, we are representing our country, our traditional foods. We bring the foods and some stories, history actually.

> And we are cooking, yeah. Everybody is some cooks. We are share how to wear [our traditional clothing], what we're cooking.
>
> Yeah, that was great. We are not believing the school wanting doing that type of thing. Because is our culture, we are doing normal at our home, but is big school and all together coming and is respect. Everybody teacher, all are respect our culture. That's why they invited us, and yeah, that is big. …Our culture is me, like this is one culture, this is me. And your culture is you. But all culture will stand up there, that is good, is awesome. (Recorded conversation, August 2020)

What we believe *Cooking for Cohesion* was designed to recognize and honor is that immigrant parents possess a wealth of linguistic knowledge that reflects their ideology, culture, social interactions, personal and collective identity, and beliefs and values (Khan, 2018). Such knowledge is unique, embedded, and strength-based. With familycentric approaches to language development such as that foregrounded through this example, we see how systematic and institutional opportunities for immigrant children to develop bilingual and biliterate skills ensure these skills flourish (Valdés, 2001). When parents are welcomed onto school landscapes, and when their knowledge and language is embraced, as valued and valuable in the teaching and learning of their children, students are more successful in maintaining their heritage language and in achieving successfully in school (Arriagada, 2005; Oh & Fuligni, 2007; Suarez, 2002, 2007).

3.2 Key Finding 2: Translanguaging Is Encouraged When Heritage Language Moves Between Home and School

3.2.1 Translanguaging in the Classroom

Co-Authoring Family Journals

When sharing with us the Family Journal project she conducted with the prekindergarten children in her classroom, excitement leapt out of Bridget's voice. She invited parents to write about their family stories in the journal at home, then she shared and celebrated each story the children brought to the classroom. The young ones were always excited to listen to the sometimes familiar and sometimes strange stories that happened in their friends' homes and to have theirs listened to in return. "Can you read it to everyone?" That was what the children would eagerly say to her, Bridget added with a big smile on her face, as if she was reliving those moments. In those moments, the semantic networks which are closely tied to the development of conceptual knowledge (Vygotsky, 1962) in the young minds of the three- and four-year-old learners expanded and extended. These semantic networks help facilitate new vocabulary learning and contribute to the reorganization and elaboration of previous conceptual information (Otto, 2014). Family journals created by the young children and their parents/grandparents carried the meanings of the conceptual knowledge and vocabulary in the most accessible, understandable, caring, and fun way.

"We are all from different places" (Bridget, Recorded conversation, July 2020). The idea of welcoming and celebrating all cultural, linguistic, religious, and geographical backgrounds in one classroom was how the project initiated in the first place. Keeping that idea in heart, Bridget embraced the translanguaging pedagogy (García & Wei, 2014) in the Family Journal project. García and Wei (2014) defined translanguaging as:

> an approach to the use of language, bilingualism and the education of bilinguals that considers the language practices of bilinguals not as two autonomous language systems as has been traditionally the case, but as one linguistic repertoire with features that have been societally constructed as belonging to two separate languages. (p. 2)

Seeing teaching and learning through a translanguaging lens means acknowledging bilingual children's language capacity and taking the language practices of bilinguals as the norm instead of the language of monolinguals (García, 2009b). These young preschoolers acquired substantial conceptual knowledge from birth to the point they entered school with their parents in their home language, even when they were seen as "English learners" with basic language abilities.

Supporting Home Language

Bridget told us that she knew how much it meant to the families to be able to keep their home language alive, given that most of the children in her classroom were "emergent bilinguals" (García, 2009b, p. 322). She understood the parents' concerns of losing their home language in the larger and dominating English world. "I promise the parents I won't ruin their language" (Bridget, Recorded conversation, July 2020). Instead, home languages were encouraged in the classroom and in the Family Journal project through a pedagogical lens of translanguaging, which posits that bilingual children have one linguistic repertoire from which they strategically select features of all the languages they know to communicate effectively (García, 2012).

The following story Bridget shared makes visible the strategical utilization of all languages in a bilingual child's linguistic repertoire:

> This little boy from Bangladesh would draw a picture in the journal and say, "Abbo." In days he would only write to "Abbo." I didn't know that word so I asked, "Oh, who's Abbo? Is that a friend?" "NO! ABBO!" And then another little girl, who also speaks Bangla and whose family was really close to his, whispered, "Abbo is Daddy." Then we all understood and started to use those words – Abbo for Daddy and Ammo for Mommy – in our class. We would try to use those words in how we were explaining things to the kids so they knew that link. (Recorded conversation, July 2020)

In this young boy's linguistic repertoire, "Abbo" was how he named his father in his everyday life, and how everyone referred to their fathers in his culture and in his home language. "Abbo" is the vocabulary label for the conceptual knowledge of "father" in his semantic networks, in which he formed a complex and sophisticated web of all members in his family and community in linguistic terms. Before the moment recalled by Bridget in the Family Journal project, his linguistic label for his father remained his one and only vocabulary—"Abbo." The second he heard the little girl's whisper, "Abbo is Daddy," his semantic network expanded, adding in a new word "Daddy" in English and the connection between the two words "Abbo" and "Daddy." Two words originated from different cultures far from each other, traveled across oceans, time, and generations, now met and emerged in the boy's linguistic repertoire. With the words, he also gathered all the wonderful memories of "Abbo" in his head and bridged them with "Daddy." That whisper also opened a door for other children in the classroom and offered an opportunity to them to collect new vocabulary and rebuild their semantic networks. All children in the classroom then had the capacity to translate this particular discourse and specific vocabulary, and then paraphrase in one language or another (García & Wei, 2014). It is a beneficial opportunity for bilingual children to actively contrast languages for vocabulary expansion (Laufer & Girsai, 2008), and gaining metalinguistic awareness which is strongly associated with foreign language learning enhancement (Rauch et al., 2012).

3.2.2 Translanguaging at Home

Navigating Daily Conversations

Translanguaging remains the most prevalent practice among bilingual students, as they flexibly draw from their rich linguistic repertoire both on and off the school landscape (García, 2009a). One of the Bangladeshi moms, Takshvi, shared with us her children's language practice in daily conversations:

> So, in our home, we mostly use Bangla, in our regular conversations. But I see that when my kids talk to each other, they use English most of the time and, with friends, they do the same thing. (Recorded conversation, January 2021)

Bilingual children navigate their linguistic repertoire in a natural, relational, and contextual way. depending on who they make conversation with, when they use the languages, and where the oral interactions happen (e.g., Chen, 2021). In the scenario described by Takshvi, her children observed the language environment of the conversation and made a judgement call based on the language used and the people who participated in the conversation. They were well aware which language resource would be best applied in the current conversational situation—Bangla with parents and English with siblings and friends. As they "consciously experience[d]" (Polanyi, 1958, p. 195) the switch of languages and made meaning during the transitions, their vocabulary in English and Bangla, as well as their ability to understand and to express themselves in both languages, expanded as they crossed the boundaries of the languages each time. What we see is that translanguaging affords these young children the opportunity to engage in flexible and informal pedagogical practices to make sense of their academic, linguistic, and social experiences, thereby strengthening their multilingual identities and enhancing their language abilities within their linguistic repertoire as a whole (Creese & Blackledge, 2010; Lewis et al., 2012).

Supporting Homework in All Languages

Translanguaging is not exclusively used by the young ones. Being encouraged by the teacher to keep their home language active, parents felt more confident to use all of their linguistic repertoire to support their children's schoolwork. Takshvi told us that when she had to teach her children something related to school, she usually preferred to speak in English. However, sometimes when her children did not fully understand her, she would explain again in Bangla.

Mishita shared a similar story of supporting her children's homework in both languages. She noted that math is usually discussed in English given the consistency of naming numbers in English and Bangla, but difficult concepts and confusing explanations invited more rounds of conversations in their home language. "Sometimes if I say something in Bangla that they don't understand," Mishita added, "I would search on Google and show them" (Recorded conversation, February 2021).

Both parents' stories demonstrate that informal language learning happens on a day-to-day basis off the school landscape. Every day at home, after school is dismissed, vocabulary learning continues in an intimate setting. When bilingual children and their parents work on homework together, the learning moves beyond their comprehension of the subject content. Languages always play a significant role in the learning process. The back-and-forth exchange of ideas and perspectives in English and their home language constantly deepen the students' understanding of the English vocabulary acquired in such a process. The translanguaging lens, consciously and unconsciously adopted by the student and their parents, offers a continuum of vocabulary learning in an informal and incidental way at home.

3.3 Key Finding 3: Learning Is Organic and Two-Way Between Parents and Children

3.3.1 Increased Communication

Parents reported that their children were not the only ones who benefited from the movement of languages between home and school. All three parents acknowledged that their improved English capacity, as a result of their engagement in the Parent Mentor program, had a positive impact on their communication with their children, both in English and in their home language. Aarvi told us that she could speak more with her children in English now that she had the vocabulary and oral communication skills and, in turn, their Bangla conversations increased because both parents and children were empowered to take advantage of all their vocabulary in English and Bangla, and transit back and forth freely (Recorded conversation, February 2021).

The language practice at home became an organic two-way learning experience. The unique circumstance in these newcomer families broke down the typical hierarchy of teaching and learning roles between parents and young children. Instead, parents and children both shouldered the responsibility of teaching and learning regarding language advancement in a natural and fluid manner.

3.3.2 Learning Through Daily Routines

It was amazingly surprising for Mishita when she and her children experienced their first winter term in Edgeview School in Western Canada. Originally from Bangladesh, a country on the equator with an average temperature of 25 degree Celsius all year round, the winter outfits the school required were confusing and overwhelming for the children. It took quite a while for her family to master the winter routine, even with help from the teacher. Everything was new. For example, where to change your shoes, where to hang your winter jacket and snow pants, how to wear your mittens, scarf and toque were all puzzling (Recorded conversation, August 2020). "What is a toque?" Mishita chuckled, reflecting on her confusion years back. No one told them that, in Canada, the hat on your head is no longer called a hat but a toque. Everyone else seemed to know this "secret shared language" (Recorded conversation, August 2020). It took many days of coming to the school landscape, observing other families' practice, talking to the teacher and other parents, and many other efforts to crack the code like this one. It was during this very hands-on, physically present, and interactive learning that her newcomer children and Mishita learned practical ways to understand and use such contextual vocabulary. The daily routines of snow boots, winter jackets, toques, and mittens sparked numerous conversations regarding such things as seasonal wear, geographical knowledge, and cultural differences. Mishita's example demonstrates how languages and vocabularies keep growing organically while flowing between home and school.

3.3.3 Engaging the Whole Family in Stories

Family story time in these bilingual families presents different challenges and opportunities than in monolingual families. "I am just afraid that my kids learn something that I couldn't. Then I will face a terrible problem" (Takshvi, Recorded conversation, January 2021). In her family story time with her son, Takshvi insisted on reading English picture books so she could also learn vocabulary with her child. The desire to keep the intimate connection and maintain an effective communication motivated her. Bedtime stories are both quality time and learning time for Takshvi. "I really enjoy reading the storybooks for them because I learn lots of words" (Recorded conversation, January 2021). She also proudly told us that in their reading time she applied the techniques of how to read picture books to/with young children that she learned during the Parent Mentor program, such as intonation variations, pointing to characters or words, and the encouragement of questions and conversations. In those moments of reading the English picture books to her son, Takshvi played two roles simultaneously—a parent who teaches her child a new language, and a student who learns new vocabulary in teaching. One story, two learners, no formal instruction.

Aarvi shared her love for children's books too. Book reading has always been a beloved mother-daughter activity in her family. "Children's books have a lot of pictures and a lot of fun things" (Recorded conversation, February 2021). Every day before bed, Aarvi and her young daughter engaged in a world of children's books, having “small conversations” in “small sentences” (Recorded conversation, February 2021). Now, as Aarvi became a beginner English reader, many times it was her daughter being the reader and teacher, telling the story in the book and passing on the vocabulary she learned in the classroom to her mom and younger brother. “It never felt like learning. But we learned, together” (Recorded conversation, February 2021). Real-life storytelling was another warmly appealing language activity in Aarvi’s family. In their family stories, mom and dad were a Queen and a King, the young ones were a Princess and Prince. Many familiar, unfamiliar, ordinary, and novel words, phrases, and sentences were created, exchanged, challenged, improved, rebuilt, and innovated in the stories happening in this Kingdom full of love, wonder, and curiosity. When asked which language the stories were told in, Aarvi said it depended on the situation. If the children came and started a story in English, it proceeded in English. If the children were curious about Bangladesh and their family history, they started a story in Bangla. At other times, stories were shared in a translanguaging style, mobilizing the linguistic repertoire of each and every member of the family.

4 Closing Thoughts

In our narrative inquiry, what was made visible is how, through a promotion of translanguaging practices, different language features that were historically and culturally separate become one new whole in speakers’ linguistic repertoires (García & Wei, 2014). As newcomer children enter Prekindergarten classrooms, they bring their diverse cultural backgrounds and transnational identities into language learning, and open space for the teacher to embrace the notion of translanguaging in English vocabulary teaching and learning. Such a notion requires a dis/positioning of the teacher—to the place of learner, to the place of co-teacher alongside immigrant parents, and to the place of immersion in a multiplicity of languages, cultures, identities. Within the concept of translanguaging it is accepted that the language practices of bilingual people are the norm. For teachers who are monolingual and who have experienced teaching language in typical and historical ways, this concept can require a challenging shift of mindset. Therefore, what comes first for a teacher in implementing a translanguaging lens in language teaching and learning, as we saw with Bridget, is surrendering one’s position as the knower and being open to learning from and with the students and their families. Translanguaging is concerned with identifying and assessing the creation and use of original and complex interrelated discourses that do not easily fit into definitions of discrete languages but are part of the complete language repertoire of the bilingual students (García & Wei, 2014). The complexity embedded in this pedagogy affords newcomer children opportunities to build strong semantic networks by drawing upon their full linguistic capacity in all their languages. English vocabulary learning, therefore, is intertwined with their continued language development and the sustainability of their home language, as well as the strengthening of identity, culture, and religion. The space for learning language is extended and expanded, beyond classrooms, to home and community settings, inviting parents, grandparents, and other family members to co-teach and co-learn. When teachers engage parents in language teaching and learning, young children have more opportunities to experience authentic, multigenerational teaching and learning in the context of school, family, and community. Parents, in the meantime, gain English language skills and a sense of efficacy in their ability to support their children as learners. Families have the potential to strengthen their relationships, in turn, as they nurture young students' vocabulary development.

Through the parents' and the teacher's stories of experience, this study has made visible the value of translanguaging pedagogy in newcomer children's vocabulary learning, in and beyond classroom settings. We believe the examples presented in this chapter invite a continued development of teaching practices centered in the strengths and capacity of bilingual students, their parents, families, and communities. We also believe they demonstrate how the engagement of parents in language teaching and learning, as they bring their linguistic and parent knowledge to bear, enriches the vocabulary learning of young children from diverse backgrounds in the classroom, at home, and in the wild of life.

References

Arriagada, P. A. (2005). Family context and Spanish-language use: A study of Latino children in the Unites States. *Social Science Quarterly, 86*(3), 599–619. https://doi.org/10.1111/j.0038-4941.2005.00320.x

Baker, C. (2001). *Foundations of bilingual education and bilingualism* (3rd ed.). Multilingual Matters.

Chen, E. (2021). The pedagogical practices of an immigrant parent: Maintaining heritage language in the home context. *LEARNing Landscapes, 14*(1), 29–43. https://doi.org/10.36510/learnland.v14i1.1026

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research.* Jossey-Bass Publishers.

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal, 94*(1), 103–115. https://doi.org/10.1111/j.1540-4781.2009.00986.x

Dewey, J. (1938). *Experience and education.* Collier Books.

Feasey, S. M. (2017). *Schools, family and community: Overlapping spheres of influence.* The road less travelled by. https://theroadlesstravelledby.com/2017/09/28/schools-families-and-community-overlapping-spheres-of-influence/

García, O. (2009a). *Bilingual education in the 21st century: A global perspective.* Wiley-Blackwell.

García, O. (2009b). Emergent bilinguals and TESOL: What's in a name? *TESOL Quarterly, 43*(2), 322–326. https://doi.org/10.1002/j.1545-7249.2009.tb00172.x

García, O. (2012). Theorizing translanguaging for educators. In C. Celic & K. Seltzer (Eds.), *Translanguaging: A CUNY-NYSIEB guide for educators* (pp. 1–6). The City University of New York.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education.* Palgrave Macmillan UK.

Hurst, E., & Mona, M. (2017). "Translanguaging" as a socially just pedagogy. *Education as Change, 21*(2), 126–148. https://doi.org/10.17159/1947-9417/2017/2015

Khan, M. (2018). *O Canada, whose home and native land? An autobiographical narrative inquiry into the critical role of curriculum in identity affirmation* (Unpublished doctoral dissertation). University of Saskatchewan.

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. https://doi.org/10.1093/applin/amn018

Lawson, M. A. (2003). School-family relations in context: Parent and teacher perceptions of parent involvement. *Urban Education, 38*(1), 77–133. https://doi.org/10.1177/0042085902238687

Lewis, G., Jones, B., & Baker, C. (2012). Translanguaging: Developing its conceptualisation and contextualisation. *Educational Research and Evaluation, 18*(7), 655–670. https://doi.org/10.1080/13803611.2012.718490

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amx039

Lin, A. (2019). Theories of trans/languaging and trans-semiotizing: Implications for content-based education classrooms. *International Journal of Bilingual Education and Bilingualism, 22*(1), 5–16. https://doi.org/10.1080/13670050.2018.1515175

Oh, J. S., & Fuligni, A. J. (2007). *The role of heritage language development in the adjustment of adolescents from immigrant backgrounds.* UCLA CCPR Population Working Papers. http://papers.ccpr.ucla.edu/index.php/pwp/article/view/977

Otto, B. (2014). *Language development in early childhood education.* Pearson.

Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. University of Chicago Press.

Pushor, D. (2009). The situation of parents in the curricular commonplaces: A place of equal rank? *LEARNing Landscapes, 2*(2), 139–154. https://doi.org/10.36510/learnland.v2i2.300

Pushor, D. (2011). Attending to milieu: Living a curriculum of parents alongside teacher candidates. In J. Kitchen, D. Ciuffetelli Parker, & D. Pushor (Eds.), *Narrative inquiries into curriculum making in teacher education* (pp. 217–237). Emerald Group Publishing Limited.

Pushor, D. (2015a). Conceptualizing parent knowledge. In D. Pushor & the Parent Engagement Collaborative II, *Living as mapmakers: Charting a course with children guided by parent knowledge* (pp. 7–19). Sense Publishers.

Pushor, D. (2015b). Walking alongside: A pedagogy of working with parents and family in Canada. In *International teacher education: Promising pedagogies* (pp. 233–253) [*International teacher education: Promising pedagogies, Part B* (C. J. Craig & L. Orland-Barak, Eds., pp. 233–253)]. Emerald Group Publishing Ltd. https://doi.org/10.1108/S1479-368720150000025008

Pushor, D. (2019). Using parent knowledge to enhance teaching and learning experiences in schools for children and youth. In T. A. Turner-Vorbeck & S. Sheldon (Eds.), *Handbook of family, school, community partnerships in education* (pp. 243–263). Wiley.

Pushor, D., & the Parent Engagement Collaborative. (2013). Feeling it in our hearts. In D. Pushor & the Parent Engagement Collaborative, *Portals of promise: Transforming beliefs and practices through a curriculum of parents* (pp. 235–243). Sense Publishers.

Rauch, D. P., Naumann, J., & Jude, N. (2012). Metalinguistic awareness mediates effects of full biliteracy on third-language reading proficiency in Turkish-German bilinguals. *International Journal of Bilingualism, 16*(4), 402–418. https://doi.org/10.1177/1367006911425819

Suarez, D. (2002). The paradox of linguistic hegemony and the maintenance of Spanish as a heritage language in the United States. *Journal of Multilingual and Multicultural Development, 23*(6), 512–530. https://doi.org/10.1080/01434630208666483

Suarez, D. (2007). Second and third generation heritage language speakers: HL scholarship's relevance to the research needs and future directions of TESOL. *Heritage Language Journal, 5*(1), 27–49. https://www.international.ucla.edu/cnes/article/72420

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Kreeft Peyton, D. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37–80). Center for Applied Linguistics.

Vinz, R. (1997). Capturing a moving form: 'Becoming' as teachers. *English Education, 29*(2), 137–146.

Vygotsky, L. S. (1962). Thought and word. In L. Vygotsky, E. Hanfmann, & G. Vakar (Eds.), *Studies in communication: Thought and language* (pp. 119–153). MIT Press. https://doi.org/10.1037/11193-007

Emma Chen is an Assistant Professor of Elementary Education with an emphasis on Children's and Young Adult Literature at Woodring College of Education, Western Washington University. Emma's research, grounded in narrative inquiry, explores transnational parent knowledge in heritage language education in home and community settings, aiming to center multilingual families' language practices in language teaching and learning.

Debbie Pushor, Ph.D. Professor in the Department of Curriculum Studies, University of Saskatchewan, Canada, has led narrative inquiries into parent engagement and leadership, a curriculum of parents, parent knowledge, and systematic parent engagement. In teacher education courses, Debbie foregrounds walking alongside parents in the education and schooling of children.

Incidental Vocabulary Learning in a Content and Language Integrated Learning Setting

Ching-Wen Wang

Abstract This research explores whether Content and Language Integrated Learning (CLIL) fosters stronger second language (L2) incidental vocabulary learning than does explicit vocabulary teaching. The study involved 27 students with a Common European Framework of Reference for Languages proficiency level of A2, a control group of 13 taught explicitly and an experimental group of 14 taught implicitly using CLIL. To examine their vocabulary knowledge, the research engaged pre- and posttests that involved translating 59 words from English to Chinese. The words—all used in the context of workplace English—were drawn from multiple sources. The results were significant, showing the experimental group's incidental learning of vocabulary exceeded that of the control group. The principal reason was likely the design of the CLIL tasks, which required students to conduct preparatory research that exposed them to a broader range of related vocabulary. Also, working in small groups fostered communication through social mediation and imaginary play, which involved the permanent presence of the L2. However, the results were limited to only one relatively narrow content setting, the workplace. Further experiments in multiple content settings would help determine the value of CLIL across a breadth of practical disciplines.

Keywords CLIL · Incidental learning · Vocabulary · Workplace English

1 Incidental Vocabulary Learning

The techniques for teaching foreign language vocabulary are numerous and wide ranging. The field is in a constant state of exploration and refinement. Recent studies include research into the effects student collaboration on English as a second language (ESL) vocabulary learning (Ariffin, 2021), methods of vocabulary study in the context of the Covid-19 pandemic (Tahir et al., 2021), and the impact of social media on learning second language (L2) vocabulary (Nguyen, 2021).

C.-W. Wang (✉)
The Language Center, Chaoyang University of Technology, Taichung City, Taiwan
e-mail: feliciawang@cyut.edu.tw

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_4

With respect to the explicit teaching of vocabulary in the ESL classroom, Tahir et al. (2021) offers a strong example of the descriptive style of vocabulary teaching. Although the present research and the Tahir study share some research procedures, such as using a pre/post-test, the content of this study was substantively different in that it does not represent explicit vocabulary teaching in the context of a traditional L2 classroom.

Another interesting recent research study was conducted by Kaivanpanah et al. (2021). In that study, both the explicit and implicit approaches were explored in one experiment. That study showed using explicit teaching in combination with modified-implicit activities results in stronger outcomes than implicit approaches alone.

Less research has been conducted about teaching attitudes with respect to implicit approaches to teaching. However, Sun et al. (2022) have conducted an interesting study comparing teachers' attitudes toward implicit and explicit approaches. By developing an implicit association test, the research was able to find an objective method of exposing teacher bias and reveal that teachers do not always say what they think. The researchers concluded that the teachers' opposing attitudes were attributable to the teachers' personal, institutional, and social contexts.

The present research is concerned with using implicit vocabulary teaching and learning in the context of the CLIL classroom. Therefore, this orientation necessarily limits the review of recent research to the implicit arena of teaching and learning. That said, no matter whether the teaching approach is implicit or explicit, vocabulary learning is the entry and fundamental element to new language proficiency in reading, listening, speaking and writing (Li, 2015; Richards & Rodgers, 2001). Wilkins (1972) asserts that with respect to language one can convey many things without grammar but they cannot convey anything without vocabulary. A great body of investigative research exists about how vocabulary acquisition occurs and the quantity of vocabulary that L2 learners need to acquire for proficient communication. For example, Nation (2006) suggests that L2 learners need to acquire knowledge of around 6,000–7,000-word families for listening and 8,000–9,000 for reading. Although some research (Schmitt et al., 2015) has challenged aspects of Nation's conclusions, the recommendation stands that L2 teachers and learners must set and meet vocabulary goals for teaching and learning.

In the more traditional, highly structured L2 learning classrooms, students usually gain lexical knowledge through the teacher's explicit teaching of vocabulary. By this method, students learn at least the surface definition of words. However, the surface knowledge of a word may not hold the depth of meaning variability the word deserves. For example, a student may learn the meaning of the word "bore" but they will not understand the difference between saying "I'm bored" and "I'm boring." More in-depth word knowledge takes time because it involves exposure to words on a fuller linguistic spectrum, such as in reading and listening activities where they may encounter a new word incidentally alongside other related vocabulary—with which they are familiar-that helps define the new word in context.

Unfortunately, due to the rising student numbers in classes—upwards of 30 where there were previously upwards of 20—and limited class time (once a week for

100 min), along with other curricular attitudes that limit a student's out of class study time, the traditional method of explicit vocabulary instruction has become inefficient. Consequently, classes cannot adequately expand the students' exposure to words, which is necessary for them to meet the desired goal of in-depth word knowledge (Schmitt et al., 2015). Thus, research has been investigating alternative approaches for learners to acquire new words, and some studies suggest that incidental learning may present a valuable opportunity for learners to acquire word knowledge that explicit teaching can no longer always cover (Al-Homoud, 2019; Laufer, 2003).

Incidental vocabulary learning is a by-product of cognitive activities that involve other aspects of comprehension, such as reading (Al-Homoud, 2019), listening or word-focused activities, such as role plays and vocabulary games (Laufer, 2003; Ramos, 2015; Tang, 2020). However, other research shows that, in addition to reading, positive outcomes for learning vocabulary can also come incidentally through task-based approaches (Sarani & Sahebi, 2012) and game-based learning (Reynolds, 2017).

Therefore, as a method of promoting incidental vocabulary learning in the presently changing classroom environment, this study looks at the emerging innovative approach to L2 teaching called Content and Language Integrated Learning (CLIL) (Coyle et al., 2010). CLIL advances the idea of using a second language to learn subject content, which in turn enables the acquisition of L2 knowledge. Thus, with CLIL, the more a student uses a second language to learn about a particular subject, the more deeply they will also learn the second language. The latter expansion of language knowledge then heightens their ability to further study the subject, and so on in a quasi-hermeneutic circle of knowledge building. In fact, by coupling content with language, CLIL necessarily implies two related concepts: (1) That language learning is the linguistic interpretation of subject content and (2) That knowledge is the learner's belief in that interpretation.

CLIL provides an ideal dual-facet teaching formulation for its dual-focused—content-language—method of learning. One facet is the 4Cs framework of Communication, Content, Cognition and Culture; the companion facet is the CLIL Language Triptych: language *of* learning, language *for* learning, and language *through* learning. The CLIL method also structures its language learning objectives by engaging the cognitive demands of content together with both the novel use of previously learned L2 elements and the use of new yet-to-be-learned L2 elements (Léon-Henri, 2015). Although teachers in the CLIL setting make the interrelationship between content objectives and language objectives explicit (Coyle et al., 2010) the approach also necessarily leaves the door open to incidental learning. Thus, not all L2 learning with CLIL is incidental because it is often fundamental to the content. However, the explicit focus of vocabulary learning does not end with a word's denotative meaning but rather continues on to engage the connotative interrelationships between vocabulary and content. In addition, as this study will show, the relationships between words in the context of content can lead to the incidental broadening of a learner's lexical knowledge.

Given the content-language learning environment of CLIL, the approach possibly offers a solution to the problems of both large class size and the time limitations placed on traditional L2 classes. As such, CLIL units may also offer a teaching alternative that deepens the L2 experience for learners that they would not have in an otherwise traditional classroom setting.

Furthermore, by removing the students from the structured class environment, the CLIL approach places them in a self-generating learning structure relative to both the content and the L2. Thus, an element of language learning in the wild occupies every phase of the CLIL activity because the L2 is a paramount presence in the learners' experience of the content.

In order to discover how effective the CLIL approach is for vocabulary acquisition, the current study investigates whether incidental vocabulary learning within the 4C's framework and the Language Triptych can accomplish the learners' L2 vocabulary goals. Although this question anticipates a quantitative result, qualitative aspects of teaching can also motivate students' incidental learning. However, the focus here is on the former.

2 Language in the CLIL Context

In recent decades, L2 learning has moved in a number of directions, all leading away from traditional methodologies, such as the grammar-translation method, the focus of which is squarely on the L2. At the other end of the spectrum would be the application of project-based learning (PBL) to L2 instruction. Using the PBL methodology moves the focus off the L2 and onto an interesting, perhaps vital problem the students attempt to solve while using the L2. Somewhere near the middle of these two extremes lies theme based L2 learning, such as English for specific purposes (ESP), where the teaching focus remains primarily on the L2 but an environment—a specific purpose—serves as a kind of backdrop against which the L2 learning takes place. CLIL falls somewhat closer to PBL on the teaching approach spectrum because it involves the students in the actual study of some aspect (content) of a real subject. Thus, with CLIL, the more traditional L2 learning backdrop takes on a more complete reality and moves to share center stage with the goals of L2 learning. In effect, the CLIL students actively engage with the content in a simulation of real world situations. However, as the label makes clear, CLIL is a dual-focused language learning pedagogy, concentrating on the *Integration* of *Content* with the second *Language* in the *Learning* environment. However, while their engagement with the content involves a simulation, their L2 engagement is both real and immediate. As Marsh (2000) explains, this approach aims to help motivate learners to *pick up* languages in a more natural way.

Through carefully designed teaching activities (Meyer, 2013), CLIL learners use higher-order thinking skills in the target language to convey their thoughts about one or more aspects of a subject, which comprises the content of a lesson (Coyle et al., 2010). In other words, the content is not the L2, per se, but rather any aspect of a subject that the learner then uses the L2 to think, read, and/or converse about. At the same time, the CLIL teacher does explicitly teach some aspects of the L2; these serve as a learning scaffold, a language for learning, that the students use as points of entry into the content. The teacher will also introduce other bits of language/word knowledge that are central to the content and which, therefore, the learners are likely to encounter when completing their CLIL tasks. Both of these aspects of L2 information stimulate the power of recognition that is central to learning of all kinds (Gadamer et al., 1986).

3 The 4Cs Framework

The 4Cs Framework—Content, Communication, Cognition and Culture—not only integrates the focal elements of CLIL but also provides lesson-plan guidelines for teachers to meet their desired content and language learning goals (Coyle et al., 2010; Meyer, 2013). For example, Content refers to the range of themes, topics, and practical aspects of a subject selected for a CLIL project. Likewise, during CLIL learning activities, Communication takes the learner beyond grammar and lexical knowledge to a deeper linguistic engagement with the content, thereby developing their contextual understanding of the target language. As a result, Communication intensifies the language acquisition experience by building fluency through engagement with the fundamental content-related language precisely when learners need to use it. Cognition in CLIL energizes the learning process through activities that stimulate higher-order thinking skills, which in turn focus or concentrate the learner's use of language on understanding the content. Finally, Culture in the CLIL framework actually refers to the *cross*-cultural perspectives that learners acquire through the integration of new content and new language. In addition, this Cultural component of the 4Cs framework promotes personalized learning, that is, the learner's independent discovery of new word and content knowledge. Personalized learning not only increases the depth of the learner's understanding but also simultaneously and incidentally builds self-awareness. Thus, this latter subjective aspect of CLIL not only sets it qualitatively apart from the objective experience of explicit vocabulary learning but also may trigger incidental vocabulary building and acquisition.

3.1 The Language Triptych

One important element of the Communication component of the 4Cs framework is the Language Triptych, which specifies the process and goals of language acquisition central to its pedagogy (Coyle et al., 2010). The first component, the language *of* learning, refers to "the language for learners to access the basic concepts and skills relating to the subject theme or topic" (Coyle et al., 2010, p. 37). The language of learning is also the first stage of CLIL planning and activity; as such, it involves analyzing the language needed for planning a lesson, such as the key words, phrases, and grammatical functions the learners will encounter. That is, while the teacher may not explicitly teach all these elements, they do lay the groundwork for the learners' incidental encounters with them during the planned CLIL activity. The second component, language *for* learning, "focuses on the kind of language needed to operate in a foreign language environment" (Coyle et al., 2010, p. 37). At this stage, the teacher needs to consider the learners' previous L2 learning history to ensure the lesson plan is compatible with their levels of competency. Thus, language *for* learning is more functional, providing a kind of language scaffold that supports discussion, task demands, cognitive strategies, and other classroom talk perhaps not always directly connected to the content. The third component, language *through* learning, "is based on the principle that effective learning cannot take place without active involvement of language and thinking" (Coyle et al., 2010, p. 37). So, in this phase the language becomes the medium of learning both the language used and the topic of the language. Here the essence of integration in CLIL is most apparent, as the language operates hand-in-hand with the content to complete the CLIL activity. Thus, the Language Triptych serves as the connector for all the objectives persisting between the content and language. However, to help guide teachers in their application and management of the Language Triptych while planning their CLIL exercises, Coyle has conceived the 3As pragmatic tool (Coyle, 2005).

3.1.1 The 3As

The 3As refer to Analyze, Add, and Apply (or Assure). The planning tool operates in three stages that correspond to the Language Triptych. Thus, the first stage involves a systematic linguistic *analysis* of the content to define the language *of* learning. The purpose is "to identify key words (including specialized contextualized vocabulary) phrases, grammatical functions for concept formation and comprehension" (Coyle et al., 2010, p. 7). The second stage concerns the language *for* learning and begins to shift the focus from the teacher's planning to the learner's experience by *adding* "language experiences…which enable the learner to operate effectively in a CLIL setting" (Coyle et al., 2010, p. 7). These additions include the language that supports "learner strategies, classroom talk, discussion, task demands" (Coyle et al., 2010, p. 7) and so on, which comprises the learning scaffold central to the language *for* learning. Finally, stage three, the "application stage" makes the CLIL experience the most advantageous for incidental learning by *assuring* that the language *through* learning has both "cognitive and cultural capital" by extending the learners' "cognitive skills and cultural awareness" (Coyle et al., 2010, p. 7).

Considered together, the 4Cs framework and its component Language Triptych suggest a classroom setting that involves carefully designed activities that guide learners through an educational journey. However, since even the most broadly educated teachers may not have enough in-depth knowledge about the Content they choose for a CLIL activity, Meyer (2013) has created the CLIL-Pyramid as a useful and dependable organizational resource.

3.1.2 The CLIL-Pyramid

It includes templates and tools to help teachers plan and organize CLIL activities by identifying quality principles and corresponding strategies (Meyer, 2013, pp. 296–307). According to the CLIL-pyramid, a valid CLIL activity needs to address six principles:

1. rich input, which triggers the target language acquisition
2. scaffolded learning, which offers the students learning support in addition to instruction
3. rich interaction and pushed output, which creates the dynamic exchange of ideas that lead to strategic discussions, peer learning, and group action
4. the added (inter-)cultural dimension, which promotes cross cultural perspectives
5. make it HOT, that is, ensure that the activity engages the higher order thinking skills of cognition, principally analysis, synthesis and evaluation (Krathwohl, 2002)
6. sustainable learning, which means ensuring a healthy learning ecosystem in which knowledge is co-created and shared (*Sustainable Learning*, 2021)

4 Incidental Learning

The decision to look at CLIL in relation to incidental vocabulary acquisition was not entirely based on finding invigorating and motivating L2 teaching methods. In fact, key research on incidental learning shows its intrinsic relationship to CLIL. As Marswick and Watkins (2001) express it, "Informal and incidental learning take place wherever people have the need, motivation, and opportunity for learning" (p. 28). CLIL precisely creates these three values: need, motivation and opportunity. Marsick, along with Volpe (1999), also created a list of incidental learning characteristics (Marswick & Watkins, 2001) that closely align with the characteristics of CLIL, making them quite compatible pedagogical methodologies. Both CLIL and incidental learning integrate with daily routines; both internal and external jolts, while not highly conscious, can trigger CLIL and incidental learning; the inductive process of both random and chance learning engages the characteristics of action and reflection that CLIL and incidental learning have in common; and, finally, the social experience of learning with others is fundamental to both CLIL and incidental learning.

While the qualities of integration and subconscious learning may seem fairly commonly understood, these and the other characteristics on the list deserve further explanation. *Integration* in the context of this research adds the dimension of informal learning to the integration of content and language in CLIL. In a sense, by establishing integration as central to the methodology in the first place, CLIL unintentionally also opened the door to the integration of incidental learning, which in turn becomes a *less conscious* aspect of both the content and language learning experience.

The notion of *triggering* incidental learning here refers to sudden or surprise events (*jolts*) that shift the learner's thinking about their situation (Marswick & Watkins, 2001). Such triggers are bound to occur during CLIL activities because while they involve structured tasks they also engage the learner in improvisational encounters. In those situations, there is a *chance* that the student might be *jolted* into using the L2 in novel ways, which could incidentally add to their depth of word knowledge.

The somewhat esoteric notion of incidental learning being an *inductive process of reflection and action* is actually quite important. While the incorporation of reflection and action are central to CLIL activities, the characteristic idea here suggests that the *influence* of those activities, especially on higher order thinking, stimulates the learning process. It also highlights the principle that CLIL experiences are not structured by explicit learning directions or instructions but rather informal and less intentional learning paths. These paths need the learner to reflect and take subsequent action to really have an effect. CLIL activities complement this idea because, although for all practical purposes they end, the learner nevertheless actually reaches no fixed conclusions; that is, CLIL supports sustainable learning rather than fixed outcomes.

Finally, the CLIL learning experience is by nature not a solitary one. It always involves students in social learning situations, through which they *learn with and because of one another*. This value is personified in the present study by the division of participants into table groups or teams. However, this value also fosters incidental learning because the students induce one another to learn from each other. Therefore, the value of the social is incidental and not explicit.

This closer understanding of incidental learning reveals a hidden but no less valuable qualitative aspect of the CLIL methodology. Furthermore, it pertains not only to the language learning focus of CLIL but also equally to the content learning experience. In other words, the incidental tacitly triggers or intensifies the integration. However, most important of all is the correlation between incidental learning and HOT, which is central to CLIL. If cognitive activity is too structured or too taken up with explicit directed learning, the options for spontaneity diminish, leaving the incidental little room for influence. Consequently, this research, which addresses an unexplored aspect of CLIL, sought to explore whether CLIL tasks, as HOT non-explicit activities that allow for greater spontaneity, can stimulate learners' incidental vocabulary learning to a level beyond what the traditional time-limited explicit teaching of vocabulary is able to accomplish.

Thus, the research goals—which also account for the methodology used—were to determine, first, if CLIL activities might move the learners' knowledge of targeted vocabulary words beyond their surface meanings and toward their effective use in real situations, and, second, if those activities led learners toward an understanding of vocabulary contextually related to the target words.

5 Methodology

5.1 Research Design

Consequently, the quasi-experimental design of this study needed to support an exploration of the effects of a theme-based CLIL task on the learners' incidental vocabulary learning. Therefore, the study used a pre-/posttest experimental instruction design that involved two groups: (1) a control group of 13 learners, and (2) an experimental group of 14 learners. The control group received the theme-based instruction using a traditional (but nonetheless engaging) lecture and discussion approach, while the experimental group received the theme-based instruction using the CLIL activity-based approach.

5.2 Participants

The study recruited participants from two sophomore general education English classes sponsored by the Language Center at a private university of technology in central Taiwan. The Language Center's curriculum intentionally steers away from teaching English as a traditional L2 subject and instead uses the *Workplace* as the content environment. The English lessons are then taught within that context, rather than as L2 lessons.

The students in these classes had an English CEFR proficiency level of A2, which was determined by both the student's score on the Taiwan Technological and Vocational Education system examination and the university's Language Center testing program. However, a number of students in each class were unable to either join or complete the study, which limits the data for this research to only those students who fully participated. The time schedules for the two classes were on Monday mornings and Thursday mornings, respectively. The overall aim (backdrop) of the class as established by the university was to help learners understand workplace English through situations and communications prompted by various related learning activities. Thus, the participants were 27 sophomores in total (9 males, 18 females), with 13 in the control group and 14 in the experimental group. The participants took the pre-/posttests to assess their proficiency with vocabulary on the general topic of business event invitations, which included related telephone communications. A two-sample t-test compared the vocabulary knowledge between two groups. No statistically significant difference in accuracy of vocabulary knowledge on the pretest showed up between the control group ($M = 10.6$, $SD = 17.98$) and the experimental group ($M = 11.28$, $SD = 10.17$); $t(25) = -0.123$, $p = 0.903 > 0.05$), indicating that all the student participants' prior vocabulary knowledge was at roughly the same level. Therefore, the researcher randomly assigned one class to be the control group and the other class to be the experimental group. The main instructional difference between the groups was that the control group received explicit vocabulary, phrase, and grammar lessons using the traditional instructional approach of lecture, discussion, and class activity, while the experimental group received the same number of hours of instruction but using a CLIL-based instructional methodology that required them to search for learning resources with the support of a handout that accompanied the CLIL activity.

5.3 Instruments

As content in CLIL "can range from the delivery of elements taken directly from a statutory national curriculum to a project based on topical issues drawing together different aspects of the curriculum" (Coyle et al., 2010, p. 28), the current study engages students to workplace situations. The experiment used a list of 59 English vocabulary words and two- or three-word terms (such as "costume ball" and "year-end party") that included 10 target words (italic) explicitly taught to all the participants (see Table 2). The words and terms on the list were chosen from the selection of textbooks recommended by the Language Center for the course—and therefore commensurate with the students' English abilities—and other texts also determined by the Language Center to be concordant with the students' English level. In a few cases, the words and terms were adjusted to conform to daily usage according to the advice of a native speaker of English who is also a Language Center associate professor. All the words related directly or indirectly to the planning of business events and related telephone communications. A vocabulary test assessed the learners' understanding of the list contents both before and after their groups' learning activities. This assessment enabled a comparison of the vocabulary knowledge the participants had already acquired before the experiment began with that acquired as a result of the CLIL unit's activities.

In addition, the vocabulary test measured the productive as opposed to the receptive skills of the participants. As Nation (2013) explains, "productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form" (p. 47). Since the test required the participants to type out a translation of each item on the list from English into their native language, it was able to measure these productive skills.

The class learning platform provides settings for randomly assigning questions while preventing students from opening other browsers during the pre/post-tests. Two evaluators—the researcher and a research assistant (who was present throughout the participants' studies)—working independently, graded the translation answers, which the class learning platform coded. Each correct translation scored one point on the test, with the total points then converted into a percentage of 100. Most of the participants completed this data collection session within 30 of the allotted 50 min. All the participants' translations on the posttest showed a high level of reliability, $r = 0.979$, $n = 27$, $p < 0.001$ (Tables 1 and 2).

Table 1 Correlations of two evaluators

		Evaluator 1	Evaluator 2
Evaluator 1	Pearson Correlation	1.000	0.979[a]
	Sig. (2-tailed)	.	0.000
	N	27	27
Evaluator 2	Pearson Correlation	0.979[a]	1.000
	Sig. (2-tailed)	0.000	.
	N	27	27

[a]Correlation is significant at the 0.01 level (2-tailed)

Table 2 59 Vocabulary words

Noun:	*reservation*	*invitation*	*servers*	*waitperson*
entertainment	costume ball	meet-n-greet	dress code	wet bar
cocktails	barbeque	disco party	catering	speakerphone
transportation	buffet	etiquette	country club	venue
reception	ballroom	lounge	luncheon	enterprise
flyer	bulletin board	gathering	interaction	dance floor
seating arrangement	announcement	follow-up	anniversary	cc (carbon copy)
year-end party	cocktail waiter	guest list	drinks on the house	
Verb:	*respond*	*confirm*	*reply*	organize
contribute	transfer	celebrate	interrupt	get in touch
RSVP (repondez s'il vous plait)	touch base (with)	commemorate		
Other:	*informal*	formal	confidential	appropriate
on behalf of someone	FYI (for your information)	TBA (to be arranged)	ASAP (as soon as possible)	etc (et cetera)

5.4 Procedure

Instructional pedagogy The pedagogical objectives of CLIL include both teaching goals and learning outcomes, with the need for vocabulary building being fundamental to both of these divisions. This study explored whether, and the extent to which, the CLIL approach could facilitate not only the learning of the 10 target words on the list but also the incidental learning of the other 49 related vocabulary items on the list. The CLIL topic (content) used for the study was business event planning and related telephone communication. The teaching goal was to develop the students' general understanding of how to plan a business event and the etiquette of conducting negotiations related to the event with people on the telephone. Since the CLIL learning process heavily emphasizes the importance of achieving learning outcomes, the researcher hoped the students would not only learn the 10 target words but also incidentally grasp the meaning and use of a significant percentage of the other 49 vocabulary items by encountering and using them in a CLIL-designed topic-activity.

Although the two classes met for the entire semester of 18 weeks, the unit of instruction designed for this study took place over only three consecutive weeks, or three 100-min-long weekly sessions. The first and last sessions included the 30-min pretest and posttest, respectively. The control group met on Monday mornings and the experimental group met on Thursday mornings. Both groups studied the assigned textbook content, as presented by the teacher, using the textbook materials and supplemental PowerPoint slides.

Supplemental to the core lesson plans for both groups was a vocabulary learning warm-up game to help the students develop their vocabulary learning sensibilities. For the game, each table group or team made a set of flash cards containing generally unfamiliar vocabulary items drawn from the textbook. The students looked up the definition of each item and noted it down on the card. After familiarizing the students with the pronunciation of each item, the teacher randomly called out the words in either English or Chinese. The students on the different teams quickly located the corresponding card, held it up and called out the word in the language that the teacher did not speak. The first two or three teams to correctly identify the word received a point token.

The CLIL unit included activities that engaged language content consisting of short expressions, past tense grammar checking, dialogue practice relating to customer service and helping customers with problems, and listening and speaking lessons. A preliminary activity used a role play model that put the students in the context of restaurant staff. The main activity then centered around a telephone conversation that focused on planning a business year-end dinner party. Both activities used worksheets (see Appendix) that served as guides for students.

Both groups of participants worked in *table groups* as teams of three to four students. As learning incentives, the students were able to earn point tokens toward the final class scores for each person on the team. For example, students who contributed to the class dynamic by answering questions put to the class received point tokens for their entire team.

The teams in both groups also participated in role play practice dialogues that included a telephone conversation. The students used the flash cards and the telephone conversation worksheet as guides. The teacher adapted the control group's role play conversations from the textbook, but the teams then used the telephone conversation worksheet to adapt the conversation further to fit situations the team members had invented for their role plays. Each team in the control group participated only with one another in the role plays; so this activity did not involve role playing between the different teams.

The research goal for the Thursday morning experimental group class was to discover whether a CLIL approach would incidentally increase the participants' word knowledge across the breadth of the vocabulary list. Thus, unlike the control group, the students in the experimental group created their role plays entirely on their own, using only the worksheet as a guide. They did not adapt their role plays from the textbook or any other existing dialogue. This adjustment added the dimension of reality to the role play simulation and required the students to use the language at their disposal in novel ways while possibly incorporating related vocabulary into their dialogues.

In addition, the experimental group's role plays included the completion of an assigned task (drawn from a set of cards containing various tasks), such as leaving or taking a message, asking for someone's telephone number, or discovering some specific information. The role-play task cards also assigned the students to their specific roles in the role play. This task itself heightened the role play simulation's feeling of reality. It also offered the opportunity for triggering jolts that might inspire the learners' incidental and intuitive use of less immediately familiar L2 word knowledge. To the same ends, the experimental group's role plays also employed a team-to-team design that had members from one table group role play with members of another table group.

Therefore, the experimental group's sessions included instructional material to facilitate the completion of the CLIL learning task. These materials included the role-play task cards—which also assigned students to specific roles—a problem record form that helped the students imagine a workplace situation that would include the telephone conversation task, and a task record form that set goals for students relative to their tasks. This latter form also provided space for the students to record their reflections both during and once they had completed their role plays. Reflections recorded during the activity allowed for their contribution to the learners' action choices. Therefore, recording their reflections during the activity was valuable to the process, a concept that the teacher had to impress upon the students.

Test administration The study included a pretest and a posttest on the students' vocabulary knowledge necessary or helpful to planning business events and having event-related telephone conversations. The students took the pretest 30 min before the beginning of the first class session. They took the posttest 30 min before the conclusion of the final class session. Reassuring the participants that these tests were not formal assessments (that is, the tests had no impact on the students' final course grades) helped alleviate any stress a student might have had over not knowing the translation of a word on the test. However, for the purposes of collecting and interpreting statistical data, the researcher recorded the number of accurate answers for each student and converted it into a comparable percentage.

Control group instruction The participants in the control group learned about the topic of planning business events and related telephone conversations primarily through question-and-answer sessions conducted by the teacher in both Mandarin Chinese and English. The teacher first posed questions to introduce the topic, then, after discussion, the class followed-up by completing the textbook exercises. In addition, the participants received the telephone conversation worksheet to help them deal with the L2 problems they would encounter when practicing the textbook's telephone conversation with their team partners. Thus, the learners' cognitive engagement was limited to the structured framework of a traditional classroom, where the concentration was on language learning with the situational content functioning as a backdrop for the language learning activities.

Experimental group instruction For the experimental group, the teacher only introduced the topic of planning business events and related telephone conversations, the purpose of asking questions was to capture and focus their attention on the content. The material the control group learned from the textbook exercises was integrated into CLIL activity tasks for the experimental group.

Table 3 abstracts how the 4Cs framework operated for the experimental group, while Table 4 shows how the supplemental planning tool of the 3As worked pragmatically to support the Language Triptych. In practice, all these values are co-functional and interdependent, such that, for example, the content and the cognition are equal forces in the activity just as the language of learning and the stage of analysis work to support the application of the language through learning.

The principles listed in Table 3 are important concepts that differentiate between the traditional classroom structure and the CLIL activity. For example, rather than the content being some aspect of language, such as word knowledge, as in the control group, the principle of new knowledge and skills applies to business event planning and telephone communication. The principle of learning through interaction is different from the learning by lecture and instruction that applied to the control group. Rather than the learning and thinking process being about completing workbook exercises, vocabulary memorization and quizzes, the CLIL cognition principles involved the students in creative or simulated real life working activities carried out in the L2.

Table 3 CLIL teaching within the 4Cs framework

4Cs	Principles	Learning content
content	new knowledge & skills	business event planning and phone communication
communication	learning through interaction	inquiring the spelling, sound, meaning, usage of the words or phrases; exchanging information; reporting problems; sharing suggestions, etc.
cognition	learning & thinking processes	planning a business event; imaging a problem in a workplace; filling out additional/necessary information on the phone conversation worksheet to prepare for the task; negotiating and solving customers' problems on the phone
culture	intercultural understanding & global awareness	thinking for others' situation; understanding others' cultures

Table 4 The 3As stages and CLIL learning procedures

3As	Principle	Design	Procedure	Minutes
Analyze	Language ***of*** learning	1. Learn words about business events and situation/problems	1. The teacher checks students' pronunciation and knowledge of definitions, particularly of the 10 target words by calling out either Chinese or English terms that the students then raced against one another to correctly repeat and earn point tokens for their team	25
		2. Learn words/expressions in a phone conversation	*2. Warm-up*. The students looked at projected images of business events and were prompted to guess the purpose for the events, leading them to discuss the reason for holding the particular event and what planning for it might entail	15
		3. Learn past tense, future tense, subjective mood	*3. Sharing*. Students shared any personal experiences of events they may have attended, including what made them special events. The students could also have described like-events they had seen in movies or on TV	20
			4. The teacher casually corrects pronunciation and grammatical errors	10

(continued)

Table 4 (continued)

3As	Principle	Design	Procedure	Minutes
Add	Language ***for*** learning	1. Practice dialogue and short expressions from the textbook	1. Students listened to the dialogue in the textbook and took turns to role play. Then students highlighted useful expressions for the phone conversation	10
		2. Imagine/predict a problem in the workplace and fill out the record form	2. Students decided on an event and details for the near future; they imagined the possible preparation problems or problems of the invitees	40
		3. Provide a website for telephone communication resources	3. Students develop a phone conversation using the terms and expressions and note any additional terms on the worksheet as they revise and rehearse the invented conversation, adding sentences or expressions they might need for cross-team interaction	40
Apply	Language ***through*** learning	1. Role play with tasks on the cards	1. Students drew lots to decide their roles for the role play activity, such as a secretary, a manager, an invitee and so on. Students used LINE, a free communication software, to share information. Table group members selected roles, such as a talker, a note taker, or a staff member in the company for the assigned task	40
		2. Complete the telephone task form	2. After completing the task, students summarized the information and confirmed it within the group; before evaluating the task and completing reflections, they shared their experiences with the other table groups. Students could volunteer to share their reflections and thoughts on how to improve the conversation for next time	40

The Language Triptych is as important to CLIL teaching as is the 4Cs framework. Furthermore, given that the students participating in the research were all part of a L2 class, the Language Triptych becomes somewhat more central to the learning outcome than if the students were in, for example, a business management class being taught in English. In the latter case, the content would likely hold a stronger place in the outcome goals.

From the pragmatic perspective, the analysis of the language of learning is perhaps the most traditional aspect of teaching with CLIL. As with the control group, it involved introducing the students to word knowledge that would eventually function in the application stage of language through learning. Nevertheless, some qualitative elements of this aspect of the teaching process were unlike the traditionally structured classroom. Principally, this stage did not stress the language knowledge in terms of what the students needed to learn to pass a test. Rather, the teacher presented the information as fundamental tools the learners would need to fulfill the tasks associated with the activity. In addition, the second and third steps in the analytical stage of the procedure set the stage for the content learning environment, which shifted from the "classroom" to the "simulated working environment."

The second stage, which *adds* language to the lesson, establishes the language for learning, serves as learning scaffolding information, and prepares the students for the creative activity involved with the simulated business setting. It also jump-starts the CLIL activity by setting the students on a planning path for their eventual completion of the CLIL tasks, steps one and two in the procedures.

Finally, the third stage of application supports language *through* learning by setting in motion the CLIL activity and its attendant tasks, which the procedure section describes. During the first two procedural steps of this third *assurance* stage, the teacher takes on a supervisory rather than instructive role, moving from team to team and assisting without testing or projecting an attitude of checking-up or trying to discipline the structure of the students' learning process. This third stage is also the climax of the lesson design, and unlike the traditionally structured control group, for which the climax was something like a test, the learning climax for the CLIL experimental group comes at the moment of accomplishment, or the realization of the task. This moment represents the high point of both the activity itself and HOT involvement, when any incidental word knowledge could contribute to the successful completion of the lesson. The third procedural step also functions as a kind of resolution to the day's lesson, when the students (and the teacher) privately reflect upon the experience, noting not only what knowledge they may have gained but also how that knowledge might have changed their interpersonal and cultural perspectives. As part of this reflection, the students noted the new or novelly used English words and phrases they recalled encountering (reading or hearing) and/or using during the day's activity. These vocabulary items would then—either consciously or unconsciously—become part of the language for learning, the language scaffolding formulated to support the following week's CLIL activity. Thus, the reflection represents the culminating opportunity in the day's exercise for the learners' recognition of incidentally acquired word knowledge (Table 4).

5.5 *CLIL Approach Re-examination*

The teaching design made use of six quality principles (Meyer, 2013). In that context, Table 5 analyzes the business planning and phone conversation ideas.

5.6 *Data analysis*

The initiating hypothesis for this study was to see if the task-oriented CLIL teaching approach was better than the traditional classroom approach at significantly facilitating the chances for incidental vocabulary learning. To check the hypothesis, the study compared the two groups of participants by separately analyzing the test results from the pretest and the immediate posttest sessions via a two-way analysis of variance (ANOVA) and t-tests.

The accuracy rate of the theme-based vocabulary words (by percentage) was calculated for both tests to investigate whether it increased for either group after the learning activity, that is, on the posttest. The number of correct words out of 59 served as numerators of ratios with denominators being the sums of the participants' accurate vocabulary answers. For example, if a student scored 36 words correctly, the accuracy rate is then 61%.

6 Results

Table 6 shows the descriptive statistics for the two groups. A two-way ANOVA was performed to test the presence of interaction effects for the two groups and the testing sessions (pretest and posttest) in terms of the accuracy rate. Significant interaction effects for the groups and the testing sessions were observed, $F(1, 25) = 13.641$, $p = 0.001$. This significant interaction effect indicates a difference in the accuracy rate for the two groups in the different testing sessions. Thus, to further investigate this result, the researcher analyzed the simple main effect on the accuracy rate using an independent sample t-test, which compared the two testing session results of the control group and the experimental group. A significant difference was detected on the posttest ($t = -5.248, p = 0.000; d = 2.04$) but not for the pretest ($t = -0.123, p = 0.903, d = 0.04$). The effect size for this analysis ($d = 2.04$) exceeded Cohen's (1988) convention for a large effect ($d = 0.80$). The results suggest that the participants had a similar vocabulary knowledge of this topic at the starting point of the learning activity, but they had different learning outcomes. This difference suggests that the different approaches affected the learning outcomes. The experimental group, which received the CLIL approach, had a better incidental vocabulary learning outcome than the control group, which received the traditional approach.

Table 5 Six quality principles

Principle	Activity design	Assessment & output
1. rich input	1. understanding dialogue from the textbook 2. learning from the worksheet 3. using websites as additional resources	Students had an abundance of language support when designing the company problems for their tasks, including vocabulary, phrases, short expressions, and situations
2. scaffolded learning	1. learning from the phone conversation worksheet 2. completing the problem record form 3. completing the task record form 4. role play cards	With the worksheet and designed forms, it offered students lines of thought, notes and guidance to complete the tasks. The role play cards triggered the students' reflections. In addition, the cards functioned as a check-list during the activity
3. rich interaction and pushed output	1. language support from the phone conversation worksheet 2. role play	Using the phone conversation as a reminder for the role play task, students were able to immediately pick out proper expressions; each party needed to complete its own interaction task according to the role play cards
4. adding the (inter-) cultural dimension	1. group discussion 2. completing forms 3. role play	During preparation for the tasks, students conversed and accepted ideas about the forms and came to the final version of the imaginary situation. The role play task also required the students to use their ability to understand and help each other with their problems
5. make it HOT	1. filling in forms to predict problems and solutions 2. helping the other party solve their problems	Students organized business events by thinking through the sequence and predicting consequences. Students chose the most suitable expressions for the role play situations
6. sustainable learning	1. creating real life situations 2. allowing students think along the lines of thought outlined on the problem record form	Students used the worksheet or forms to set up a business event and look-up useful conversational expressions. After planning the activity and practicing the expressions, the learners may acquire the latter as speech memories for future use

Table 6 Descriptive statistics of the improvement rate between groups and testing sessions

Groups	*N*	Pretest		Posttest			
		M	*SD*	*M*	*SD*	*t*	*p* (2-tailed)
Control group	13	10.600	17.984	10.153	4.705	0.084	0.935
Experiment group	14	11.285	19.178	27.357	10.902	−4.662	0.000

To determine whether any significant differences existed within each group between the mean scores on the pretest and the posttest, a paired *t*-test looked into the vocabulary learning outcomes between the two different testing sessions. The control group's performance showed no significant differences between the two testing sessions ($t = 0.084$, $p = 0.935$); whereas, the experimental group's performance showed significant improvement ($t = -4.662, p = 0.000$). The results suggest that the control group had less exposure to vocabulary words for this topic than did the experimental group.

In sum, the results from the independent sample *t*-test and the paired *t*-test suggested that the accuracy of the incidental vocabulary knowledge in the experimental group outperformed the control group in the posttest. Thus, the CLIL approach may have accelerated the participants' incidental vocabulary learning chances and enhanced their memorization of the words.

7 Discussion

7.1 Effective Learning

The purpose of the current study was to investigate whether the application of CLIL as a teaching approach has any impact on the incidental learning of vocabulary when students are learning about a specific topic. The aim of the sophomore English classes established by the university is to provide a general understanding of workplace English and to promote communication skills for the workplace environment. As long as the L2 teachers keep within the broad limitations of this topic, they have the option of choosing their teaching approaches. While some may choose traditional L2 drills, studies show that learning through tasks presents similar or better outcomes (Reynolds, 2017; Sarani & Sahebi, 2012). The findings of this study show that the performance of the participants in the experimental group using the task-based CLIL approach was remarkably better than that of the participants in the more traditionally taught control group. Thus, the experimental group's significant improvement in incidental vocabulary learning ability must have been the result of using the CLIL teaching approach.

7.2 Cognition, Vocabulary and Memory

As previously indicated, cognition is principle among the 4Cs for any CLIL learning plan, with HOT being central to the students' understanding, planning, and organizing of the CLIL tasks. Cognitive psychology research suggests that the human brain processes information, including memory, dynamically on a construct of levels (Craik, 2020). It carries out the L2 learning process in much the same way, coding, storing, using, and eventually reproducing the L2 information. This L2 mental processing includes the manipulation or processing of new word knowledge to facilitate the learning of new content (Craik & Tulving, 1975). Thus, with each cycle of content and language integrated learning, L2 information processing results in the storing of vocabulary word knowledge in the learner's memory. Since no single encounter with a word presents its complete meaning and usage, it is likely that the storage of vocabulary knowledge continues to grow with each learning cycle.

Some studies suggest that the demands of a task govern the attention of the participants' carrying it out. Likewise, incidental learning requires some degree of conscious attention from learners (Huckin & Coady, 1999; Newton, 2013). However, this latter awareness is awakened by the learning task, rather than the other way around. In other words, with CLIL the learner does not go looking for a task in order to apply a list of vocabulary words; rather, the demands of the CLIL task lead them to encounters with new vocabulary. The necessary attention the learner pays to these new words in order to use them effectively while completing the task leads to the storage of the word in the learners' memory. The point is that incidental learning happens as the result of the need to know something, which is not the same thing as accidental or chance learning.

7.3 Imagery and Play

Another less frequently acknowledged aspect of cognition that CLIL calls into play is the imagination, or creative thinking, particularly during the rehearsal and execution of the CLIL activity.

Since this research and the CLIL activity associated with it were both designed in the context of a L2 learning environment, engagement with the content—the workplace or business practice—necessarily involved the imaginary activities of simulation and mental representation. If on the other hand this research and the CLIL activity associated with it had been conducted in the context of a business management class being taught in English, then it is possible that the activity could have involved a real world rather than simulated setting. In that case, creative thinking may not have played as strong a role in the execution of the CLIL activity.

Simulation and representation are imitative activities that engage creative thinking in ways that are complementary rather than inimical to critical thinking. As Aristotle has taught us, the first things people learn come to them through imitation. People take pleasure in imitative representations, and they enjoy images because they learn by looking at them (*Aristotle, 1448b5-21*). Thus, creative thinking is an aspect of HOT that CLIL not only engages but also possibly adds to its ability to inspire incidental learning. Equally important is the idea that creative imitative thinking also introduces a playful language learning element into the CLIL process. Caon (2020) theorizes that with playfulness in language learning "strong intrinsic motivational inducements become integrated with affective-emotive, cognitive and social aspects of the learner." Heras and Lasagabaster (2015) have also shown that CLIL has a positive effect on these motivational and affective inducements. In particular, their research showed "that the CLIL module had a similar positive effect on both male and female students' learning of the technical content-related vocabulary" (p. 70).

The influence of creative thinking and playful language learning on cognition in CLIL should not be underestimated. Every word and expression that a student encounters triggers HOT in the formation of concomitant mental imagery. This imagery does not only comprise mental pictures of things, but of each moment in a whole situation. As psychologist Jerome L. Singer (1973) expresses it, "there is a complex interaction between representation in the brain of material stored according to sensory modality and also material stored in terms of a summarizing verbal label" (p. 190) Thus, the imaginary situations that CLIL activities create incidentally engage L2 verbal associations that the learner uses to build their understanding of the content.

In addition, the interplay of mental imagery and verbal representation develops the learner's L2 mental lexicon, which includes not only an arrangement of words and their definitions but also something of the mental imagery that informed the learners' first encounter with this vocabulary knowledge.

Thus, the demands of the CLIL activity tasks governed the attention of the research participants. These tasks focused the participants' imaginations on new situations that in some cases incidentally involved the manipulation or processing of new word knowledge. Processing this new word knowledge helped them construct mental representations of those situations. The results of the posttest suggest that the experimental group had successfully stored some aspects of this new, incidentally acquired, vocabulary knowledge. However, to deepen the participants' acquisition of the new vocabulary knowledge will require further processing that involves using the new words in new CLIL tasks that engage both critical and creative HOT.

7.4 Benefits of Task-Oriented Learning

One of the factors improving incidental vocabulary learning among the participants in this study may have been that the tasks and accompanying worksheets required them to discover additional words to facilitate both their discussions and negotiations with one another. As Caon (2020) notes, this kind of "social mediation…places the student at the centre of the learning process as students are considered resources and origins of learning, actively involved in building their knowledge base" (p. 446).

In addition, the CLIL activities enabled incidental vocabulary learning to take place on an unconscious level because the students had to do some research to complete their tasks, which would naturally have exposed them to new words or known words used in a novel way. The same opportunity for incidental learning exists in the cultural aspect of the 4Cs framework because the cultural use of language, especially turns of phrase and idioms, exposes the learner to deeper aspects of word knowledge.

To complete the CLIL task, each table group had to come up with their own approach, which included imagining and then simulating the workplace context. However, in addition to their discussions and negotiations, it was essential for the learners to treat language as a key element in the actual execution of the task itself. Thus, in each phase of the activity, the learners were discovering and using words as speech acts, in which case the language itself plays a proactive role in the social mediation.

Although the CLIL tasks are pragmatic modes of learning, they are also teleological activities in which the process leads to the end goals. As Newton (2013) found in his study of incidental vocabulary learning, with communication tasks "attention is clearly related to purpose" (p. 17). Therefore, the more important a word is to accomplishing the goal, the more likely a learner will be to acquire it as the result of some task-oriented learning activity. As Newton (2013) found, "learners gave priority to what they knew or what the task required them to know and did not spend time on words that could be avoided (p. 19). Therefore, both when the learner discovers a word and how they use it during their performance of the task, play important roles in the depth of their acquisition of it.

The task-based activities of CLIL use the learning scaffolds of activity forms and worksheets that offer the participants hints and guidance for researching the language they will need. These learning scaffolds also reduce learning anxiety about falling into the unknown, much the way a building scaffold insures the safety of a worker. Consequently, these learning scaffolds focus the learners' attention on the immediate purpose of the language they are using for not only communicative meaning but also achieving their end goals. Thus, rather than using language drills and structured practice activities, CLIL tasks place the L2 within a structured thinking and learning framework. Through planning discussions, negotiations, and the actual execution of content and language related tasks, CLIL gives learners opportunities to utilize L2 words in the simulated context of the content.

7.5 Limitations

The findings of this study demonstrate the potential the CLIL approach has for the incidental learning of vocabulary. However, the study itself had limitations. First, even though the focus of the research was on using CLIL as a teaching approach, conducting the study as part of an ongoing course limited it to regular class hours and required learning topics. Therefore, the research did not allow for the collection of further retention data. In addition, the arrangement of the CLIL tasks had to coincide with the students' other related workplace topics.

Also, relative to the classroom situation, the students' schedules and commitments to their major courses do not allow them the time necessary to do much out of class homework. They find out of class project work particularly onerous. This situation means that to facilitate the experience of language learning in the wild, the CLIL content had to be something they could engage in within the normal schedule of their lives. Thus, learning about the workplace was very useful in the present context. That said, it would be useful for further research to choose content more challenging to the students' classroom comfort zone. In addition, engaging with the content in a real, rather than simulated environment would better meet the conceptual aims of learning in the wild without diminishing the imaginative element.

The research was also subject to a few other limitations. First, this study did not include a questionnaire to survey the participants' feedback on the CLIL approach. Such a survey would have been useful as a comparative tool to the participants' reflections.

Second, a greater number of participants in both groups would have increased the amount of usable data. Again, this limitation was the result of conducting the study within the context of an ongoing course.

Finally, one cannot really learn a language outside the context of some content. Most often, the content of language classes is superficial, because it is calculated to reinforce the L2 language learning process. In the present study, learning about the workplace was the content. But in fact, this choice was a contrivance of the existing course requirement that it should teach Workplace English. In the future, it would be advantageous to conduct the study in the context of a course in which English was not the focus of the class. For example, a journalism or a history class, or any of the STEM or humanities subjects would offer substantive content for L2 learning with CLIL. The difficulty is that L2 teachers are limited to teaching subject content about which they have acquired an intuitive understanding. In that case, perhaps it would be useful to train interested content teachers in CLIL and see if the results are even better.

Appendix

Role: Maître d'

The manager is in charge of making sure the restaurant runs smoothly. As a manager, you will need to give clear instructions to everyone involved and step-in to help if a customer has a complaint. (from: https://www.fluentu.com/blog/english/english-for-restaurant-staff/)

Task:

1. Make a menu for the restaurant.
2. Find out why the customers are here.
3. Find out what the customers like / don't like.
4. Sell customers the special meal ($2000 per person) on the menu for next time.

Useful phrases / sentences:

1. Please accept my apologies.
2. What seems to be the problem?
3. Please accept this dessert, on the house.

Words you would use:

Role: Waiter / Waitress

Waiters (servers) are some of the most visible employees of a restaurant. As a waiter, you'll see and speak to many customers every day. Good English skills are important! There's plenty of repetition in this position, which you'll learn quickly as you work. (from: https://www.fluentu.com/blog/english/english-for-restaurant-staff/)

Task:

1. Find out the customer's allegation(s)/allergies.
2. Find out the customer's diet.
3. Recommend an expensive bottle of wine or special drink on the menu.

Useful phrases / sentences:

1. My name is ____, and I'll be your waiter (server) today.
2. May I take your order?
3. Today's special is...
4. Can I get you something to drink?

Words you would use:

Role: customer A

(from: https://7esl.com/restaurant-english/#Restaurant_English_Ordering)

Task:

1. Get the right table.
2. Order food / drinks.
3. Make clear any special dietary requirements (vegan, Muslim...)
4. Sell your company's products to customer B.

Useful phrases / sentences:

1. We'd like a table for ___ people, please.
2. I'd like ____ for as a starter / an appetizer / the main course.
3. What are today's specials?
4. What do you recommend?
5. Does ___ contain beef / nuts / dairy products...?

Words you would use:

Role: customer B

(from: https://7esl.com/restaurant-english/#Restaurant_English_Ordering)

Task:

1. Deal with food problems.
2. Complain about food / waiter (server)...
3. Get the bill (check).
4. Bargain down the price of customer A's products to at least 90% off.

Useful phrases / sentences:

1. Excuse me, but I didn't order this.
2. Excuse me. Does this contain ___?
3. Can I get a ___ (size) ____ (noun), please?
4. Can I have the check, please?

Words you would use:

Role: observer

Words:

Maître d'	Waiter (Server)
Customer A	**Customer B**

Tasks for each role:

Maître d'	Waiter (Server)
Customer A	**Customer B**

Activity 1: Phone Conversation

phase	sentences you may use	sentences you want to use
greetings	Hi, this is 人 from 地 speaking. Do you have a moment to talk? How may I help you? / What can I do for you? Who's calling, please? I'm afraid I don't have time to talk right now.	
purpose	I'm calling about 事/名詞. The purpose of my call is to 動詞.	
look for someone	Can I speak to 人, please? May I speak with 人, please? I'd like to speak with 人.	
transfer a call	Could you please connect me to 人? Can you put me through to 人? Is 人 around? I'll put you on hold while we transfer your call. Hold on please. Let me check that for you.	
message	Would it be possible to leave a message for him/her? Can you let him/her know I called? Can I leave a message? Can I take a message? We're planning to V...	
problems	What would be a good time to call? Would you like to meet with our 職稱(人)? I'm afraid S + V. We need to have some wiggle room. We are hoping to V / that S + V. We have no idea. Did you receive my email? I put it out... (群組發信了。)	
solutions	Call me if there's any change. You can always change your mind. We can live with that. You can always change your mind down the road.	
farewells	Have a great/nice day. Is there anything else I can help you with? Be sure to contact me if you have any questions. Can I call you back?	
confirm	Sorry, I didn't catch what you just said. Can you please speak a little more slowly? Can you speak louder? Could you repeat that, please? Would you say that again?	

caller	
receiver	
information	
problem	
solution	
reflection	

caller	
receiver	
information	
problem	
solution	
reflection	

LINE ID: 1A 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 1A Ask for another person. Invite others to the company party. Invitation: date, time, bring some snacks, bring only one friend dress code: green shirt and red pants Get the person's email address.
LINE ID: 2A 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 2A Ask for another person. Invite others to the company party. Invitation: date, time, bring some snacks, bring only one friend Dress code: green shirt and red pants Get the person's email address.
LINE ID: 3A 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 3A Ask for another person. Invite others to the company party. Invitation: date, time, bring some snacks, bring only one friend Dress code: green shirt and red pants Get the person's email address.
LINE ID: 4A 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 4A Ask for another person. Invite others to the company party. Invitation: date, time, bring some snacks, bring only one friend Dress code: green shirt and red pants Get the person's email address.
LINE ID: 5A 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 5A Ask for another person. Invite others to the company party. Invitation: date, time, bring some snacks, bring only one friend Dress code: green shirt and red pants Get the person's email address.

LINE ID: 1B 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 1B Transfer a call to another person. Explain: your problem. Explain: your friend can't speak Chinese and ask for some assistance. Ask for transportation/pick-up service. Get the person's phone number.
LINE ID: 2B 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 2B Transfer a call to another person. Explain: your problem. Explain: your friend can't speak Chinese and ask for some assistance. Ask for transportation/pick-up service. Get the person's phone number.
LINE ID: 3B 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 3B Transfer a call to another person. Explain: your problem. Explain: your friend can't speak Chinese and ask for some assistance. Ask for transportation/pick-up service. Get the person's phone number.
LINE ID: 4B 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 4B Transfer a call to another person. Explain: your problem. Explain: your friend can't speak Chinese and ask for some assistance. Ask for transportation/pick-up service. Get the person's phone number.
LINE ID: 5B 組員: 組員: 組員: 組員: Android: Cube ACR	Task: 5B Transfer a call to another person. Explain: your problem. Explain: your friend can't speak Chinese and ask for some assistance. Ask for transportation/pick-up service. Get the person's phone number.

References

Al-Homoud, F. (2019). Explicit teaching and incidental learning of vocabulary: Where is the pendulum swinging to? *International Journal of English Linguistics, 9*(6), 445–453. https://doi.org/10.5539/ijel.v9n6p445

Ariffin, A. (2021). Effects of students collaboration on ESL learners' vocabulary development. *Asian Journal of University Education, 17*(1), 177–191.

Aristote. (1958). *Aristotle: On poetry and style.* (G. M. A. Grube, Trans.). Liberal Arts Press.

Caon, F. (2020). Motivation, pleasure and a playful methodology in language learning. *EL.LE, 9*(3), 437–460. https://doi.org/10.30687/elle/2280-6792/2020/03/006

Cohen, J. (1988). *Statistical power analysis* (2nd ed.). Erlbaum.

Coyle, D. (2005). *CLIL planning tools for teachers.* Academia.edu. https://www.academia.edu/23915456/CLIL_Planning_Tools_for_Teachers_4Cs_Curriculum_Guidance_3As_Lesson_Planning_Tool_Matrix_Audit_Tool_for_Tasks_and_Materials

Coyle, D., Hood, P., & Marsh, D. (2010). A window on CLIL. In D. Coyle (Ed.), *CLIL: Content and Language Integrated Learning* (pp. 1–12). Cambridge University Press.

Craik, F. I. M. (2020). Remembering: An activity of mind and brain. *Annual Review of Psychology, 71*(1), 1–24. https://doi.org/10.1146/annurev-psych-010419-051027

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*, 268–284.

Gadamer, H.-G., Walker, N., & Bernasconi, R. (1986). *The relevance of the beautiful: And other essays*. Cambridge University Press.

Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research, 19*(1), 70–88. https://doi.org/10.1177/1362168814541736

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *SSLA, 21*, 181–193.

Kaivanpanah, S., Akbarian, I. h., & Salimi, H. (2021). The effect of explicit, implicit, and modified-implicit instruction on EFL learners' vocabulary learning and retention. *Iranian Journal of Language Teaching Research, 9*(2), 129–146.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

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.

Léon-Henri, D. D. P. (2015). *CLIL in the Business English classroom: From language learning to the development of professional communication and metacognitive skills.* ELTWorldOnline.com (Special Issue on CLIL), 1–26.

Li, K. Z. (2015). A study of vocabulary knowledge and reading comprehension on EFL Chinese learners. *Studies in Literature and Language, 10*(1), 33–40. https://doi.org/10.3968/6263

Marsh, D. (2000). *Using languages to learn and learning to use languages.* Retrieved 1 October, 2017. http://archive.ecml.at/mtp2/clilmatrix/pdf/1uk.pdf

Marsick, V. J., & Volpe, M. (1999). The nature and need for informal learning. *Advances in Developing Human Resources*, *1*(3) 1–9. https://doi.org/10.1177/152342239900100302

Marswick, V. J., & Watkins, K. E. (2001). Informal and incidental learning. *New Directions for Adult and Continuing Education, 89*, 25–34.

Meyer, O. (2013). Introducing the CLIL-Pyramid: Key strategies and principles for CLIL planning and teaching. In M. Eisenmann & T. Summer (Eds.), *Basic Issues in EFL-Teaching and Learning* (2nd ed., pp. 295–313). Universitatsverlag Winter.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82.

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

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research, 17*(2), 1–24. https://doi.org/10.1177/1362168812460814

Nguyen, N. T. T. (2021). A review of the effects of media on foreign language vocabulary acquisition. *International Journal of TESOL & Education, 1*(1), 30–37.

Ramos, F. D. R. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *Profile, 17*(1), 157–166.

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.

Richards, J. C., & Rodgers, T. (2001). *Approaches and methods in language teaching*. Cambridge University Press.

Sarani, A., & Sahebi, L. F. (2012). The impact of task-based approach on vocabulary learning in ESP courses. *English Language Teaching, 5*(10), 118–128.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212–226. https://doi.org/10.1017/S0261444815000075

Singer, J. L. (1973). *The child's world of make-believe: Experimental studies of imaginative play*. Academic Press.

Sun, C., Wei, L., & Young, R. F. (2022). Measuring teacher cognition: Comparing Chinese EFL teachers' implicit and explicit attitudes toward English language teaching methods. *Language Teaching Research, 26*(3), 382–410. https://doi.org/10.1177/1362168820903010

Sustainable Learning. Center for Learning in Practice. (2021, April 19). https://clip.careyinstitute.org/sustainable-learning-framework/

Tahir, M. H. M., Albakri, I. S. M. A., Khalid, P. Z. M., Mokhtar, M. M., Zaini, M. F., & Yusoff, M. Z. N. M. (2021). A review of explicit vocabulary instructions for ESL learners during COVID-19 pandemic. *Malaysian Journal of Social Sciences and Humanities, 6*(8), 384–393.

Tang, Z. (2020). A review on studies into incidental vocabulary acquisition through different input. *English Language Teaching, 13*(6), 89–95.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

Ching-Wen Wang is currently a lecturer of English in the Language Center at Chaoyang University of Technology in Taiwan. Her research interests include pronunciation, phonological awareness, the use of technology in language learning, and second language acquisition. She is also a PhD candidate in teaching Chinese as a second language in the Department of Chinese as a Second Language at National Taiwan Normal University.

Vocabulary Learning from Digital Gaming

The Good Gaming (GG) List: Key Vocabulary in Videogames

Julian Heidt, Geoffrey G. Pinchbeck, and Michael P. H. Rodgers

Abstract Recent technological advancements have meant that the classroom is no longer necessarily the only source of input for foreign language learners with other sources such as videogames gaining attention. The potential for gaming as a valuable source of language input may lie in its popularity, as it has been shown to be one of the most common extracurricular activities amongst learners. However, little is presently known about the vocabulary in videogames (Schmitt, 2019) and how knowledge of certain words may facilitate language learning through playing games. To address this, a pedagogical word list for videogames was created using a multimethod design, utilizing both quantitative and qualitative approaches. A corpus of ten videogame scripts (5.7 million tokens) was compared to a reference corpus of scripted, spoken English, to identify words that are significantly more frequent in videogames. This resulted in a 484-word list of common game language vocabulary. To facilitate how these words might be taught and learned, a sample of 100 words in the list were then coded from an evaluative perspective for whether they served a diegetic or ludic function. The implications of knowledge of this word list for incidental vocabulary learning, comprehension, and pedagogy are discussed.

1 Introduction

Playing videogames is often the object of derision as a time-wasting activity, and this negative association has likely been a barrier to its legitimization as a language learning strategy in formal educational settings. It should be noted, however, that similar pejorative sentiments have previously been associated with watching

J. Heidt · G. G. Pinchbeck · M. P. H. Rodgers (✉)
Carleton University, Ottawa, ON, Canada
e-mail: Michael.Rodgers@carleton.ca

J. Heidt
e-mail: JulianHeidt@cmail.carleton.ca

G. G. Pinchbeck
e-mail: Geoff.Pinchbeck@carleton.ca

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_5

television (e.g., Will, 2001) and even reading for enjoyment (e.g., Patri, 1938). Taken at face value, gameplay possesses many affordances that makes it ideal for language learning: visual-input, interaction, knowledge through situated meanings, and embodiment of the player within the game (Gee, 2004, 2006). Seen from this perspective, it is perhaps not surprising that gaming has been identified as a valuable source of extramural language input. However, research is sparse about the ways in which this input can be best exploited for learners. One approach to supporting learning through this type of input is to provide learners with specialised videogame vocabulary to facilitate language learning through gaming. To identify such specialised words, we were therefore interested in examining both: (1) the types of English vocabulary encountered while gaming, and (2) the contexts in which different gaming words are used.

2 Background and Rationale

Most research that has investigated the use of videogames in language learning has employed quasi-experimental designs, examining games as the *medium* for learning as opposed to investigating their lexical make-up (cf. Chen & Yang, 2013). Several studies have also shown correlations between language proficiency in additional language learning contexts and time devoted to gaming (De Wilde et al., 2019; Sundqvist & Wikström, 2015; Sylvén & Sundqvist, 2012). The findings that link gaming to language learning might not be all that surprising given how more formal language learning, especially in foreign-language learning contexts, can often be de-motivating (Dörnyei, 1994), and games may offer a way to motivate English as a Foreign Language learners. The suggestion that gaming might promote vocabulary acquisition, in particular, is also consistent with vocabulary research theory. The intensive engagement that gaming elicits aligns with the *involvement load hypothesis*, which posits that the more cognitively engaging the input is for the learner, the more likely that they will acquire and retain new vocabulary (Laufer & Hulstijn, 2001). There is, however, very little known about how language and vocabulary might be learned while playing games and as Schmitt (2019) recently pointed out, there is a "…need to understand the specialist vocabulary in games…" and determine whether "…different games promote different types of vocabulary" (p. 268).

The importance of the relationship between the lexical make-up of texts and the texts' suitability as a source of L2 vocabulary input has been recognized for some time. Therefore, a good entry point for the investigation of games as input is to conduct a corpus coverage study. *Lexical frequency profiling* (LFP) studies are estimates of the number of words a learner needs to understand in a text in order to comprehend the text as a whole, and this is often referred to as *lexical coverage* (Nation, 2006). Nation (2006) estimated that for a learner to be able to fully understand a text without guidance, the learner should encounter no more than one unknown word in every 50 words (i.e., 98% known words), but with guidance and scaffolding, they might be able to manage up to one unknown word in every 20 (i.e., 95% known words).

Therefore, LFP studies report how many word families are required to reach 95% and 98% coverage from different types of language input. These coverage levels usually include both proper nouns (PN) and marginal words (MW) in their totals (Nation, 2006; Webb & Rodgers, 2009a, 2009b). The assumption is that learners with the vocabulary knowledge to obtain these coverage levels will have enough contextual information of a given text that unknown words can be understood and may not substantially hinder comprehension (Nation, 2006). The more words required to reach these coverage thresholds, the more difficult the texts are estimated to be.

In an earlier study, Nation (2006) determined that 95% of the text in several classic novels is covered by words within the first four 1000-word bands (4000 word families total) of his British National Corpus word family lists (when proper nouns were assumed to be understood by learners), and that 98% coverage was obtained using words up to the 9000-word band. In an LFP study on two different songs corpora, Tegge (2017) found that 95% coverage was reached at the 3000-word band and 98% at the 8000 band. Similarly, in the case of movies and television, 95% coverage was shown to be reached at the 3000 band (Webb & Rodgers, 2009b; Rodgers & Webb, 2011), and 7000 to 8000 word families were able to cover 98% of these texts. Thus, this same approach was used in a recent study using a commercial off-the-shelf (COTS) games corpus (Rodgers & Heidt, 2021). In their study, 4000 to 5000 word families were required to cover 95% of the game corpus text and 10,000 word families for 98% coverage. This is notably higher than that required to cover the same percentage of text of songs, television, movies, and classic novels. As shown in Table 1, a comparison of these coverage studies indicates that the texts in games might be more lexically sophisticated than that found in other sources of language input commonly used in language learning contexts. If *extensive playing* were to be considered as a feasible pedagogical strategy, analogous to that of *extensive reading* (see Nation, 2015), it would seem reasonable to first explore and better understand the nature and identity of the vocabulary encountered in COTS games.

2.1 Videogame Corpus Compilation

A videogame corpus can be compiled from at least two different types of sources: (1) closed captions (CC) of the audio text encountered while playing games, and (2) the text that is encoded in the game program files. In the first type of corpus, text is extracted from CC of 'playthrough' videos, which are screen recordings of one instance of an entire game as played from beginning to end. The CC are generated automatically when playthrough videos are uploaded to sites such as YouTube. The text extracted from these CCs can then be compiled as a corpus of the game narrative and/or dialogue that a gamer would have heard or read while playing. The second type of corpus is composed of all of the text that is encoded in the game program files. This program file text often includes the CC text provided by the game manufacturer, as well as all of the written text that might ever be encountered on-screen.

Table 1 Lexical coverage reported by various LFP studies

Frequency band	Reading (Nation, 2006)	Songs (Tegge, 2017)	Movies (Webb & Rodgers, 2009b)	Television (Rodgers & Webb, 2011)	Games (Rodgers & Heidt, 2021)
1,000	82.93	88.96	89.76	89.16	83.64
2,000	90.14	93.18	9402	93.56	90.19
3,000	93.28	**95.13**[a]	**95.83**[a]	**95.45**[a]	93.01
4,000	**95.06**[a]	96.32	96.89	96.51	94.79
5,000	96.13	97.15	97.52	97.21	**96.01**[a]
6,000	96.88	97.42	97.96	97.64	96.75
7,000	97.43	97.79	**98.22**[b]	97.91	97.23
8,000	97.90	**97.99**[b]	98.45	**98.17**[b]	97.69
9,000	**98.22**[b]	98.09	98.62	98.35	97.96
10,000	98.46	98.17	98.78	98.51	**98.16**[b]
Corpus Size	121,099	180,892	2,841,887	1,330,268	5,744,388

Note Coverage figures include proper nouns and marginal words; [a] reaching 95% coverage; [b] reaching 98% coverage

The LFP study conducted by Rodgers and Heidt (2021) described above used corpora compiled from game program files, and the findings indicated that, from a lexical perspective, it might be difficult for beginner or intermediate learners to understand game texts. However, the authors also indicated that there were likely two different types of vocabulary in games and suggested that only one of these accounted for the apparent difficulty of game texts. When a smaller corpus compiled from the CC texts of a YouTuber's complete playthrough of one of the games in the corpus (Grand Theft Auto V) was used, these texts reached 95.74% coverage with only the first 3000 word families, and 98.25% at the 8000 word level, which aligns with the lexical thresholds of movies, television, and songs, and are considerably lower than those found for texts extracted from the game program files (see Table 1 above). Additionally, the entire CC audio text of playthrough corpus (82,370 tokens) accounted for less than 10% of the entire text encoded in the game (949,165 tokens: see Table 2).

2.2 *Pedagogical Word List Compilation and Validation*

Pedagogical word lists have often served as useful heuristics for teachers and materials designers and the extent of their usefulness can be assessed using lexical frequency profiling. Essentially, a pedagogical word list that represents a target register should provide the greatest amount of coverage in the fewest words possible. For example, the 570 word families in the Academic Word List provides 10%

Table 2 Videogame corpus from Rodgers and Heidt (2021)

Name	Genre	Release year	Tokens
Sam & Max	Graphic adventure	2007	139,719
Walking Dead	Graphic adventure	2012	221,526
Phoenix Wright Trilogy	Life simulation	2019	751,347
Sims 4	Life simulation	2014	323,555
Divinity: Original Sin 2	Action role-playing	2017	1,095,755
Monster Hunter World	Action role-playing	2018	205,219
Fallout 4	Action role-playing	2015	1,029,969
Skyrim	Action role-playing	2017	960,322
Grand Theft Auto (GTA) V	Action-adventure	2013	949,165
Resident Evil 6	Action-adventure	2019	67,811
Total Tokens: 5,744,388			

coverage of academic texts (Coxhead, 2000). Therefore, if language learners of English were to gain the greatest benefits from a games word list, the following three criteria must be balanced. First, the list should achieve a substantial coverage of a games corpus. Second, it should comprise the fewest number of words possible to minimize learner study time. Finally, the words on the list should not already be known to the students. In previous research on lexical coverage across a range of genres, knowledge of 3,000 word families was necessary to achieve 95% coverage, which is widely considered the minimum level for text comprehension (c.f. Laufer & Ravenhorst-Kalovski, 2010). This is the case for such text genres as spoken discourse (Adolphs & Schmitt, 2003), and as was presented in the Background and Rationale section, movies (Webb & Rodgers, 2009b), television (Webb & Rodgers, 2009a), and pop songs (Tegge, 2017). Thus, for word lists where the goal is to identify potentially unknown words for learners, words from beyond this level of 3000 word families is recommended. These recommendations were used as a guideline for the list created in this study, and where it was necessary to stray from them, a rationale has been provided in the methodology section.

2.3 Functional Classification of Game Vocabulary

The vocabulary found in games can be classified into two different functional categories. First, words that are central to the story can be classified as ‘diegetic’ (Castelvecchi, 2020). This type of vocabulary features heavily in the dialogue and narration of a game. The second category of language that is key to the manipulation of the game is called ‘ludic’ (Crystal, 2001). This vocabulary might be encountered as text that provides identifying and descriptive information about the characters, creatures, objects, tools, or weapons that a player might encounter, engage with, or manipulate while playing. For example, if the word *melee* were encountered in a game as in "this item increases melee damage", *melee* would be serving a ludic function, where *melee* is a specialized term for a type of damage that the player can inflict. If the same word form were used as part of dialogue or narration, such as 'there is a melee going on', that would be an example of diegetic language. Based on their lexical coverage study, Rodgers and Heidt (2021) hypothesized that since the vocabulary found in a game playthrough was more accessible to language learners, the low-frequency vocabulary encoded in the game software texts might primarily be comprised of ludic language. They then proposed that if a word list of such ludic vocabulary that was commonly used across many different games could be identified and compiled, it might serve as a useful pedagogical resource for teachers and/or learners.

3 Methodology

The creation of a pedagogical word list of gaming vocabulary was conducted in four steps. First, a multi-step, corpus-linguistic keyness analysis was used to identify words that are statistically more frequent in a game corpus as compared to a suitable reference corpus, but which are also not uniquely associated with any one game. Second, a lexical frequency profile of the game corpus was conducted using the game keyword list to provide an initial proof of concept for the construct of game vocabulary and the potential pedagogical usefulness of a game word list. Third, a selected, shortened list of the general gaming keywords were then classified as ludic or diegetic by raters. Finally, the words that were classified as ludic were then subclassified further according to how they are used in games.

3.1 Corpora Used in This Study

This study used the same videogame corpus created in Rodgers and Heidt's (2021) study (see Table 2). The games included in this corpus were a sample of convenience, which was limited by the availability of decompiling software (computer programs used to extract the language) specific to these games. Although the ten games used in the corpus are not balanced in terms of token counts (see below), they nevertheless represented a range of gaming genres. The corpus text includes every word encoded and potentially available to be seen by a user, depending on which parts of the game are accessed during play and the attention of the gamer. This approach is supported by Masso (2009) who suggested that to create any kind of videogame corpus the researcher should attempt to play all the different routes found in the game and experience as much as they possibly can to get a representative sample of the language used. There is a noticeable difference in the number of tokens between the different games. For example, Divinity: Original Sin 2 has almost eight times as many tokens as Sam & Max. However, this is a function of the amount of game play available. Sam & Max takes approximately 12 h to complete whereas there is over 100 h of gameplay in Divinity: Original Sin 2.

The $\mathrm{SUBTLEX}_{\mathrm{US}}$ is a 51-million-word corpus of American television shows and films (Brysbaert & New, 2009), and it was used as the reference corpus against which the vocabulary in the Rodgers and Heidt's (2021) videogame corpus could be compared. $\mathrm{SUBTLEX}_{\mathrm{US}}$ was chosen because the register of television and movies is similar to that of games; they both consist primarily of narration and dialogue, and the texts in games and television/movies are both accompanied by visual media. Moreover, this corpus is freely available.

3.2 Creating an Initial Gaming Word List Based on Keyness

As we found no precedent for the word family lexical unit to be used in keyness analyses studies, we conducted all keyness analyses using flemma-based word lists (Paquot & Bestgen, 2009). Word frequency lists for the gaming corpus and for each of the 10 games' sub-corpora were first generated in AntConc (Anthony, 2019) using a flemma-based lemma-list, such that the 5.7 million tokens comprised 43,540 flemma groups, of which 13,551 were hapax forms. A *flemma* is a word grouping that falls in between the *lemma* and the larger *word family* grouping. While a 'lemma' is "a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling" (Francis & Kučera, 1982, p. 1), a 'flemma' is a base form and inflected forms, regardless of word class (Pinchbeck, 2014, 2017). The f/lemma groupings differ from the 'word family', which consists of a freestanding base form and a range of affixed forms, including both inflectional forms *and* derivational forms (Bauer & Nation, 1993).

Milička's (2012a, 2012b) free *KeyWorder* software was used in two steps to identify flemmas that are more frequent in the target (i.e., gaming) corpus as compared to the $\mathrm{SUBTLEX}_{\mathrm{US}}$ reference corpus. In brief, KeyWorder provides the minimal ratio (MR) for all words by dividing the frequency of a word in the target corpus by the *minimal expected frequency* (MEF) of a word given its frequency in a reference corpus. The MEF is the lower frequency limit of the confidence interval as determined by the Fisher exact test for a given alpha value. There were 5373 'key' flemmas identified in this way in the entire game corpus as compared to the reference corpus.

The goal of this study was to find vocabulary typical of games in general, and to exclude vocabulary that was prevalent in only one game. For example, a videogame set in medieval Europe would likely include words such as *castle*, *knight*, *serf*, and *lord*, etc. These words would be indicative of only this specific game rather than of all games in general. Therefore, vocabulary of each game was compared to that of the entire game corpus as a reference in *KeyWorder*. Words with a high MR value for any single game sub-corpus were removed from further word list development. As an additional word-filtering step, words retained for the general videogame list had to be found in at least six of the 10 game sub-corpora. After these steps were followed, of the 5373 key flemma initially identified, 484 game keyword flemma remained. This list is available as an open-source resource (Flemma List Link[1]).

3.3 Game Text Coverage Provided by Game-Specific Words

As an initial test of how useful the game keyword list might be from a pedagogical perspective, a lexical frequency profile analysis was performed to determine the proportion of the game corpus covered by the 484-word initial keyword vocabulary list. Coverage of a significant proportion of the corpus would mean that explicit learning from the list would more likely enable English-learner gamers to gain better access to the language input afforded by games. This would therefore justify further development of the list and it would validate the list as representative of the general game language construct.

First, since LFP analyses are traditionally conducted using word families, the members of the 484 flemmas in the word list that are within the same word families were combined, which resulted in a new list comprised of 457 word families. These word families covered 6.06% of the entire game corpus. This is a smaller coverage figure than Coxhead's (2000) Academic Word List (AWL) provides to her original academic corpus, but the game keyword corpus is smaller (484 words vs. 570 AWL words) and the methods used to develop each list were very different. The method used to compile the AWL means that the AWL contains words that are also frequent in other registers of English, whereas the game keyword list is focused on game vocabulary.

While the third recommendation outlined in the Pedagogical Word List Compilation and Validation section was only to include words from beyond the first 3000 word families on Nation's BNC/COCA lists, a choice was made to have a small portion of the list created here contain words from within the 3000 most frequent words. Given the method used to identify key words in the corpus, we believed that high frequency words that were identified may have a specialised usage in games and were therefore included. However, with the recommendation in mind, we also wanted to determine the unique coverage provided by the game keyword list after only the coverage of the 235 families from the first 1000- to 3000-word lists of the BNC-COCA was removed. This analysis revealed that 5.41% of the game corpus was covered by the 222 word families from the game keyword list that are on BNC/COCA lists 4000 or higher. The difference in coverage between this analysis and the first analysis indicates that 0.65% (i.e., 6.06% - 5.41%) of the total game words (tokens) fall into the 3,000 most common word families in English. Thus, most of the coverage provided by the game keyword list is due to the 222 word families from the 4000 band and higher, which perhaps identifies a niche vocabulary. As previously pointed out, the most frequent/dispersed 5,000 words of English (including proper nouns

[1] https://www.iris-database.org/details/ckt6u-aRM4l.

and marginal words) provides 95% coverage of the entire videogame corpus (see Table 1) and it is not until 10,000 word families that 98% coverage is achieved. This is a larger number of words necessary than that required in other genres examined in lexical coverage research. However, the 5.41% coverage provided by the word families from the videogame list (with only those words beyond the 3000-word list) when added to 90.19% coverage from 2,000 word families (plus proper nouns and marginal words) provides over 95% coverage—the minimum level of coverage that has been recommended for language learning through the other genres. This indicates the potential pedagogical value of this word list. The videogame list might support and/or scaffold learning through gaming in a manner similar to the way that the AWL has been used to support language learners reading academic texts. Although 5.41% of the game corpus represents a substantial coverage, we also recognized that the universe of games is heterogeneous, and it is likely that different genres of games might contain large variations in vocabulary use. For this reason, we also tested the coverage of the keyword list on the largest sub-corpus of the game corpus, that of Action Role-playing Games (ARPG) genre. The four ARPG games of the corpus (*Skyrim, Monster Hunter World, Fallout 4*, and *Divinity 2*) represented 3.2 of the corpus' total 5.7 million tokens. The keyword list made from the ARPG corpus comprised 543 flemma (Flemma List Link[2]). The coverage provided by the ARPG keyword list covered 9.63% of the ARPG sub-corpus (see Table 3), over and above the coverage provided by the words in Nation's first 1000- to 3000-word lists. This latter result is in line with Coxhead's AWL 10% coverage of her academic corpus of 3.5 million tokens (Coxhead, 2000). These results suggest that while there may not be a clear and distinctive set of general words that can be found across all games, the language found in subgenres of the games, at least for ARPG games, might be comprised of a more homogeneous vocabulary.

In any case, the keyword vocabulary in the 10-game corpus is specialized and lexically rich enough to be worthwhile for learners to invest their time acquiring. However, it is also important to remember that a learner can choose to read or not read the actual game text, and the vocabulary presented will depend on the gamer's specific path chosen in the game. Such factors will determine the potential degree of vocabulary acquisition that takes place.

Finally, it was recognized that the large size of the initial 484-flemma game keyword list may not be practical for pedagogical purposes. Therefore, to produce a shorter word list that would be more manageable, the 100 words with the largest MR values from the initial keyword analysis were selected for further analyses. The vocabulary composition of the final version of the Good Gaming (GG) Word list is shown in Table 4.

The section that follows describes the steps we took to classify the function of the GG word list words in the context of games.

---

[2] https://www.iris-database.org/details/ckt6u-aRM4l.

Table 3 Lexical coverage of only the ARPG keyword list and ARPG games

Word Band	Tokens	Coverage (%)	Cumulative coverage (%)
1,000	2,401,135	72.95	72.95
2,000	164,537	5.00	77.95
3,000	80,375	2.44	80.39
ARPG word list	301,626	**9.16**	**89.55**
4,000	63,031	1.92	91.47
5,000	42,721	1.30	92.77
6,000	26,780	0.81	93.58
7,000	16,978	0.52	94.10
8,000	13,597	0.41	94.51
9,000	9,993	0.30	94.81
10,000	7,327	0.22	95.03
11,000	6,202	0.19	95.22
12,000	5,080	0.15	95.37
13,000	3,055	0.09	95.46
14,000	1,781	0.05	95.51
Total Tokens	3,291,265		

3.4 Coding the Words for Ludic Properties in the GG Word List (Qualitative)

3.4.1 Sample Selection for Evaluation Coding

Each word on the 100-word GG word list went through a process of evaluative coding. Evaluation coding is a form of assessment that stems from the researcher's own evaluative perspective (Saldana, 2015). It is a system that reflects initial questions that structured the evaluation. For this study, words were classified as ludic or diegetic based on the word's usage in context in the KWIC (keywords in context) lines. The coding of all words was done by the first author, but an inter-rater reliability analysis was conducted to assess the consistency of the original coding. Following this, the functions of all words that were categorized as ludic were sub-categorized further.

The first step in the process of evaluative coding was to sample examples of contexts in which each word was used in games, since evaluation of every instance of all 100 words in the game corpus was not practical. A systematic sampling method was employed using Anthony's (2019) *AntConc* software that allows the user to "thin" the data by selecting every nth row of concordance lines. For this analysis, the goal was to generate 10 concordance lines for each of the 100 words on the GG list, and this required that only every three rows, or more, of the original output were displayed, depending on how many concordances were available. This sampling method was

Table 4 The GG word list categorized as Ludic Words or Diegetic words

Ludic Words		Diegetic Words							
1	melee	1	teleport	21	lemme	41	flame	61	delve
2	recoil	2	trader	22	composure	42	scrawl	62	path
3	tutorial	3	artifact	23	courtyard	43	hideout	63	volume
4	npc	4	nearby	24	flee	44	vigilant	64	hunger
5	gameplay	5	retrieve	25	glare	45	mist	65	rogue
6	optional	6	foe	26	grumble	46	décor	66	leather
7	autosave	7	teleportation	27	thankfully	47	weaponry	67	skull
8	current	8	mastery	28	return	48	glow	68	pillar
9	overwrite	9	infuse	29	shard	49	power	69	ensure
10	disable	10	successfully	30	brewing	50	formidable	70	fiery
11	dialog	11	stealth	31	culprit	51	quarry	71	vase
12	highlight	12	crimson	32	cobble	52	craftsmanship	72	madness
13	attacker	13	wary	33	jagged	53	feat	73	poison
14	potent	14	pause	34	cloak	54	sigh	74	reveal
15	rename	15	useful	35	masterwork	55	cleanse	75	beneficial
16	regeneration	16	stagger	36	skeletal	56	creation	76	remain
17	tier	17	relic	37	deity	57	upwards	77	loot
18	navigate	18	fearsome	38	armoury	58	amethyst	78	skilled
19	resurrect	19	scythe	39	unleash	59	otherworldly	79	grime
20	projectile	20	agility	40	assassin	60	fury	80	capacity

done to minimize sample bias that would have resulted from using the first 10 lines output by the software's default settings. A more detailed description of the methods used for the evaluative coding is provided in Heidt (2020).

3.4.2 Inter-rater Analysis of Game Vocabulary Coding

All of the raters chosen for this study were familiar with the games included in the corpus and were able to understand the limited context provided by the KWIC lines. All raters were first trained on five words with four concordance line examples each (20 examples total) which illustrated ludic versus diegetic usage. The supplementary raters were also provided with the researcher's rationale for categorizing those five words as ludic or diegetic. Raters were permitted to ask questions about any of the evaluations performed in the training sessions. Training words were also marked for game context in which they were used; for example, "turn water and blood surfaces into oil", was tagged "(skill description)". During the coding session, raters could choose "A. LUDIC, B. DIEGETIC, or C. REQUEST CLARITY." If option "C" was requested, a surrounding concordance would then be retrieved and given to the rater if applicable.

Following the training session, the PI rated the 100 GG words in 1000 concordance lines, and then the three supplemental raters each evaluated 10 lines for a subset of 28 words. Inter-rater reliability for the 280 KWIC lines (n = 4) resulted in 87.86% agreement with a free-marginal kappa value of 0.76 [0.70, 0.81]. Cases of disagreement were resolved through deeper examination of the context and discussion between the raters. This figure is in line with a high reliability (Mackey & Gass, 2015; Saldana, 2015).

Finally, each word on the GG list was assigned as primarily ludic or diegetic. Words that were used as ludic function in at least five out of the 10 KWIC lines examined were classified as ludic and words that did not meet this criterion were labelled as diegetic. The reduced GG list with all ludic and diegetic words identified can be viewed in Table 4.

3.4.3 Pattern Coding

A second cycle of evaluative coding is often done in qualitative research to ensure reliability or to addresses gaps in the initial coding (Saldaña, 2015). The initial gap that is resolved through this second cycle of coding, pattern coding, is *how* the words are being used ludically (see Table 5). Pattern coding "organizes the corpus into sets, themes, or constructs and attributes meaning to that organization" and that it is "appropriate for second cycle coding" (Saldaña, 2015, p. 6). These 20 ludic words on the GG Word List were classified into three categories: *Game System, Game Property,* and *Game Information.* Game *System* words inform the player about alterations that can be made regarding the game system (the console or personal computer the game is being played on). Game *Property* words are the words used in the game to refer to attributes found within the games themselves, which convey some important part of the gameplay. Finally, Game *Information* words are words similar to the game system, but instead of relating to the system, they related to the game itself. This is the language used by the game to inform the player of the systems within the game that enable various features of play. Table 5 shows the primary usage pattern for 10 ludic words and the number of times each word was used in this way in the coding sample.

4 Implications of Current Research

The findings of this study suggest that there is a distinctive vocabulary associated with games, and representatives of this vocabulary have been provided in three open-source word lists: the GG Word List, the full game keyword list, and the ARPG keyword list. The findings also illustrate how various game words can function in both diegetic and ludic contexts, and the latter type has been further subcategorized into three patterns. This information could inform strategies for teaching

Table 5 Examples of the pattern coding (how Ludic words are used)

Word	Ludic Type	How the word is being used	Ludic Type Occurrences (/per Ludic uses)	Ludic Uses (/10)
Current	SYSTEM	System informs the state of controls (buttons) or settings	3/5	5
Overwrite	SYSTEM	System informs of a consequence for an intended action regarding saving or game settings/alterations	10/10	10
Disable	SYSTEM	System informs that certain settings can be disabled	7/8	8
Melee	PROPERTY	A game property: melee damage. Collocates with "attack" and "damage"	8/9	9
Recoil	PROPERTY	A game property: gun/bow knockback. Associated to gun recoil. Collocates with close-gun related terms such as "scope, shooting, aim, suppressor, reload time"	7/7	7
Attacker	PROPERTY	A game property: Refers to an in-game enemy	6/6	6
Tutorial	INFORMATION	Game informs player regarding tutorial settings and tooltips	10/10	10
NPC	INFORMATION	Game informs player on non-playable characters and their functions	10/10	10
Gameplay	INFORMATION	Game informs player about gameplay options and experiences	9/9	9

and/or enhancing learners' incidental acquisition of the vocabulary while gaming and suggests directions for future research.

In the 100-word GG Word List, 80 words were found primarily to be diegetic and 20 were ludic. Words in this study were categorized as ludic if at least five out of 10 usages were ludic. For some overtly marked game words such as *autosave*, *recoil*, and *gameplay*, it was clear that these words primarily have a ludic function, as these terms are commonly used in games to refer to concepts that are outside of the game narrative. The small number of ludic words in this list overall might suggest that gaming does not have a common sophisticated technical word list that can be described but might use more general language for these purposes. It could also be that ludic functions are realized through other linguistic and/or non-linguistic means, apart from the single words. A list of 20 ludic words is nevertheless a simple, manageable number for teaching and learning, especially for those students who are unfamiliar with gaming, or who are used to different game-genres.

A main finding of this study is that although many game keywords can be used in a ludic context, the categories of ludology and diegesis are not mutually exclusive for individual word forms. It is also important to consider polysemy, even when considering two different ludic contexts. For example, in the game *Monster Hunter World*, ‘surfacing’ is a state into which certain monsters can be configured (to come out of the ground), and certain items known as ‘Screamer Pods’ can force them to surface. However, in the game *Divinity 2*, ‘surface’ is used to refer to the environment that the players are on. This environment can be changed by the player using skills and items. For example, a surface can be made icy, and an enemy can slip, or a puddle of water can be turned into oil, which can then be ignited. Based on insights gained through the process of the pattern coding in this study, it became evident that it is not entirely the word that dictates a ludic or diegetic function, but the context in which the word is used in the game as well. One consequence of this observation is that it may be difficult for teachers to target ludic words for games with which they are not personally familiar. While making the distinction between ludology and diegesis is instructive in the research of the language of games, it is yet unclear how much of a role it might play in text comprehension and vocabulary acquisition. In the example above, simply knowing the diegetic context of the word ‘surface’ might be sufficient to understand their ludic functions (i.e., to enact a monster surfacing). Despite the small sample of concordances analyzed in this study, our findings illustrate how words can be understood in these different contexts. By giving some context to the ludic and diegetic words, practitioners might use these lists to teach ludic words to learners who may not have had previous experience with certain game genres and might even encourage their use, as games have been shown to be a good extramural source of language for vocabulary acquisition.

The discovery that game vocabulary might be a valid construct suggests that there is a potential for videogame word lists in language learning and teaching. There are several pedagogical implications that relate to the use of game word lists. First, gaming word lists might be used by materials developers to create videos that language learners could watch prior to embarking on a new game adventure. Such videos might highlight a sample of game playthroughs that highlight the usage of a short list of words that are central to either the story and dialogue (diegesis) or to the manipulation of the game (ludology). Such videos might be recommended by language teachers who become aware that some of their students are gamers. These types of activities might be analogous to extra-curricular extensive viewing and/or extensive reading in language courses. Word lists might also be used in conjunction with reflective written work that uses a learner’s gaming experience for content or as part of discussion-focused tasks based on the situations encountered in the videogames.

This study suggests that game-genre does have an effect on the kind of vocabulary that is found in games. Reinhardt (2018) suggested that games may include a diverse range of language that would be otherwise difficult to learn in a traditional classroom, and might be more accessible than when encountered in other media such as books, songs, movies, and/or television, etc. Based on the coverage analyses described in this study, it may be more beneficial to use gaming as an extramural activity for intermediate to advanced language learners to expand their vocabulary, as games have a wider range of lower frequency vocabulary that may not be present or might be less accessible in other media.

The results of this study also indicate several areas in which future research might take place. While previous experimental-designs have been very useful for understanding the activity of gaming, they tend to limit our understanding of the broader processes of game-design and its affordances, particularly with respect to language acquisition. Further research might examine how learners engage with the language of games, or how communication is conducted in multiplayer games between players, and whether ludology or diegesis affect learners' comprehension of a given game or game-genre. This research might be done by performing longitudinal studies both in and out of the classroom. While research suggests a strong link between gaming and language proficiency, we do not yet understand how this happens or what language skills are particularly affected. Finally, there is also a lack of closed caption corpus data for games, and this type of corpus might lead to further analyses of the registers and genres of gaming texts.

Longitudinal studies with those adult language learners who would be most likely to engage in gaming as a source of extramural language input could examine how gaming could be conducive to the acquisition of vocabulary. With the word lists provided in this study, learners might be able to make otherwise difficult words more salient, while using the rich contexts that games provide. This line of research might also examine what kinds of games might be the best for language learning. Games that do not use a lot of vocabulary, such as the popular genre Multiplayer Online Battle Arena (MOBA) game *League of Legends* (*LoL*), would likely not be ideal to play alone. The affordance that a game like LoL might have is the communication that often takes place during the game with the player's teammates, which could help develop language problem-solving and more fluent communication. The language used for this type of communication is typically more accessible to lower-level learners, as spoken conversation contains more high frequency words, fewer word families are required to reach 95% lexical coverage (Adolphs & Schmitt, 2003). In other words, it is not the vocabulary itself found in the game, but rather, the act of communication and negotiation of meaning that might be the most helpful (Sykes & Reinhardt, 2013). When a learner plays a COTS game, a language teacher might encourage them to consider whether they will be playing alone or with others for the incidental purpose of acquiring vocabulary.

5 Limitations of the Present Research

Research on the language of gaming is still exploratory, and there are several areas where methodology will likely improve as interest in this area increases and as the technologies of corpus compilation and corpus linguistic analysis progress. First, one issue of concern to those in the field of vocabulary research is how to define what a 'word' is. In this study, the keyness analyses used the 'flemma' to count words. The rationale for this is that polysemy is less of an issue when derivative forms are not included in the same word groupings, particularly when keyness analyses are performed (Paquot & Bestgen, 2009). In contrast, the lexical frequency profiles in this study used word families to allow for comparison to other LFP studies, as these studies had also used the word-family as a counting unit. These decisions were made for practical reasons, and we look forward to a more comprehensive examination of which word counting unit might be more appropriate for which learners and for the specific purposes for which word lists are used. A second issue that might be seen as a limitation of this study is that the game corpus that was used was not tagged for parts of speech (PoS). Had the corpus been tagged, and we had conducted the study using lemmas, it might have been easier to classify words according to their functions in different contexts across different games and their genres. In the example provided above, the word 'surface' as used in *Divinity 2*, is primarily a noun, but in *Monster Hunter World*, it is used as a 'verb'; thus, these two different functions would have been distinguished simply with the PoS tag. We look forward to progress in the quality and level of information that future COTS games corpora will provide. A third limitation is that this research is based on a single, small, and underbalanced corpus of the language in videogames. For comparison purposes, performing a lexical coverage analysis on a second or parallel corpus of games with our word list would have been ideal and increased our ability to discuss the validity of the list. However, at present, there are no other videogame corpora that could serve this purpose. It is hoped that as interest in gaming and language learning increases more corpora will be created such that these cross-corpora analyses can be completed.

6 Conclusion

The research presented here represents an initial response to Schmitt's (2019) call and provides some directions for future research in this area. The main findings from this study are: (1) games provide a rich range of vocabulary, as is seen by their lexical frequency profiles, (2) vocabulary usage in games can be categorized in ways that are possibly useful for research and pedagogy, and (3) certain varieties of games may have common vocabulary that players might learn more rapidly by playing within the same game genres over a period of time. We further speculate that the lexical diversity of game texts may make them suitable in terms of language learning for intermediate to advanced learners and that the motivational affordances of gaming can be harnessed by language educators.

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425–438.

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Bauer, L., & Nation, P. (1993). Word Families. *International Journal of Lexicography, 6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of anew and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Castelvecchi, S. (2020). On "Diegesis" and "Diegetic": Words and concepts. *Journal of the American Musicological Society, 73*(1), 149–171.

Chen, H., & Yang, T. (2013). The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive Learning Environments, 21*(2), 129–141. https://doi.org/10.1080/10494820.2012.705851

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Crystal, D. (2001). *Language and the Internet*. Cambridge University Press.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism, 23*(1), 171–185. https://doi.org/10.1017/s1366728918001062

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*(3), 273–284. https://doi.org/10.1111/j.1540-4781.1994.tb02042.x

Francis, W., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin.

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Routledge.

Gee, J. P. (2006). Are video games good for learning? *Nordic Journal of Digital Literacy, 1*(03), 172–183.

Heidt, J. S. (2020). *"GG" Good Gaming Wordlists: Insight into Single-player commercial off-the-shelf games and action-roleplay game vocabulary* [Unpublished M.A. Thesis]. Carleton University. https://curve.carleton.ca/f11aa1b9-1e8a-4915-86bc-635bf915ed0a

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

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Mackey, A., & Gass, S. (2015). Second language research. *Erlbaum*. https://doi.org/10.4324/9781410612564

Masso, C. I. (2009). Developing a methodology for corpus-based computer game studies. *Journal of Gaming & Virtual Worlds, 1*(2), 143–169. https://doi.org/10.1386/jgvw.1.2.143/7

Milička, J. (2012a). Keyworder. [Computer Software]. Institute of Comparative Linguistics, The Faculty of Arts, Charles University. http://www.milicka.cz/en/keyworder/

Milička, J. (2012b). Minimal ratio: An exact metric for keywords, collocations etc. *Czech and Slovak Linguistic Review, 1*, 62–71.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27(1), 136–145. http://hdl.handle.net/10125/66705

Paquot, M., & Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. *Language and Computers, 68*, 247–269. https://doi.org/10.1163/9789042029101_014

Patri, A. (1938, June). Too many novels are read by young people. *The St. Petersburg Times.*

Pinchbeck, G. G. (2014, March). *Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 Academic English* [Roundtable presentation]. American Association for Applied Linguistics Annual Conference 2014. https://app.box.com/s/2772oofxkg6jyzmd3l7j

Pinchbeck, G. G. (2017). *Vocabulary use in academic-track high-school English literature diploma exam essay writing and its relationship to academic achievement* [Unpublished Ph.D. Thesis]. University of Calgary. https://doi.org/10.11575/PRISM/28660.

Reinhardt, J. (2018). *Gameful second and foreign language teaching and learning: Theory, research, and practice.* Springer.

Rodgers, M. P. H. & Heidt, J. (2021). Levelling up comprehensible input and vocabulary learning: The lexical profile of video games. In Werner, V. & Tegge, F. (Eds.), *Pop culture in language education.* Routledge.

Rodgers, M. P. H., & Webb., S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly, 45*(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Saldaña, J. (2015). *The coding manual for qualitative researchers (3rd edition).* Sage.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274. https://doi.org/10.1017/S0261444819000053

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

Sylvén, L., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL, 24*(3), 302–321. https://doi.org/10.1017/S095834401200016X

Sykes, J. E., & Reinhardt, J. (2013). *Language at play: Digital games in second and foreign language teaching and learning.* Pearson Higher Education.

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System, 67*, 87–98.

Webb, S., & Rodgers, M. P. H. (2009a). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S., & Rodgers, M. P. H. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Will, G. F. (2001, June 21). TV for Voyeurs. *The Washington Post*.

Julian Heidt completed his M.A. in Applied Linguistics and Discourse Studies at Carleton University and is a TESL-accredited teacher. His research focus is on video games and how they can be used as a pedagogical tool for informal language learning and acquisition.

Geoffrey G. Pinchbeck completed PhDs, in Education and Medical Biochemistry and is an Assistant Professor and Coordinator of TESL programs in the School of Linguistics and Language Studies at Carleton University, Canada. His research focuses on the role of vocabulary knowledge in writing and reading proficiency and on corpus-based approaches to language program curriculum design and materials development.

Michael P.H. Rodgers is an Associate Professor within the School of Linguistics and Language Studies at Carleton University. His academic interests concentrate on language learning through

media. In particular, his research has focused on comprehension of television programs viewed with and without captions, as well as vocabulary acquisition through viewing videos, lexical coverage of television, movies and videogames, and the effects of imagery on the word learning process.

Involvement Load and Vocabulary Acquisition in Digital Game-Based Tasks

Amin Rasti-Behbahani

Abstract This study investigated the effectiveness of digital game-based vocabulary learning (DGBVL) task (It is worth mentioning that, in this study, the term task refers to vocabulary learning activities)-induced levels of involvement load (LIL) on vocabulary acquisition. Firstly, based on the criteria defined in the *Technique Feature Analysis* (TFA) checklist, three DGBVL tasks were designed to induce low, moderate, and high LIL. After that, 30 Persian-speaker teenagers were recruited as participants. Next, participants were randomly assigned a DGBVL task for completion in pairs. Finally, participants' vocabulary acquisition was evaluated by both receptive and productive delayed post-tests three weeks later. Moreover, think-aloud data were collected during and after task completion. The results showed that: 1) all three DGBVL tasks supported vocabulary acquisition; 2) the LIL mattered, and a high LIL produced more vocabulary acquisition; 3) the structural elements of DGBVL tasks could alter the predictability of LIL; and 4) the structure of the DGBVL tasks may invoke different mental processes and therefore vocabulary learning. Pedagogically, this study suggests that prospective teachers may use pre-teaching techniques in lieu of incorporating glosses into DGBVL tasks provided to students. However, if glosses are to be used, teachers should not use multiple-choice glosses to avoid confusion.

Keywords Digital Game-based learning · Incidental Vocabulary Learning · Involvement Load Hypothesis · Multiple-choice glosses · Single-word glosses · Pre-teaching · Task · Video game

Most learners acquire vocabulary incidentally (Webb, 2020). Incidental vocabulary learning (IVL) is learning a word while engaging in a language learning activity not specifically focused on vocabulary (Hulstijn, 2001). Despite it being slow and incremental (Schmitt, 2010), "incidental vocabulary learning gains are not only meaningful but central to L2 lexical development" (Webb, 2020, p. 231) because, concerning the multidimensionality of knowing a word (Nation, 2001),

A. Rasti-Behbahani (✉)
Department of Language and Communication Studies, University of Jyvaskyla, Jyvaskyla, Finland
e-mail: amin.rasti@outlook.com; aminrasti@du.edu.om

College of Arts and Applied Sciences—Foundation Program, Dhofar University, Salalah, Oman

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_6

IVL may enhance the acquisition of many pieces of information such as association, grammatical functions, collocations, register, etc. in addition to the form-meaning relationship (Webb, 2020). In addition, many techniques, such as glossing, pre-teaching, mnemonic devices, dictionary-work, and so on, accelerate and enhance IVL (Sökmen, 1997). However, the search for more effective techniques continues. Using digital game-based vocabulary learning (DGBVL) tasks is a recent technique (Bahari, 2020; Chen et al., 2018; Martinez et al., 2022), which is the focus of this study. DGBVL is the use of digital games for teaching and learning vocabulary (Rasti-Behbahani & Shahbazi, 2020). Although the DGBVL literature generally supports DGBVL tasks, some studies have shown that the type and the design of the DGBVL tasks can influence their effectiveness. For example, deHaan et al. (2010) found that a DGBVL task with a high level of interactivity can hinder word form recall. However, replicating the same study but implementing a different task design, Ali Mohsen (2016) found that a high level of interactivity can enhance IVL gains and form recall. Ali Mohsen (2016) concludes that by evoking various cognitive processes, each DGBVL task design can affect vocabulary acquisition differently. Similarly, Yang et al. (2020) found that a DGBVL task with a dynamic task design can enhance vocabulary acquisition more effectively. They explained that an effective change can balance the complexity and difficulty of a DGBVL task with the cognitive capacity of language learners. Hence, they concluded that the effectiveness of a DGBVL task depends on the balance between its design and cognitive complexity or the number of provoked mental processes. Also, an effective DGBVL task design offers a balanced and optimal cognitive complexity.

Levels of involvement load (LIL) is a cognitive-motivational construct that can balance the cognitive complexity of tasks for effective IVL (Laufer & Hulstijn, 2001). While DGBVL task-induced LIL was observed through data collection involving questionnaire techniques (Reynolds, 2017), to the author's knowledge, its effect on IVL through DGBVL tasks has not empirically been investigated. Filling this gap in the literature can offer new perspectives on how to design DGBVL tasks with optimal cognitive complexity effectively.

1 Involvement Load Hypothesis and Incidental Vocabulary Learning

"Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) was designed to predict the effectiveness of instructional activities on incidental vocabulary learning" (Yanagisawa & Webb, 2021, p. 1281). According to Laufer and Hulstijn (2001), the ILH posits that providing rich and numerous opportunities for processing new words can effectively enhance their incidental acquisition. Furthermore, they assume that the "retention of words when processed incidentally is conditional upon the following factors in a task: need, search, and evaluation" (Laufer & Hulstijn, 2001, p. 14). In other words, different combinations of involvement in need, search, and evaluation can define the processing load of different vocabulary learning tasks. Such a combination is called ‘involvement’ (Laufer & Hulstijn, 2001). Need is the motivational part of the ILH and refers to the learner’s feeling about the importance of an unknown L2 word (Laufer & Hulstijn, 2001). Search, a cognitive part of the ILH, is the learner’s attempt at finding either the meaning of an unknown L2 word or the form of a concept in the L2 by consulting a dictionary or a teacher, and evaluation, another cognitive part, refers to any internal or external comparison of a word. The ILH assumes that the higher the weight of task-induced LIL, the larger the gain in vocabulary acquisition (Laufer & Hulstijn, 2001).

In practice, Hulstijn and Laufer (2001) compared participants’ “short- and long-term retention of ten unfamiliar words […] in three learning tasks (reading comprehension, comprehension plus filling in target words, and composition-writing with target words)” (p. 539). They found that the composition-writing with target words task that induced the highest involvement load enhanced retention of the target words more effectively than the other two tasks. Hence, they concluded that there was a relation between task-induced involvement load and the amount of retention. Similarly, Jing and Jianbin (2009) used the ILH to explain vocabulary acquisition through listening tasks. They also concluded that higher LIL could boost the retention of vocabulary items.

However, the ILH has been criticized for its quantifying method (Folse, 2006; Jahangiri & Abilipour, 2014; Keating, 2008; Kim, 2011; Martínez-Fernández, 2008). Its opponents argued that due to imprecise measurement of the components, especially evaluation, the predictive power of the current indexing method is not reliable. Reflecting on the criticisms against ILH, Nation and Webb (2010) emphasized that a new quantifying method is necessary because “the involvement load hypothesis does not include many features that other researches have shown to be important when designing vocabulary teaching techniques” (p. 7) such as the frequency of exposure (Schmitt, 2010). Nation and Webb (2010) therefore introduced a checklist that they called “Technique Feature Analysis” (TFA) (Table 4). It comprises five main categories: Motivation (which also contains the notion of need), Noticing, Retrieval (which also covers the concept of search), Generation, and Retention. The maximum checklist score is 18. If a vocabulary acquisition task provides an opportunity for triggering any of the components, the respective components will receive one point. Hence, higher scores in the checklist mean greater task-induced LIL.

The superior accuracy of TFA over the ILH quantifying method has also been argued. Hu and Nassaji (2016) showed that because TFA scores a task based more on contributing factors, especially retrieval and generation (Gohar et al., 2018), in comparison to ILH, the predictive power of TFA is more reliable than the ILH. Therefore, DGBVL tasks-induced LIL were scored by TFA in this study.

2 Digital Games and Incidental Vocabulary Acquisition

DGBVL, or learning vocabulary by playing a digital game as a main task (Rasti-Behbahani & Shahbazi, 2020), has been shown supportive of IVL. For example, Ebrahimzadeh and Alavi (2016) reported that DGBVL tasks can support the retention of form-meaning knowledge. Janebi-Enayat and Haghighatpasand (2019) found that the incidental acquisition of both the receptive and productive knowledge of words can be enhanced effectively by DGBVL tasks. Sundqvist (2019) found that playtime correlated positively with the incidental acquisition of both receptive and productive knowledge of both frequent and infrequent words. Rasti-Behbahani and Shahbazi (2020) found that a DGBVL task can enhance the incidental acquisition of components of word knowledge, especially the productive recognition of form-meaning. Bahari (2020) found that playing digital games regularly can increase gamers' depth and size of vocabulary knowledge incidentally. Overall, in recent meta-analyses, the potentials of DGBVL tasks are well-recognized (Chen et al., 2018; Chiu et al., 2012; Martinez et al., 2022). Moreover, it was found that adventure DGBVL tasks can improve vocabulary acquisition more effectively than other genres because their challenges can increase motivation (Chen et al., 2018).

Furthermore, how DGBVL tasks may enhance vocabulary acquisition has also been discussed. In general, digital games by benefiting from their inherent elements such as interactivity, rules, goals, challenge, risk, fantasy, curiosity, and control present content via rich images, animations, videos, visuals, audio, and interactive dialogues (Pivec et al., 2003; Janebi-Enayat & Haghighatpasand, 2019). Hence, they allow a high and spaced frequency of occurrence, variety in presentation mode, richness of input, and contextualized exposure to unknown words and in authentic contexts (Ebrahimzadeh & Alavi, 2016; Hwang & Wang, 2016; Janebi-Enayat & Haghighatpasand, 2019; Rasti-Behbahani, 2021). Duly, they can increase motivation; decrease anxiety; and foster interactions among learners which are all well-known and effective factors in supporting IVL (Pivec et al., 2003; Ebrahimzadeh & Alavi, 2016; Hwang & Wang, 2016; Janebi-Enayat & Haghighatpasand, 2019; Rasti-Behbahani, 2021).

Regarding the importance of DGBVL task design (Ali Mohsen, 2016; Chen et al., 2018; deHaan et al., 2010), Yang et al. (2020) designed a vocabulary learning digital game whose difficulty (cognitive complexity) level was being dynamically balanced based on the user's performance in the game. In other words, the design of the tasks in the game was dynamically changing to become more suitable for the learners' cognitive capacities. In a quasi-experimental study, they found that their digital game could enhance low-achieving students' vocabulary acquisition more effectively than the conventional DGBVL tasks. Yang et al. (2020) emphasized the importance of task design and its complexity in DGBVL tasks and effective IVL.

Moreover, Reynolds (2017) observed that DGBVL tasks can involve learners in components of ILH, including search, then need, and finally evaluation. He emphasized that "how useful one particular game can be for inducing vocabulary acquisition depends on the amount of task-induced involvement" (Reynolds, 2017, p. 482).

However, Reynolds (2017) did not investigate the effect of DGBVL task-induced LIL empirically. Also, to the best of my knowledge, this issue has not been addressed in the DGBVL literature. Therefore, investigating the effect of DGBVL task-induced LIL on IVL is timely. This study can fulfill the gap in the DGBVL literature by answering the following questions.

Research Questions

1. To what extent do DGBVL tasks make significant differences in vocabulary acquisition?
2. To what extent does the difference in DGBVL task-induced LIL make a significant difference in vocabulary acquisition?

3 Material and Methods

This is an empirical study (pre-test/treatment/delayed-post-test). Both quantitative and qualitative data were collected. DGBVL task-induced LIL was the independent factor and the rate of vocabulary acquisition was a dependent factor. Furthermore, the vocabulary acquisition was incidental because the participants' goal was to play a game, and they were not forewarned about upcoming vocabulary tests (Hulstijn, 2001).

3.1 Participants

30 Persian speakers (male = 14, female = 16; age = 13–15) were randomly recruited from a private English learning institute in Behbahan, Iran. The results of a standard paper-based *Vocabulary Size Test (Receptive, 14,000)* showed that selected participants' vocabulary sizes ranged between 2100 and 3300 words ($M = 2406.67$, $SD = 347.338$). Hence, their English proficiency level was between A2 and B1 in CEFR[1] (Milton, 2010).

3.2 Materials

3.2.1 The Digital Game

The selected digital game is called *Haunted Hotel: Death Sentence Collector's Edition*, published by Big Fish Games, Inc. In this game, a gamer plays as a detective who must solve a series of crimes by finding specific hidden objects and solving

[1] Common European Framework of Reference for Languages.

puzzles. A commercial *adventure* digital game was selected because it "enables students to use a high-end and attractive product", which draws teens' interest (Reinders & Wattana, 2012, p. 163). Moreover,

> adventure games use intrinsic motivation. Intrinsically motivating games incorporate learning […] in[to] a virtual world. Game characters have to solve a certain problem and can proceed further only after solving the problem. In this case, the problem is part of the game and [sic] players are motivated to provide a solution in order to continue […] the game. (Pivec et al., 2003, p. 218)

Crime cases and puzzles in the game can be solved by finding and using specific objects in specific places (see Fig. 1). The objects can be either found in different game scenes or created by combining specific objects (see Fig. 2).

Fig. 1 Using specific items in specific places (*Note* A rock is being used to break the glass)

Fig. 2 Item creation (*Note* Glue and fractions [found objects] are combined to create/fix a door Handle)

3.2.2 Game-Guide

The game-guide was downloaded from the official website of the game and its first chapter was used for this study. To integrate the game-guide into this study, the picture guides were first removed. The sentences were then numbered. Without informing participants about the importance of target words, they were presented in bold font. The bolded target words could reduce task complexity and accelerate task completion by attracting participants' attention to them and their definitions on the page margins (Jahangiri & Abilipour, 2014; Schmitt, 1997). The game-guide was provided to reduce time spent on the task and prevent the negative effects of extensive interactivity (deHaan et al., 2010; Rasti-Behbahani & Shahbazi, 2020).

Target Words

Forty-four nouns, from the first chapter of the game-guide, were selected. Knowing them was necessary for participants to solve puzzles in the game. However, to control the effect of repetition (Schmitt, 2010), the frequency profile[2] of the selected target words was checked. Out of forty-four nouns, twenty low-frequent names of inanimate objects such as *magnifier* (K4),[3] *skull* (K5), etc. were chosen because they were repeated only twice throughout the game-guide. Moreover, considering the participants' vocabulary size, I assumed that selecting target words from the K4 band and higher would increase the chance of them being new to the participants. The results of pre-tests confirmed this later. Only nouns were selected because nouns are easier to acquire than other word classes (Nation, 2001). Limiting learning to nouns would also reduce the burden of word difficulty in the learning process (Ellis, 1995; Sökmen, 1997). Moreover, Rasti-Behbahani (2017) found that a DGBVL task can accelerate the acquisition of nouns more quickly than other lexical word classes.

3.2.3 Task Design

To induce low, moderate, and high LIL, target words were modified by techniques such as single-word (meaning) and multiple-choice glosses and pre-teaching. Glosses were used because they make texts more comprehensible and prevent wrong guesses (Nation, 2001). Furthermore "(a) glosses trigger[…] a search for concrete meaning and firm form-meaning mapping; (b) a lack of glosses corresponds[…] with global text processing, skipping of words and shallow meaning mapping" (Rott & Williams, 2003, p. 45). Also, the pre-teaching technique increases the chance of uptake by focusing learners' attention on the vocabulary items of interest (Schmitt, 2010). It is notable that to decrease the degree of intention in the DGBVL tasks (Webb, 2020),

[2] https://doi.org/www.lextutor.ca/vp/comp/

[3] K stands for frequency level, or band.

Table 1 Lo-LIL task game-guide text

15. Click on the key **cover**; Click on the key in your bag, and put it on the door	سرپوش

Note سرپوش means cover

Table 2 Mo-LIL task game-guide text

	فَرش
15. Click on the key **cover**; Click on the key in your bag, and put it on the door	تیکه خُرده
	سرپوش

Note فَرشmeans carpet; تیکه خُرده means debris; and سرپوش means cover

the importance of the target words, and why they had been either bolded, glossed, or pre-taught were not explained.

Table 4 shows the DGBVL task-induced LIL for each task. Single-word glosses induced a low level of involvement load (Lo-LIL) and provided few opportunities for processing target words by triggering the components 2, 4, 5, 12, 15, and 18 in TFA. Hence, the Lo-LIL score was 6 (Table 1).

Multiple-choice glosses induced a moderate level of involvement load (Mo-LIL) and triggered a few more components in TFA: 2, 4, 5, 6, 7, 10, 12, and 18. The Mo-LIL score was 8 (Table 2).

Pre-teaching the target words induced a high level of involvement load (Hi-LIL). The Hi-LIL group participants received a word list containing a Persian definition and a sample English sentence for each target word. Pre-teaching triggered many more components in TFA: 2, 4, 5, 6, 7, 9, 10, 11, 12, 15, and 18. Hence, the Hi-LIL score was 11 (Table 3).

All tasks could trigger component 2 because adventure digital games are motivating (Pivec et al., 2003); 4 because target words were in bold font; 5 because vocabulary learning techniques were implemented; 12 because participants were exposed to the target words twice in two different sentences; and 18 because words with related meanings, like left and right, were not taught together (Nation & Webb, 2010). However, Lo-LIL and Hi-LIL tasks triggered component 15 because providing Persian definitions enhanced a successful linking of form and meaning. Moreover, the Mo-LIL and Hi-LIL tasks triggered component 6 because participants needed to either recognize or recall the definitions. Also, the negotiation was expected as

Table 3 Hi-LIL task game-guide text

15. Click on the key **cover**; Click on the key in your bag, and put it on the door

they did the task in pairs. The Mo-LIL and Hi-LIL tasks triggered components 7 and 10 because each target word was encountered twice in the game-guide. The Hi-LIL task triggered component 9 because there were no Persian definitions on the page margins and component 11 because target words had been pre-taught two days before the main task. The tasks were first scored by the author. Later, the scores were reviewed, discussed, and confirmed by two applied linguistics professors. There was a general instruction for all tasks: *please read the text and play the game to complete the first level* (Table 4).

3.2.4 Instruments

For eliciting data, a Vocabulary Size Test, achievement tests, think-aloud protocols, and an interview were used.

Vocabulary Size Test (VST)

To ensure the homogeneity of the selected participants, a Vocabulary Size Test (receptive, bilingual, 14,000) (Nation & Beglar, 2007) was administered. The VST is a multiple-choice test with 140 questions. For better accuracy, all 140 questions were administered. Moreover, I adapted a bilingual VST for this study. In so doing, I translated only the alternatives into Persian (Nation, 2012). When translating, I was especially careful regarding cognates (Beglar, 2010) because the Persian language is an Indo-European language and shares many similarities, especially in vocabulary, with other members of this family, such as English (Yule, 2014). I administered a bilingual VST because "in a foreign language context, participants who know the conceptual meaning of an L2 word being tested may be disadvantaged if they misunderstand a definition containing a noun/verb phrase […], owing to their insufficient knowledge of grammar or syntax" (Beglar, 2010, p. 254).

Achievement Tests

Two achievement tests[4] were designed. Both receptive and productive achievement tests measured participants' ability to recognize (by a multiple-choice test) and to recall (by fill-in-the-blanks) the Persian definitions and English forms of the target words respectively. Each test contained 20 questions (one point for each correct answer); had a maximum score of 20; and was administered as both a pre-test and a delayed post-test (Table 5 and Table 6).

The face validity of the tests was evaluated by submitting them to four applied linguistics professors. They confirmed the suitability of the tests for the participants provided that I made some small changes in the instructions. After pilot-testing the

[4] They are accessible from this link.

Table 4 TFA scores for each task

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

Table 5 Question 4 in the receptive test

4. I hung a **portrait** of my hero on a wall:
a) قاب عکس b) فَرش c) کبریت d) جمجمه

Table 6 Question 6 in the productive test

6. They were poor; they were living in a small **sh**…………

achievement tests by administering them to 9 volunteers who were in the same age group at the same institution but not participants in this study, the internal consistency for each test was calculated by Cronbach Alpha (receptive test = 0.63, productive test = 0.59). To increase the internal consistency of the tests, more adjustments were made to the tests by replacing items that were poorly correlated with other items until an acceptable level of internal consistency (Taber, 2018) was achieved (receptive test = 0.76, productive test = 0.72). The formal participants recruited for the study were not forewarned about the achievement tests.

Think-Aloud Protocols

Concurrent think-aloud protocols (Ericsson & Simon, 1980) were also applied. To reduce possible protocol effects, the instruction was composed carefully (Ericsson & Simon, 1980) using the following words: '*verbalize whatever goes through your mind when you are playing, and it is acceptable to discuss these thoughts with your partner*'. During the task, I was not physically present in the room to observe participants directly. However, I randomly walked into the participants' rooms to remind them of verbalizing their thoughts. I was absent because the presence of the researcher in the same space could impair the validity of the data (Ericsson & Simon, 1980).

Interview

To enrich the elicited concurrent think-aloud data, an "exit interview[5]" (Charters, 2003, p. 73) was also conducted as recommended by mainstream think-aloud researchers (Ericsson & Simon, 1993). Verbalized data may differ in both quality and quantity owing to various factors such as individual differences, the protocols used, etc.; hence, to attain a deeper understanding of the collected concurrent think-aloud data, an exit interview was given (Charters, 2003; Ericsson & Simon, 1993).

Two types of questions were asked in this semi-structured exit interview. The first type included general questions aimed at eliciting more information about the processes and strategies that participants used for either learning or understanding the target words during their tasks. The second type included questions that were adapted from the study by deHaan et al. (2010) and checked or controlled the effect of the multimedia learning factors. The interview was conducted individually and in Persian.

---

[5] It is accessible from this link.

3.2.5 Procedure

This study was conducted in three phases. In the first phase, 63 potential participants sat for the VST, and both receptive and productive achievement tests (pre-test). Next, participants whose vocabulary sizes were lower than 2000 were removed because I wanted to ensure that the selected participants can comprehend the game-guide text. Then, out of the remaining 44 participants, 30 were randomly selected for this study. After that, the participants were asked to select their pairs themselves assuming that it might enhance their willingness to negotiate and their motivation to complete the DGBVL tasks. Then, each pair was randomly assigned to Lo-LIL (N = 10), Mo-LIL (N = 10), and Hi-LIL (N = 10) tasks.[6] Later, two pairs in each group were randomly selected for collecting think-aloud data (N = 12). The participants completed the tasks in pairs because "such tasks are an integral part of second language learning" (Cohen, 1987, p. 90).

In the second phase of the study, the think-aloud pairs were invited to a 30-min warmup session two days before their main task (Ericsson & Simon, 1980). During this session, they were expected to learn how to control the game; cooperate with their partners; become accustomed to being recorded; verbalize their thoughts; and discuss their ideas with each other while playing Chapter 3 of the game. After the warmup session ended, I delivered the target word list to the participants, including non-think-aloud pairs, in the Hi-LIL group and asked them to memorize all 20 words beforehand and they returned the list to me the following day. Non-think-aloud participants in Lo-LIL and Mo-LIL groups were not supposed to do anything before the main task session. Two days after the warm-up session, all participants came to a computer lab. I instructed them for almost 10 min on how to play the game and complete the task. In computer labs in three different rooms, all participants played the game in pairs and tried to complete the first chapter of the game by following the game-guide instructions. The think-aloud pairs' actions and voices were also both video- and audio-recorded. I was there to both assist with technical issues and remind the think-aloud participants to verbalize their thoughts. All participants spent between 70 and 90 min on their tasks because there were many unskippable video cut-scenes in the game.

In the third phase, immediately after the think-aloud participants had finished their tasks, I interviewed them individually for about 20 min each, whereas non-think-aloud participants left the computer lab right after finishing their tasks. The interviews were audio-recorded. The participants were not forewarned about the exit interview because this could have affected their cognitive processes during the main task performance (Ericsson & Simon, 1993). Finally, three weeks after the main task, the delayed-post-tests were administered in the same setting because later retention of target words is a more desirable learning outcome; besides, Schmitt (2010) asserts that "a delayed posttest of three weeks should be indicative of learning which is stable and durable" (p. 157). It should be mentioned that because test items were arranged from easy to difficult, and "receptive tests are much easier than productive

[6] Their groups were also named after their tasks.

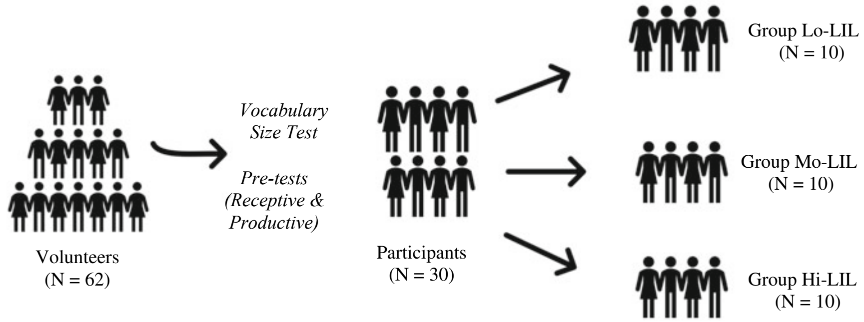

Fig. 3 Procedure (Phase 1)

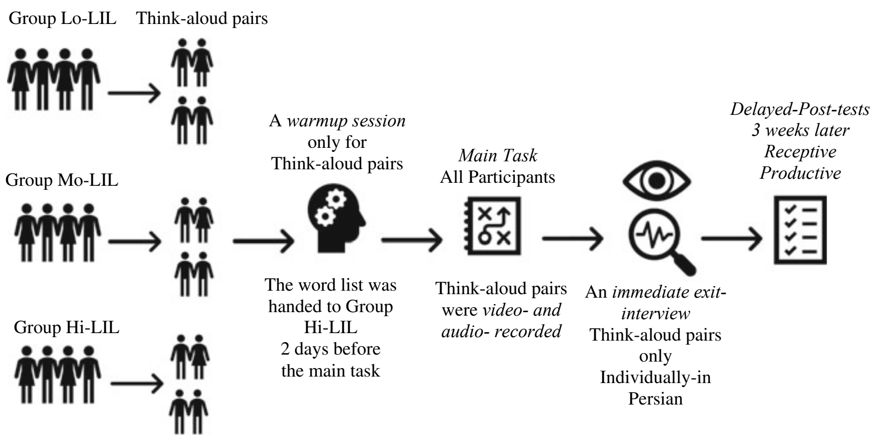

Fig. 4 Procedure (Phase 2 & 3)

tests" (Nation, 2001, p. 33), the receptive test preceded the productive test (Figs. 3 and 4).

3.2.6 Data Analysis

SPSS version 23 was used to implement *Two Related-Samples Wilcoxon Signed Rank Tests* and the *Kruskal–Wallis Test* for analyzing the quantitative data. These formulas were selected because the data were not distributed normally and the number of participants was small (Hall, 2015).

The qualitative data were analyzed by inductive content analysis (Dörnyei, 2007) because there is fragmented knowledge on the effect of DGBVL task-induced LIL (Elo & Kyngäs, 2008). Therefore, the following steps were taken: (1) the sentence was selected as the most suitable unit of analysis; (2) to become familiar with the data, the data were immediately transcribed, read, listened to, and watched several times; (3) recurring patterns and themes in the data sheets were coded; (4) the emerged patterns and their codes were extracted and recorded on separate sheets; (5) related patterns were categorized in the same groups; (6) tags were assigned to the units based on the most recurring pattern in each group; (7) the patterns were abstracted by labeling them after concepts in applied linguistics; (8) to explain the qualitative results, the findings were categorized and explained. To improve the reliability of the qualitative findings, an applied linguistics professor reviewed the whole process to check the quality of the analysis and presentation of findings (Church et al., 2019).

4 Results

To report the results, first, I describe the quantitative findings, and then the qualitative.

4.1 Quantitative Results

The Kolmogorov–Smirnov test showed that the test scores were not normally distributed (See Table 7). Therefore, the nonparametric two Related-Samples Wilcoxon Signed Rank Test and the Kruskal–Wallis Test were used to analyze the quantitative data (Hall, 2015).

RQ1: To What Extent Do DGBVL Tasks Make Significant Differences in Vocabulary Acquisition?

The Two Related-Samples Wilcoxon Signed-Rank Tests were implemented for comparing mean ranks of participants' performance in both productive and receptive pre-tests and delayed-post-tests (Table 8).

The results indicate that the DGBVL tasks have enhanced the acquisition of both receptive and productive knowledge of the target vocabulary items regardless of their differences in LIL with large effect sizes ($r_{\text{(receptive)}} = 0.87$, $r_{\text{(productive)}} = 0.85$).

Table 7 The results of normality test

	Kolmogorov–Smirnov		
	W	*df*	*p*
Receptive Gain Scores (Post–Pre)	0.180	30	0.014
Productive Gain Scores (Post–Pre)	0.149	30	0.038

Table 8 The results of the Wilcoxon signed-rank test

		N	Mean Rank	Sum of Ranks	*Z*	*p*
Receptive	Negative Ranks	0	0.00	0.00	4.790	0.000
	Positive Ranks	30	15.50	465.00		
	Ties	0				
	Total	30				
Productive	Negative Ranks	0	0.00	0.00	4.709	0.000
	Positive Ranks	29	15.00	435.00		
	Ties	1				
	Total	30				

Table 9 Results of the Kruskal–Wallis Test

Tests	Groups	Mean Rank	*N*	χ^2	*df*	*p*
Receptive Gain Scores (Post–Pre)	Group Lo-LIL	12.90	10	9.240	2	0.010
	Group Mo-LIL	11.30				
	Group Hi-LIL	22.30				
Productive Gain Scores (Post–Pre)	Group Lo-LIL	14.35	10	10.687	2	0.005
	Group Mo-LIL	9.75				
	Group Hi-LIL	22.40				

However, one of the participants did not acquire productive knowledge of the target words by completing the DGBVL tasks.

RQ2: To What Extent Does the Difference in DGBVL Task-Induced LIL Make a Significant Difference in Vocabulary Acquisition?
To find the differences among groups' performances in both receptive and productive tests, the Kruskal–Wallis test was used (Table 9).

The results show that, in both receptive and productive delayed-post-tests, group Hi-LIL scored the highest and that Group Mo-LIL surprisingly scored the lowest. Also, their differences at $p \leq 0.05$ are statistically significant with large effect sizes ($\eta^2_{(receptive)} = 0.19$, $\eta^2_{(productive)} = 0.24$) (Fig. 5).

4.2 *Qualitative Results*

Ten patterns emerged through the qualitative analysis of data. Based on their features, patterns were later grouped into two broader categories: *universal moves*, consisting of specific efforts (or moves) taken by *all* participants for completing DGBVL tasks, and *exclusive strategies*, consisting of group-specific strategies that participants implemented mainly for dealing with the target words.

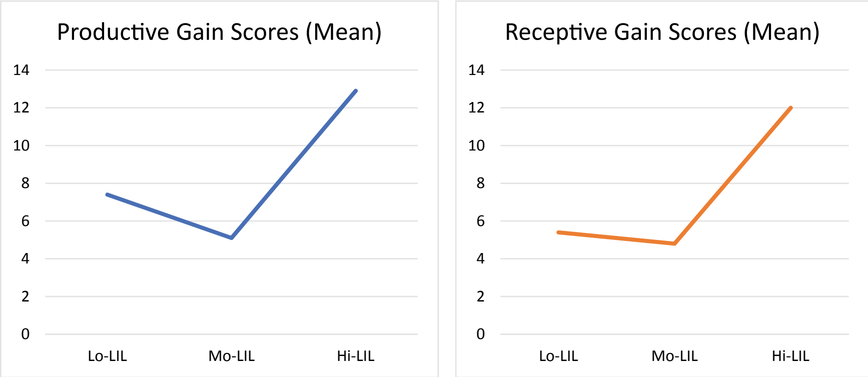

Fig. 5 Productive gain scores (Mean) graphical presentation of group performances in both receptive and productive tests (*Note* Lo-LIL stands for Low level of involvement load; Mo-LIL, Moderate level of involvement load; Hi-LIL, High level of involvement load)

4.2.1 Universal Moves

Five emerged categories that were not exclusive to any group or task were placed in a group called universal moves and labeled as *information search*, *negotiation*, *turn-taking*, *trial-and-error*, and *review*.

Information Search

To complete the DGBVL tasks, participants searched for information about the target words or their progression in both the game-guide and the game. These instances were coded as *information search*. For example[7]:

A: But it noted where to find the glue. Look at this page once more.
B: We must have glue somewhere.
A: Click on the ladder; what does the ladder mean? Click on the glue.
A & B: {In-game information search}.

Negotiation

After finding the correct information, the participants had to develop either strategies or plans for solving puzzles. Hence, participants negotiated their choice of strategy and implemented it. These instances were coded as *negotiation*. For instance, participants are negotiating over the target word *ladder*.

A: Click on the ladder, what was ladder?
B: I can imagine the shape.

[7] All examples are English translations.

A: Wasn't it "leather"?
B: I don't remember the "ladder".
A: {she reads the game-guide}.
B: Maybe, by ladder, it means {Persian word for the the first-aid kit}.

Turn-Taking

After negotiations, participants evaluated their ideas by taking turns. These instances when participants took turns either by controlling the game with mouse clicks or by interacting with the game-guide text were coded as *turn-taking*. In the following example, participants took turns and swapped possession of the mouse.

A: Go, go, go; Give it to me {exchanging the mouse}; {A clicks to test his idea}
B: What did you take?
A: Nothing, it didn't work.
{B takes the mouse and A reads:} Click on the hammer.
B: So, you must click on the hammer.
A: Aaha, do you know what it wants? {He takes the mouse and clicks for testing his idea}

Trial-And-Error

When none of the plans or ideas had succeeded, the final option was trial-and-error. Participants were randomly clicking on the objects on the screen till they click on a specific object by chance. These instances were coded as *trial-and-error*. In the following excerpt, participants relied on trial-and-error to find a door *sign.*

A: click on the door sign
B: Where is the door sign? Isn't that the door sign?
A: {A is clicking randomly and frequently}
B: doooor siiiign {murmuring}; Isn't that the door sign?
A: Nope

In their exit-interview, some of the participants indicated that they had used trial-and-error.

Researcher: Do you use any tricks, when playing a digital game, to help you with unknown words?
A: Trial-and-error
B: We found some words that we didn't read or know before by trial-and-error.

Review

Finally, participants reviewed previously solved problems regularly to reorganize their thoughts. These instances were coded as *review*. In the following example, the participants were reviewing the number and the places of *nails* that they needed to find.

A: Well, it said go back down…
B: Then click on the hammer in your bag, click on the portrait; what was portrait?
A: {Persian definition}
B: Aha
A: Behind that, click also on the plank 1/3 and Isaac's reel; It mentions behind; Ok, we removed its nails either
B: 2/6, we need more
A: Behind that…, Isaac…., we found it too. Ok! we did ém all.

4.2.2 Exclusive Strategies

The differences in the task design led each group to employ strategies that were task-/group–specific. These strategies were specifically for overcoming target word-related challenges such as finding, recalling, and guessing their definitions.

Lo-LIL Task

Participants with the Lo-LIL task employed a specific strategy that I called *input enhancement* strategy when they encountered the target words. In using the strategy, they either voiced the target words or enunciated the Persian definition on the page margin. For example, participant A, who was reading the game-guide in English, enhanced the input by enunciating the Persian definition instead of the *closet.*

A: Click on the hammer in your bag; click on the red {Persian definition of Closet}
B: {Persian definition of Closet}

In the exit interview, the implication of the input enhancement strategy was mentioned.

Researcher: What do you think you have learned about the words in this game? Meanings? Pictures? Spelling?
A: Pronunciations and definitions the most.
B: Both pronunciations and definitions.

The analysis of the think-aloud data indicated a relationship between the universal moves and strategies in the Lo-LIL group (See Fig. 6). The participants interacted firstly with the text of the game-guide to obtain the instructions (*Information search*) (See Fig. 6, step 1). They then analyzed the digital game on the computer screen (See Fig. 6, step 2). When searching for information, the Lo-LIL group participants employed the *input enhancement* strategy to deal with the target words. Then, they *negotiated* over the information that they obtained to plan their future actions. They *took turns* controlling the problem-solving process. After each success, they *reviewed* their actions before moving to the next task. If participants' plans were

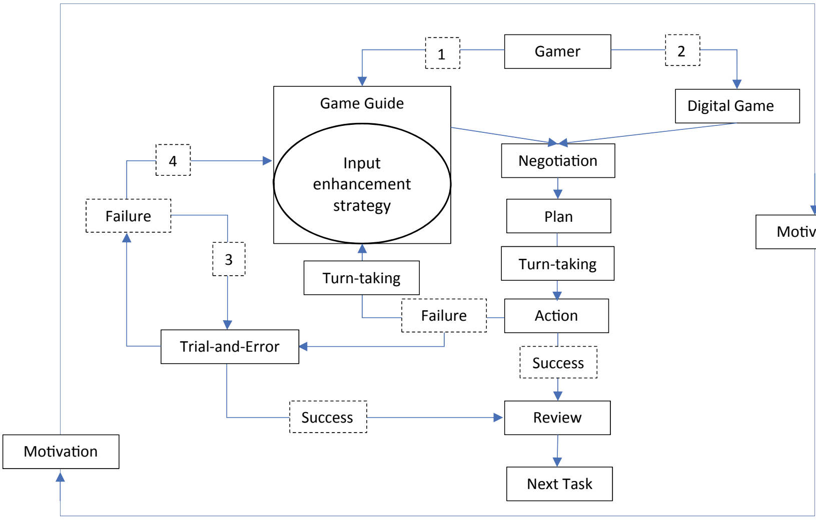

Fig. 6 Lo-LIL group learning approach (*Note* The numbers in Figs. 6, 7, and 8 represent steps taken by participants)

unsuccessful, they performed *trial-and-error*. If their trial-and-error moves failed, they either retried until they succeeded (See Fig. 6, step 3) or occasionally rechecked the game-guide (See Fig. 6, step 4).

Mo-LIL Task

Data analysis showed that the Mo-LIL group participants were employing a strategy called *inferencing from contexts* (Rott & Williams, 2003). In inferencing from contexts, participants were looking for contextual clues in both the game-guide and the game scenes to infer meanings of the target words. For example, participants A and B read the text and considered surrounding contextual information to guess the meaning of the target word *closet:*

A: Click on the hammer in your bag; click on the red closet.
B: Click on the red closet.
A: Red means {Persian definition}; Ok, click here.

Also, the Mo-LIL group participants mentioned this in their exit-interview.

Researcher: Do you use any tricks when playing a digital game to help you with unknown words?
A: For example, when we read click on something and it said it was red or yellow, it was helpful to guess what it meant; then, we looked for red things only on the screen.
B: I used the game-guide sentences near the unknown words.

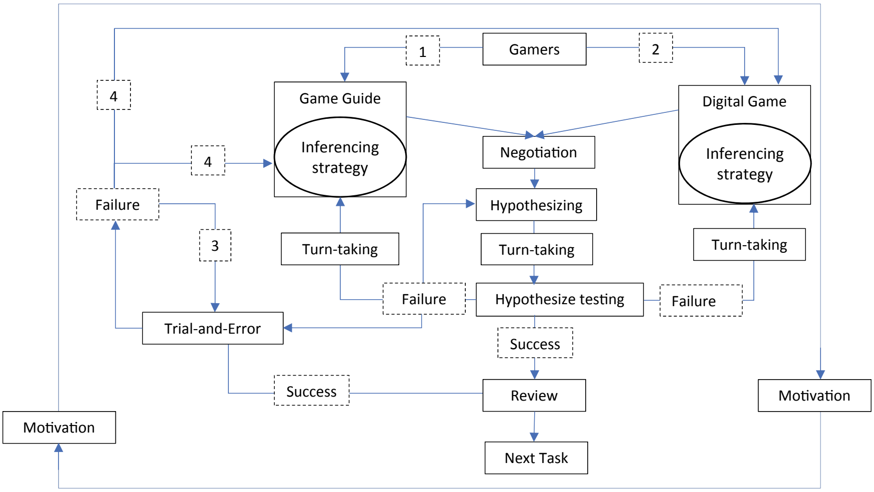

Fig. 7 Mo-LIL group learning approach

There was also a relation between Mo-LIL participants' moves and strategies (see Fig. 7). They first *searched for information* in the game-guide (See Fig. 7, step 1) then in the game (See Fig. 7, step 2). Meanwhile, they were trying to *infer* proper meanings by contextual clues. Next, they *negotiated* their guesses and developed hypotheses. Then, they tested their hypotheses in random *turns*. If they were successful, they *reviewed* their actions and moved to the next tasks. If they failed, they tested the rest of their hypotheses; referred to both the game-guide and the game to develop new hypotheses; or relied on the *trial-and-error* move as their final option. If the *trial-and-error* failed, they either repeated it (See Fig. 7, step 3) or referred to the game-guide and the game (See Fig. 7, step 4) for obtaining more information.

Hi-LIL Task

Hi-LIL group participants used *memory search* and *feedback request* strategies. They tried *memory search* first when they encountered the target words. For example, participants A and B were searching their memories for the word *latch.*

A: *Click on the latch. What was the latch? {Silently Thinking} Aha! It means {Persian Definition}.*
B: What was the latch? {while thinking, she whispered} Latch, latch, latch {Persian Definition}.

However, sometimes, to be reassured of their memory-search outcomes or recall the meanings, participants *requested feedback* from their partners.

A: What was the hook?

B: Hook?
A: {Thinking silently}; I am not sure, does it mean {Persian Definition}?
B: Yeah, I think it was {Persian Definition}.

They also mentioned this in their exit-interview.

Researcher: What did you do if you couldn't remember/recognize the meaning of the unknown words in this game-guide?
A: I had to either think a lot about that to recall its definition or ask my partner to progress.
B: I asked my partner for help.

Moreover, they indicated the implication of a specific strategy that I call *word association* to deal with unknown words other than target words in their exit-interview. In implementing this strategy, they referred to the semantic association of contextually known words to infer either what an unknown word meant or which object they had to use.

Researcher: What did you do when you met an unknown word in the game-guide of this digital game?
A: Apart from the words we read, we found new words by considering other words and objects.
B: For example, when we didn't know what to do with the shovel when we found soil, we guessed "yeah, we must use the shovel".

There was also a relation between Hi-LIL group participants' moves and strategies (See Fig. 8). Their main source of *information search* was the game-guide. They used *memory search*, or *feedback request* strategies to deal with the target words; moreover, they used *word-association* strategy to deal with other unknown words. After reading the game-guide (See Fig. 8, step 1) and checking the game (See Fig. 8, step 2), they *negotiated* to develop plans and *took a turn* to implement them. Next, if their plans were successful, they would *review* their previous actions and move to the next task. Otherwise, they *took turns* and read the game-guide to develop new plans. Rarely (dashes in Fig. 8 imply this) did they solve the problems by the *trial-and-error* move (See Fig. 8, step 3 & 4).

4.2.3 DGBVL Tasks as Motivational Sources

The analysis of individual exit interview data showed that the DGBVL tasks were also highly motivating. Factors such as story, challenge, and curiosity motivated participants to complete their DGBVL tasks enthusiastically.

Researcher: What do you focus on when you play a digital game?
Lo-LIL (A): To win, to find the objects and move to the next levels, to see the end of the story.
Lo-LIL (B): Trying to solve the mystery: whose skeleton was that in the room?
Mo-LIL (A): To win!

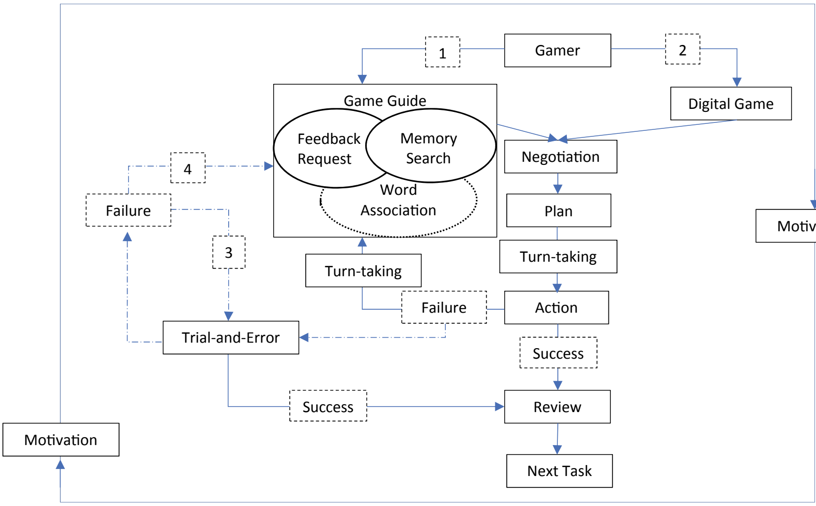

Fig. 8 Hi-LIL group learning approach

Mo-LIL (B): Solving problems and puzzles, opening the doors, and seeing the end.
Hi-LIL (A): I wanted to see the end and find out what would happen.
Hi-LIL (B): What happens in the end, what was behind the red garage door?

5 Discussion

This study investigated the effectiveness of the DGBVL task-induced LIL on vocabulary acquisition. In answering the first research question, it was found that the *three* DGBVL tasks enhanced IVL. This outcome not only supports the previous findings (Ebrahimzadeh & Alavi, 2016; Janebi-Enayat & Haghighatpasand, 2019; Rasti-Behbahani & Shahbazi, 2020; Sundqvist, 2019) but also adds that DGBVL tasks, regardless of their differences in LIL, can be supportive of IVL. An explanation for this outcome can be the role of context. Previous studies indicate that digital games present content via rich images, animations, visuals, audio, and interactive dialogues (Hwang & Wang, 2016). Hence, in this study, those inherent elements might have created a context that was supportive of IVL by providing opportunities such as frequent and contextualized exposures to and multimodal presentation of the target words (Ebrahimzadeh & Alavi, 2016; Janebi-Enayat & Haghighatpasand, 2019). Qualitative findings support this argument by showing that context assisted Mo-LIL and Hi-LIL participants in either inferring or recalling some of the target words' meanings.

Another explanation can be the role of motivation. It was discussed that the inherent elements of digital games could also promote motivation and keep learners continuously active in a DGBVL task (Pivec et al., 2003; Rasti-Behbahani, 2021). Moreover, Ebrahimzadeh and Alavi (2016) found that those elements correlated with effective IVL. Besides, qualitative findings in this study showed that elements such as story, challenge, and curiosity motivated the participants and kept them active. Hence, the DGBVL tasks might have been motivating enough to enhance participants' acquisition of the target words.

In answering the second question, results showed that effective IVL gains in DGBVL tasks, as discussed by Ali Mohsen (2016) and Reynolds (2017), highly depend on task design and task-induced LIL. However, results showed that inducing a higher LIL does not guarantee a higher vocabulary gain in DGBVL tasks. For instance, Mo-LIL group participants' rate of target word acquisition was the lowest. Also, the results suggest that the search component (Reynolds, 2017) must be optimized in DGBVL tasks because it may hinder IVL as the Mo-LIL task did. There is an explanation for this finding. The qualitative results showed that the Mo-LIL task engaged participants fully in search for meaning; hence, they needed to employ inferencing techniques that could invoke several mental processes (Rott & Williams, 2003). However, some of those invoked processes are unnecessary; can possibly increase the complexity of a task (Yang et al., 2020); and lead to confusion (Martínez-Fernández, 2008). Furthermore, inferencing techniques do not always initiate form-meaning links (Martínez-Fernández, 2008). Thus, because of the unoptimized search component in the Mo-LIL task and employing inferencing techniques, participants may have been lost in their searches; puzzled by many alternatives; and may have floundered in their decision-making and hypothesis-testing. Therefore, they did not manage to invoke the necessary processes for form-meaning link generation. Additionally, the existence of possibilities such as trial-and-error might have reduced any chances of form-meaning link formation because trial-and-error could have assisted the Mo-LIL participants to solve puzzles. Consequently, they did not feel the need to engage in searching for or inferring meanings. Thus, the target words may not have been processed richly enough to be effectively retained by Mo-LIL participants.

In contrast, Hi-LIL group participants acquired the target words more effectively. Based on the think-aloud data, a possible reason is that they employed the most appropriate strategies. According to Rott and Williams (2003), both the choice and quantity of strategies used can be the defining factors in effective vocabulary gain in vocabulary acquisition tasks. The qualitative findings show that Hi-LIL participants not only used more strategies (quantity) than other groups but also employed possibly strategies that reduced their mental load and assisted them in their searches for meanings effectively (quality). Hence, Hi-LIL group participants acquired the target words effectively because they possibly had qualitatively and quantitatively enough opportunities to process both forms and meanings.

Another explanation for the superiority of the Hi-LIL group can be the effect of frequency of exposure. Frequency of exposure, or repetition, is known as an effective factor in vocabulary learning (Folse, 2006; Nation, 2001; Schmitt, 1997, 2010). In this study, the Hi-LIL participants had a greater chance of both frequent and spaced exposure to the target words than the other participants. They were exposed to them before the main task in their target word list, in sample sentences, and later in the game-guide. The higher frequency of exposure might have increased their level of awareness which in combination with an appropriate LIL could promote vocabulary acquisition (Martínez-Fernández, 2008). Moreover, it might have limited the scope of participants' searches for meaning which thereby could have optimized the component of search and reduced the complexity of the Hi-LIL task. Thus, the participants had more opportunities to process the target words and to become aware of their different aspects and features. This in turn could lead to deeper memory traces; generate stronger form-meaning connections; and help the learners acquire the target words much more effectively (Hulstijn & Laufer, 2001; Rott & Williams, 2003).

Finally, like studies that question the validity of ILH, the findings in this study partially support the premise of ILH (Folse, 2006; Keating, 2008; Kim, 2011; Martínez-Fernández, 2008). Moreover, this study supports the argument about the inaccuracy of quantifying methods of LIL (Keating, 2008; Kim, 2011) because although the DGBVL task-induced LIL was scored by TFA, a more precise method of measuring, the results showed that currently available ILH quantifying methods are still inaccurate. Hence, it seems reasonable to argue that the performance of the group Mo-LIL participants was poor because LIL was lower in group Mo-LIL's task than in the tasks of groups Lo-LIL and Hi-LIL. However, this was not revealed owing to the inaccuracy of the indexing method.

6 Conclusion

The generalizability of the outcomes of this study was confined by its limitations. The main limitation concerns the target words. In the present instance, the target words were selected only from concrete nouns. Thus, caution is advised in generalizing the findings of this study to other lexical classes and abstract nouns. The influence of the lexical category is a possible topic for future studies. Moreover, the low number of target words and participants were other limitations that can be addressed in future studies.

Overall, this study shows that the rate of incidental vocabulary acquisition from DGBVL tasks highly depends on LIL. Moreover, the DGBVL task design, target word characteristics, invoked mental processes, the type of search for meaning, and strategy choices can influence the effectiveness of DGBVL tasks-induced LIL (Rott & Williams, 2003; Ali Mohsen, 2016; Yang et al., 2020). Such factors seem to influence the complexity of DGBVL tasks (Yang et al., 2020) and persuade learners both to invoke different cognitive processes (Ali Mohsen, 2016) and to follow different learning approaches for completing their DGBVL tasks. Therefore, to induce an effective LIL and achieve a desirable outcome from a DGBVL task, implementing a proper vocabulary learning technique in designing a DGBVL task is critical. Moreover, because DGBVL and pencil-and-paper contexts are not identical, studying the effect of various vocabulary learning techniques is also recommended.

Pedagogically, inferencing techniques are not encouraged in the DGBVL task designs even though they can induce a relatively high LIL (Rott & Williams, 2003). On the contrary, applying pre-teaching techniques and introducing the target words, their meanings, and their uses combined with higher exposures to target words are highly recommended. Moreover, implementing game-guides creatively, like in this study, can provide precious opportunities for both researchers and teachers to optimize DGBVL tasks for their respective purposes.

Acknowledgements I would like to express my sincere gratitude to Mrs. Maryam Shahbazi for her invaluable assistance in conducting this study.

References

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.

Bahari, A. (2020). Game-based collaborative vocabulary learning in blended and distance L2 learning. *Open Learning: The Journal of Open, Distance and e-Learning, 1–22.* https://doi.org/10.1080/02680513.2020.1814229

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing, 27*(1), 101–118.

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice, 12*(2), 68–82.

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.

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), E104–E107. https://doi.org/10.1111/j.1467-8535.2012.01295.x

Church, S. P., Dunn, M., & Prokopy, L. S. (2019). Benefits to qualitative data quality with multiple coders: Two case studies in multi-coder data analysis. *Journal of Rural Social Sciences, 34*(1), 1–14.

Cohen, A. D. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), *Multilingual matters, No. 30. Introspection in second language research* (pp. 82–95). Multilingual Matters.

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford University Press.

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.

Ebrahimzadeh, M., & Alavi, S. (2016). Motivating EFL students: E-learning enjoyment as a predictor of vocabulary learning through digital video games. *Cogent Education, 3*(1), 1–14.

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implication for CALL. *International Journal of Computer Assisted Language Learning (CALL), 8*(2–3), 103–128.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* MIT Press.

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

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.

Hall, J. L. (2015). *A guide to doing statistics in second language research using SPSS and R.* Routledge.

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

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

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

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.

Jahangiri, K., & Abilipour, I. (2014). Effects of collaboration and exercise type on incidental vocabulary learning: Evidence against involvement load hypothesis. *Procedia-Social and Behavioral Sciences, 98*, 704–712.

Janebi-Enayat, M., & Haghighatpasand, M. (2019). Exploiting adventure video games for second language vocabulary recall: A mixed-methods study. *Innovation in Language Learning and Teaching, 13*(1), 61–75.

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

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

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 100–140.

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

Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task, and type of item. In *Selected proceedings of the 2007 second language research forum* (pp. 210–228). Cascadilla Press.

Martinez, L., Gimenes, M., & Lambert, E. (2022). Entertainment video games for academic learning: A systematic review. *Journal of Educational Computing Research, 60*(5), 1083–1109. https://doi.org/10.1177/07356331211053848

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (pp. 211–232). Second Language Acquisition and Testing in Europe Monograph Series 1.

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

Nation, I. S. P., & Webb, S. A. (2010). *Researching and analyzing vocabulary.* Heinle: Cengage Learning Boston, MA.

Nation, I.S.P (2012). The vocabulary size test. https://www.victoria.ac.nz/lals/about/staff/publications/paulnation/Vocabulary-Size-Test-information-and-specifications.pdf on 13 october 2017.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003, July). Aspects of game-based learning. In *3rd International Conference on Knowledge Management, Graz, Austria* (pp. 216–225).

Rasti-Behbahani, A. (2017, March 21–23). *The effect of video games on learning lexical word classes* [Conference presentation]. Nordic Educational Research Association Conference 2017, Aalborg University.

Rasti-Behbahani, A., & Shahbazi, M. (2020). Investigating the effectiveness of a digital game-based task on the acquisition of word knowledge. *Computer Assisted Language Learning, 1–25*. https://doi.org/10.1080/09588221.2020.1846567

Rasti-Behbahani, A. (2021). Why digital games can be advantageous in vocabulary learning. *Theory and Practice in Language Studies*, *11*(2), 111–118. https://doi.org/10.17507/tpls.1102.01

Reinders, H., & Wattana, S. (2012). Talk to me! Games and students' willingness to communicate. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 156–188). Palgrave Macmillan.

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.

Rott, S., & Williams, j. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language, 15*(1), 45–75.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199–228). Cambridge University Press.

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

Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 237–257). Cambridge U.P.

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Journal of Language Learning & Technology, 23*(1), 87–113.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning, 34*(5–6), 751–777.

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), 1–21. https://doi.org/10.1016/j.compedu.2020.103808

Yanagisawa, A., & Webb, S. (2021). Involvement load hypothesis plus: Creating an improved predictive model of incidental vocabulary learning. *Studies in Second Language Acquisition, 44*(5), 1279–1308. https://doi.org/10.1017/S0272263121000577

Yule, G. (2014). *The study of language* (5th ed.). Cambridge University Press.

Amin Rasti-Behbahani is a postdoctoral researcher at the department of languages and communication studies, University of Jyvaskyla. He is also an English lecturer at the college of arts and applied sciences – foundation program, Dhofar University. His main research interests are digital game-based language learning and vocabulary acquisition.

Gaming in a Foreign Language: L2 Vocabulary Processing in a Single-Player Role-Playing Game

Kevin Reay Wrobetz

Abstract Commercial role-playing games (RPGs), when played in a target language (TL), represent a unique form of media which may be conducive to learning the TL. However, the gameplay mechanics of commercial RPGs are not explicitly designed to coincide with the most pedagogically effective second language acquisition (SLA) processes. Indeed, players who wish to utilize commercial RPGs to improve their linguistic proficiency in the TL may benefit from modifying how they interact with the game content both during and after gameplay. For such players who do not have experience with playing a commercial RPG in the TL, concrete information on how gaming behavior affects linguistic gains in the TL is vital. To examine how commercial RPGs may affect TL vocabulary acquisition, Square Enix's *Final Fantasy IX* was played by a native English speaker in Japanese to analyze (1) the type of vocabulary learned during gameplay, (2) the ratio of vocabulary learned to the total gameplay time, and (3) how the vocabulary learned during gameplay was reactivated during and after gameplay.

Keywords Digital game-based learning (DGBL) · Second language acquisition (SLA) · Vocabulary acquisition · *Final Fantasy IX* · Role-playing games (RPGs)

Proficiency in a language is much more than the sum total of knowledge about grammar and vocabulary. Indeed, as has been suggested in several studies, language proficiency is achieved by using the acquired linguistic knowledge of grammar and vocabulary to interact with both native/proficient target language (TL) speakers and the culture of the TL (e.g., Gass, 2010; Mackey et al., 2013). Serving the role as the primary medium of communication for nearly every aspect of social interaction, proficient TL use fundamentally requires the speaker to interact with and participate in the social structure with which the TL is associated (Kinginger, 2010; Norton, 1997). Fortunately, we are currently living in the age of information which provides easy access to a diverse range of media in any number of languages, and, as research

K. R. Wrobetz (✉)
Department of Business Administration, Kobe Gakuin University, Kobe, Japan
e-mail: kwrobetz@ba.kobegakuin.ac.jp

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_7

has suggested, digital games are one type of media that can be used as a language-learning tool (Lai et al., 2012; Nietfeld, 2017; Peterson, 2010; Peterson et al., 2022). Furthermore, interaction with digital games in the TL may aid SLA by providing the impetus to make meaningful social connections to the TL (Cornillie et al., 2012; Sørensen & Meyer, 2007). This study specifically analyzed how single-player role-playing games (RPGs), when played in the TL, may be used to elicit the acquisition and retention of TL vocabulary. This study also investigated how the social sphere of the TL may affect the reactivation and long-term retention of TL vocabulary acquired through gameplay.

When considering the possibility of utilizing RPGs as a language-learning method outside of the classroom, even in environments isolated from the social sphere of the TL, digital RPGs played in the TL represent a pedagogically potent SLA tool from both sociocultural and psycholinguistic perspectives. From a sociocultural perspective, the interactive gameplay mechanics, particularly in massively multiplayer online role-playing game (MMORPGs), may benefit SLA (Bryant, 2006; Peterson, 2012; Rankin et al., 2009; Sylvén & Sundqvist, 2012). This SLA mechanism can come from inter-player linguistic exchanges required to progress the game forward such as might occur with group-oriented quests in MMORPGs, a process in line with the social aspects of learning at the foundation of sociocultural theory (Lantolf, 1994). From a psycholinguistic perspective, digital games which are conducive to "spaced gaming," or spreading the total gameplay interval over extended periods of time, may potentially lead to better long-term retention of vocabulary acquired through gameplay by capitalizing on well-researched memory benefits such as the spacing effect (Appleton-Knapp et al., 2005; Bjork, 2014; Janiszewski et al., 2003). Single-player RPGs, especially titles which are focused on a central narrative, are particularly conducive to spaced gaming due to the long number of hours typically required to play through a game. The single-player RPG *Final Fantasy IX* (Square Enix, 2000) used in this study requires, when averaging several different gameplay styles, approximately 48 h to complete (HowLongToBeat, 2021). The potential benefits to SLA that have been identified in RPG gameplay mechanics coupled with long gameplay intervals make commercial single-player RPGs such as *Final Fantasy IX* compelling candidates to serve as efficacious language-learning tools.

The efficacy of any potential pedagogical benefits garnered from playing RPGs to increase TL proficiency should by no means be considered as universal for all players. Commercial games are first and foremost games and are not necessarily designed to be efficacious language-learning interventions. It is therefore important for research to identify how the SLA process functions when playing a single-player RPG outside formal educational environments. To this end, the author of this research (a native English speaker) played the single-player RPG *Final Fantasy IX* (FFIX) in Japanese to investigate how the gameplay mechanics affected the acquisition/retention of Japanese vocabulary acquired from gameplay and how the vocabulary was reactivated both during and after gameplay. The research was guided by the following research questions: (1) What type of vocabulary will be learned from playing FFIX? (2) How will vocabulary acquired from playing FFIX be reactivated? and (3) What factors will influence the retention of vocabulary acquired from playing

FFIX? The following section details how this study was structured to answer these research questions.

1 Method

1.1 Positionality of the Author

It is the position of the author that digital games present both foreign language educators and learners with unique and efficacious options to improve their skills in the TL. In particular, the past research of the author has focused on how RPGs, when played in the TL, may elicit TL vocabulary acquisition and retention. With that in mind, the goal of the present study was not to examine the efficacy of single-player RPGs to elicit TL vocabulary acquisition and retention but rather to examine the processes involved in vocabulary acquisition and retention when playing a single-player RPG in the TL (for the case of the author, in Japanese). The author did not have any preconceived hypotheses about the outcome of this study and attempted to construct the tools to measure how vocabulary acquired through gameplay is reactivated and retained in as unbiased a manner as possible.

1.2 Game Selection

To select the most pedagogically potent single-player RPG title, the following factors were considered: (1) both the current linguistic level and the linguistic potential of the player, (2) the potential of the game title to invoke internalized motivation to continue playing, and (3) accessibility of the digital game title on mobile platforms to increase the frequency of sessions. Square Enix's *Final Fantasy IX* was chosen as this study's focus for three reasons. The first reason was the author's familiarity with both FFIX and other titles in the Final Fantasy series. Having played through FFIX in English upon its release in North America in 2000 and a handful of other Final Fantasy titles in the late 1990s, both the story and gameplay mechanics were well understood at the start of this study. While familiarity with the linguistic content of FFIX makes for familiar gameplay, the bulk of the linguistic content featured in the story and dialogue of the game remained outside of the author's productive skill level in Japanese. The second reason was the deep level of nostalgic attachment that the author had towards the game. The third reason was the widespread access of the game on gaming platforms including a number of mobile gaming platforms. The platform chosen for this study was the Nintendo Switch to capitalize on the benefits of "dead time" (Stockwell, 2010, p. 106). The above-mentioned criteria were all considered by the author to be desirable features to maximize the potential for TL comprehension, increase the length of TL exposure, and elicit higher frequencies of gaming sessions.

1.3 Japanese Proficiency and Potential

When looking at how single-player RPGs affect vocabulary processing, the player's level of TL proficiency, characterizing innate linguistic ability in the TL, as well as the relationship of the player to the TL social sphere, characterizing the linguistic potential in the TL through interaction with more capable peers (usually native speakers of the TL), must be considered (Pulido, 2003; Serrano et al., 2012). At the time of this study, the author had achieved a Japanese level of advanced high to superior as described in the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines and had passed the N1 qualification of the Japanese Language Proficiency Test (JLPT) (Breiner-Sanders et al., 2000). The author had lived and worked in Japan from 2011 to 2013 and from 2015 onward to the time of this study, regularly engaged in the technical aspects of Japanese linguistics through professional activities such as translation, and was married to a Japanese national. The author's social sphere therefore ensured regular contact with numerous native Japanese speakers (i.e., more capable peers) and media sources throughout this study.

1.4 Cataloging Vocabulary and Scoring In-Game/Out-of-Game Reactivation

To ascertain how single-player RPGs affect TL vocabulary processing, addressing research question one regarding what type of vocabulary will be learned playing FFIX, the author used dictionaries and online resources to look up the meaning and phonetic reading to unknown Kanji (Japanese logograms adopted from Chinese) for every unknown word which came up during 94 h of gameplay over a period of one year from March 2019 to March 2020. All 1,311 words acquired from gameplay were catalogued chronologically into 13 lists of 100 words each (with the 14th list being composed of 11 words) to track the time of acquisition and how the reactivation of the vocabulary was correlated to the time of acquisition. Therefore, as unknown words were encountered during gameplay, they were immediately assigned to a vocabulary list on a chronological basis. Once a vocabulary list was populated with 100 words, a new vocabulary list was started. The vocabulary lists were grouped into lists of 100 to chronologically track when the words were acquired during gameplay as well as to serve as groupings for delayed vocabulary tests to further track how well the acquired words were retained. Once catalogued into a chronological vocabulary list, each word was divided into one of three vocabulary types: "Review Vocabulary A," "Review Vocabulary B," and "Novel Vocabulary." Review Vocabulary A refers to words for which the phonetic reading of the Kanji was known but the English definition was not (e.g., 思い上がる, whereby the author knew the correct phonetic Kanji reading of おもいあがる but not the English definition of "to be conceited"). Review Vocabulary B refers to words whose English definition was known, but the phonetic Kanji reading was not (e.g., 育む, whereby the author knew the English meaning "to rear" but not the correct phonetic Kanji reading of はぐくむ). Novel Vocabulary refers to words whose phonetic Kanji reading and English definition were both unknown (e.g., 晒す, whereby the author neither knew the phonetic Kanji reading of さらす nor the English definition of "to expose"). To address research question two, the vocabulary lists were monitored for vocabulary reactivation.

Vocabulary reactivation refers to any moment that the author noticed any catalogued word again either during gameplay (in-game reactivation) or during daily life (out-of-game reactivation). The author did not review the catalogued vocabulary lists at any time during this study and strived to track the reactivation of catalogued vocabulary with a scoring system which would reflect the natural noticing of acquired vocabulary in contexts different to that in which they were learned. Once a word had been catalogued as being reactivated, the instance of reactivation as well as the circumstance of the reactivation (namely in-game or out-of-game reactivation) were recorded by assigning the catalogued vocabulary a reactivation score of 1. For instances of out-of-game reactivation, the specific circumstances of each reactivation were also catalogued (e.g., while engaging with TL media). Once each word was scored for reactivation, scoring for other instances of reactivation was not allowed. For example, a catalogued word that had come up more than once during gameplay only had a reactivation score of 1 no matter how many times that particular vocabulary came up during gameplay. Similarly, vocabulary could not achieve an out-of-game reactivation score higher than 1 even if the reactivated vocabulary came up numerous times in different circumstances outside of gameplay (e.g., in a Japanese news broadcast and during conversation with a Japanese native). The only exception to this rule is with duplicate vocabulary or otherwise words that were catalogued into the vocabulary lists more than once. These duplicate vocabulary entries occurred when the definition and/or phonetic Kanji reading of a previously catalogued word was forgotten. If a forgotten word came up again during gameplay and was identified as an unknown word, the definition and/or phonetic Kanji reading were looked up and subsequently catalogued into a vocabulary list. Upon cataloguing new words into the vocabulary lists, an autonomous search was automatically run on all the previous vocabulary lists to search for identical entries (i.e., duplicate vocabulary). For words that were catalogued more than once (i.e., duplicate vocabulary), an automatic in-game reactivation score was assigned to only one of the duplicate words, meaning that other duplicate vocabulary was excluded from further in-game reactivation scoring. However, if duplicate words that were excluded from in-game reactivation came up outside of gameplay (e.g., in a Japanese news broadcast), then those duplicate words were allowed to receive an out-of-game reactivation score of 1. In this sense only duplicate words could receive a maximum reactivation score of 2, a score of 1 for in-game reactivation and another score of 1 for out-of-game reactivation.

The vocabulary reactivation scoring system described above which limits the reactivation score to 1 for all vocabulary except for duplicates was implemented to avoid artificially inflated reactivation scores. The aim of the reactivation scoring system employed in this study was not to measure how many times the acquired TL vocabulary was recalled but rather to measure the potential sources of reactivation (e.g., in-game reactivation versus out-of-game reactivation). Many of the words used during the game are repeated when the game mechanics that prompt those specific words are repeated. Bearing this in-game repetition of specific vocabulary in mind, should the reactivation scoring method have allowed for multiple scoring on a single word, then the in-game reactivation scores would have been inflated to a number that might not accurately represent true reactivation of vocabulary. Specifically, if a catalogued word is used throughout the game to perform a task specific to disproportionately frequent game mechanics (e.g., a word that is present every time the player opens an option menu), then a reactivation scoring method that accounts for this frequency of in-game reactivation could be artificially inflated by many thousands of occurrences. Similarly, if the out-of-game reactivation scoring system allowed for catalogued vocabulary to receive a reactivation score when the TL vocabulary acquired during gameplay was used by the author outside of gameplay, then the possibility exists for the author to employ the same TL vocabulary acquired during gameplay numerous times in the same contexts outside of gameplay (e.g., in a conversation with a native speaker), resulting in a reactivation score with significantly inflated out-of-game reactivation scores by way of selective word choice.

1.5 Delayed Vocabulary Testing and Quantitative Analysis Procedure

To measure the effects of vocabulary reactivation (or lack thereof), addressing research question three regarding which factors will affect the long-term retention of vocabulary acquired through gameplay, 14 delayed vocabulary tests were employed for each of the 14 vocabulary lists. Each of the 14 delayed vocabulary tests were taken on separate days between the period of March 3, 2021 and March 21, 2021. The period of delay therefore ranged from 24 months after the compilation of vocabulary list one in March 2019 to 12 months after the compilation of vocabulary list 14 which marked the conclusion of gameplay in March 2020. The vocabulary tests consisted of computer-generated, randomized questions (so as to avoid reactivation from author-created tests) prompted with the Japanese vocabulary and required the correct English definition for Review Vocabulary A, the correct phonetic Kanji reading for Review Vocabulary B, and both the correct English definition and phonetic Kanji reading for Novel Vocabulary. Correctly answered questions were scored with a value of 1, and partially correct answers were scored with a value of 0.5. Answers to questions using Review Vocabulary B could only be scored as being either correct or incorrect, but answers to questions using Review Vocabulary A or Novel Vocabulary could be scored as being partially correct. Since Review Vocabulary A required a correct English definition, answers which were judged to be close to the English definition provided for each catalogued word were scored as being partially correct. For example, the vocabulary test response to the question prompted with the original vocabulary entry "漸く" was "in the end" which is arguably different from the correct definition of "finally, at last" but is similar enough to the correct definition so as not to be scored as being entirely incorrect. Answers to questions using Novel Vocabulary, requiring both correct English definitions as well as correct phonetic readings of each Kanji, could be counted as being partially correct if at least one of the required parts of the definition (i.e., the English definition or the phonetic Kanji reading) was correctly answered. The delayed vocabulary tests were autonomously scored with a pre-programmed spreadsheet and reviewed by the author for the correctness of the registered score. Answers which were scored as being incorrect by the computer program but which were determined to be correct were manually adjusted by the author.[1] The results of these 14 delayed vocabulary tests were then subjected to quantitative analyses.

To analyze the effects of the reactivation process, again addressing research question three regarding which factors will affect the long-term retention of vocabulary acquired through gameplay, percentages of correct answers for reactivated vocabulary and non-reactivated vocabulary from each of the 14 vocabulary lists were compared with statistical analyses. All duplicate vocabulary were removed from the statistical analyses to avoid artificially inflating the vocabulary test scores of words which were catalogued only once throughout all 14 vocabulary lists. However, the delayed test results from the initial word catalogued in any duplicate vocabulary grouping were included in the final analyses. The test results for the removed duplicate vocabulary were presented in the results section as an independent variable. The test results for reactivated vocabulary and non-reactivated vocabulary from each of the modified 14 vocabulary lists were subjected to two-sample, two-tailed *t*-tests to analyze the test results for statistically significant differences in each of the vocabulary types (Review Vocabulary A, Review Vocabulary B, and Novel Vocabulary) as well as for the aggregate vocabulary sample. Finally, linear regression analyses were carried out on each of the vocabulary categories as well as on the aggregate vocabulary sample to analyze for statistically significant correlations between the time period till testing and the vocabulary test results for reactivated and non-reactivated vocabulary. The following section will detail the results from the data analyses presented above.

[1] The vocabulary test data including the vocabulary lists, vocabulary types, reactivation scores, responses to the vocabulary test questions, and vocabulary test scores have been made available via this hyperlink https://drive.google.com/file/d/1YriNiHqPTxPAgdVhb7OhqzyvovN_4SYs/view?usp=sharing.

2 Results

2.1 *Vocabulary Type Acquired Through Gameplay*

The breakdown of vocabulary types from each of the 14 vocabulary lists is presented in Fig. 1, and the total percentage of each vocabulary type from all 1,311 words is presented in Fig. 2. The majority of the unknown vocabulary encountered throughout the year-long, 94-h process to complete FFIX is made up of Review Vocabulary A and B for a total of 74.29% (including Review Vocabulary A, 62.7%, and Review Vocabulary B, 11.59%), leaving only 25.71% of the vocabulary to make up the Novel Vocabulary type. There were, however, fluctuations throughout gameplay, the most notable examples of which being from vocabulary lists 1, 5, and 14 with Novel Vocabulary percentages reaching 38%, 40%, and 45% respectively. Overall, the majority of catalogued vocabulary were words and phrases that the author had encountered previously through formal study of Japanese and interaction with the TL culture.

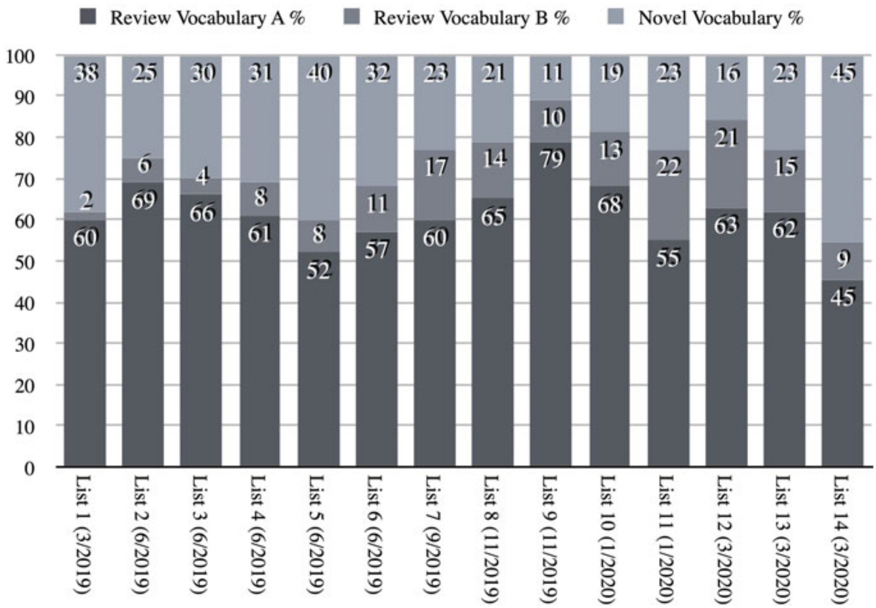

Fig. 1 Percentages of vocabulary type per vocabulary list

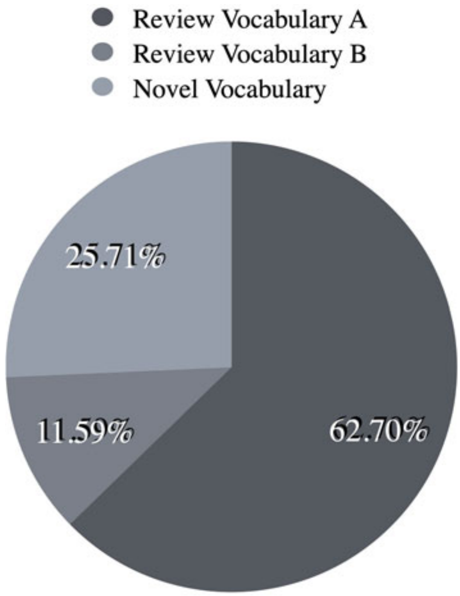

Fig. 2 Total sample percentages of vocabulary type (reactivated and non-reactivated)

2.2 Ratio of Acquired Vocabulary to Total Gameplay Time

The average ratio of acquired vocabulary to the amount of time spent playing the game was approximately 13.95 words per hour of gameplay. There are unique features of single-player RPG gameplay that must be considered when looking at the TL vocabulary load put on the player. In contrast to other sources of TL input such as textbooks, literature, or film in which the stream of TL input remains constant when interacting with the TL source, RPGs tend to have linguistically up-regulated intervals of TL input during story-heavy portions of gameplay and linguistically down-regulated intervals during portions of the game primarily centered on the execution of gameplay mechanics that do not require linguistic input (e.g., physically moving the characters from point A to point B). Such intervals of linguistic up-regulation and down-regulation can be seen in the x-axis time indices in Fig. 1. There are relatively longer intervals of time between the time indices on vocabulary lists 1 and 2 (3/2019–6/2019), 6 and 7 (6/2019–9/2019), 7 and 8 (9/2019–11/2019), 9 and 10 (11/2019–1/2020), and 11 and 12 (1/2020–3/2020) during which the TL input from the game has been down-regulated enough to require comparatively more time to reach the 100 word limit for a single vocabulary list. Intervals of linguistic up-regulation during gameplay can be seen in lists 2–6 (6/2019), lists 8 and 9 (11/2019), lists 10 and 11 (1/2020), and lists 12–14 (3/2020) during which the TL input from the game has been up-regulated enough to reach the 100-word limit in comparatively less time to form a single vocabulary list. Due to the game being played at regular intervals throughout the study period, the differences observed in the rate of vocabulary acquisition may be explained by periods of linguistic up- and down-regulation built into the gameplay mechanics.

2.3 In-Game/Out-of-Game Vocabulary Reactivation

There was a total of 330 instances of vocabulary reactivation throughout the study period. These instances of vocabulary reactivation are presented in Table 1. After adjusting for duplicate vocabulary, there were a total of 100 instances of in-game reactivation accounting for 30.3% of total reactivation. There were a total of 230 instances of out-of-game reactivation accounting for 69.7% of total reactivation. Of the 230 instances of out-of-game reactivation, there were a total of five subcategories accounting for the various contexts during which this reactivation occurred. These out-of-game reactivation instances occurred when the author was engaged in freelance translation work (23 instances, 6.96%), conducting professional duties teaching EFL courses at a Japanese university (30 instances, 9.09%), actively engaged in the production of the TL in writing or speaking (62 instances, 18.79%), engaged in the consumption of TL media such as television programs (86 instances, 26.06%), and engaged in conversation with native TL speakers during which the native TL speakers used a catalogued vocabulary word (29 instances, 8.79%). These subcategories of out-of-game reactivation are presented with the in-game reactivation category in Fig. 3.

The percentages of reactivated vocabulary relative to vocabulary type is similar to the percentages in each vocabulary type (See Fig. 2) and are presented in Fig. 4. The similar ratios of reactivated vocabulary types to the total sample of catalogued vocabulary types suggests that no one vocabulary type is reactivated more than another with 61% of the reactivated vocabulary coming from Review Vocabulary A (making up 62.7% of the total sample), 9% of the reactivated vocabulary coming from Review Vocabulary B (making up 11.59% of the total sample), and 30% of the reactivated vocabulary coming from Novel Vocabulary (making up 25.71% of the total sample).

Table 1 Vocabulary reactivation scores and categories

Categories	Total	Reactivation (%)	Subcategories	Subtotal	Reactivation (%)
In-game	100	30.3	In-game	100	30.30
Out-of-game	230	69.7	Translation	23	6.97
			In-class/professional	30	9.09
			Personal use	62	18.79
			Media	86	26.06
			Native contact	29	8.79
Total	330	100	6	330	100

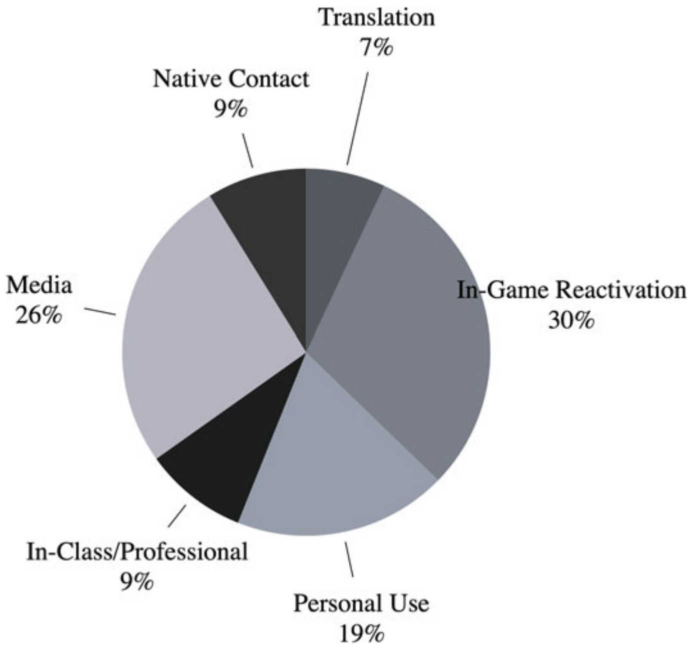

Fig. 3 Reactivation categories

The percentage of reactivated vocabulary acquired through gameplay relative to the non-reactivated vocabulary is presented in Fig. 5 and demonstrated that the majority of vocabulary (75%) made up the non-reactivated category. Figure 6 illustrates the reactivation rate throughout the study period across all 14 vocabulary lists as well as the total number of duplicate words throughout the study period. A decaying rate of reactivation was identified during the study period with an R^2 value of 0.83 indicating a high correlation for the polynomial regression analysis expressing a decay in reactivation relative to the elapsed time from the cataloguing of the vocabulary. Figure 7 illustrates the reactivation rate of Review Vocabulary A relative to the total sample reactivation score and a decaying trend of reactivation throughout the study period with an R^2 value of 0.61 indicating a strong correlation for this decaying reactivation trend. Figure 8 illustrates the reactivation rate of Review Vocabulary B relative to the total sample reactivation score and a trend expressing an increase in the Review Vocabulary B sample size throughout the study period with an R^2 value of 0.75 indicating a strong correlation for this model. The decaying trend of vocabulary reactivation observed in the aggregate and Review Vocabulary A samples was not observed in the Review Vocabulary B samples. Finally, Fig. 9 illustrates the reactivation rate of Novel Vocabulary relative to the total sample reactivation score and a decaying trend of reactivation throughout the study period with an R^2 value of 0.72 indicating a strong correlation for this trend relative to the gameplay interval.

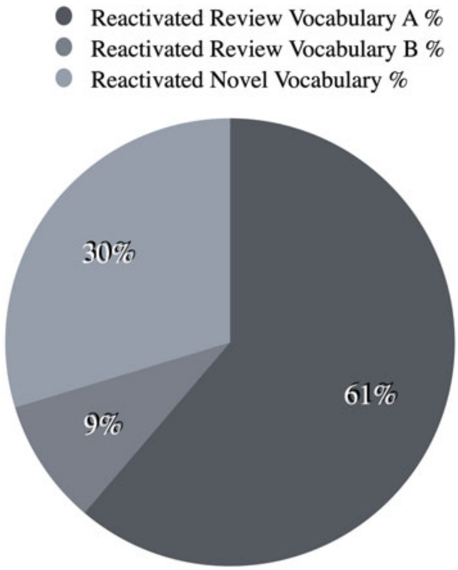

Fig. 4 Reactivation ratio by vocabulary type

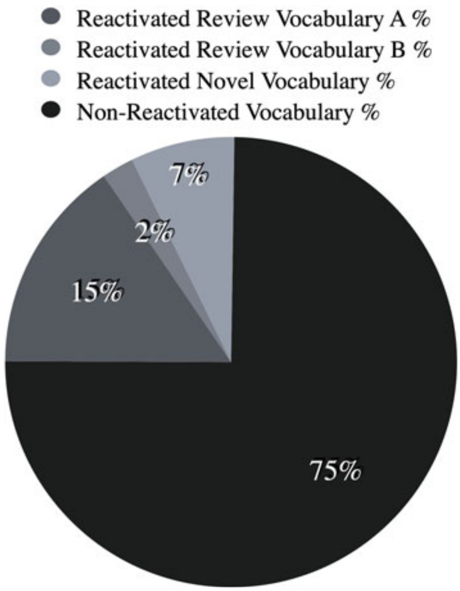

Fig. 5 Reactivated vocabulary by vocabulary type relative to non-reactivated vocabulary

2.4 Delayed Vocabulary Test Results

The averages from the delayed vocabulary tests for each of the vocabulary types are presented in Fig. 10. The vocabulary test score averages reveal that reactivated

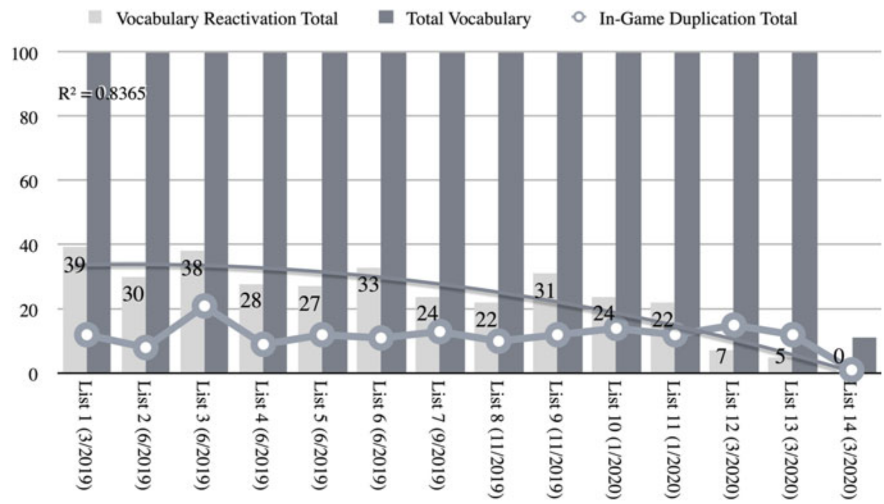

Fig. 6 Vocabulary reactivation across the total gameplay interval

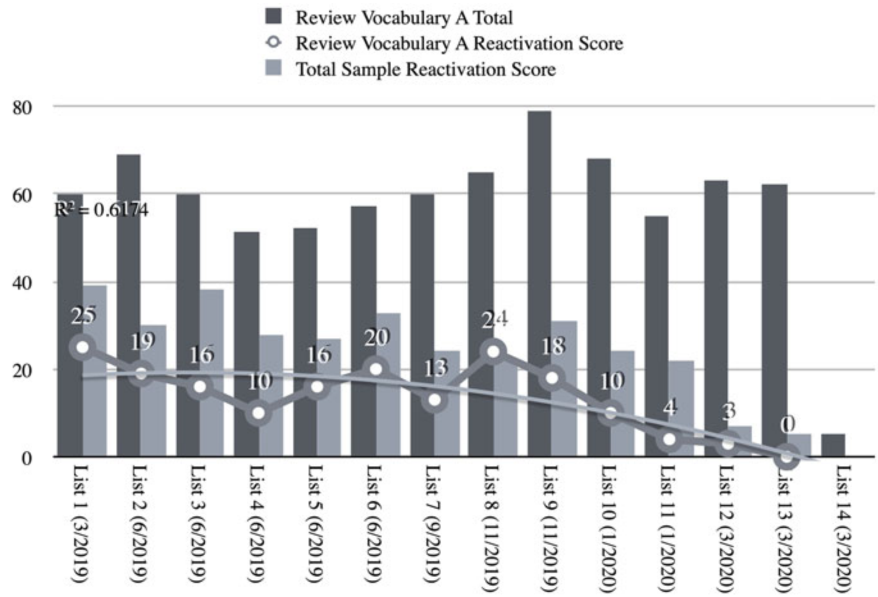

Fig. 7 Reactivation of Review Vocabulary A across the gameplay interval

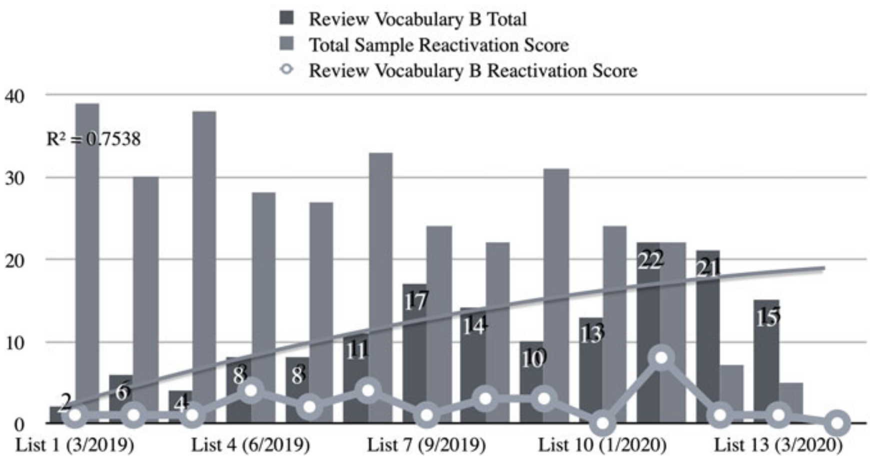

Fig. 8 Reactivation of Review Vocabulary B across the gameplay interval

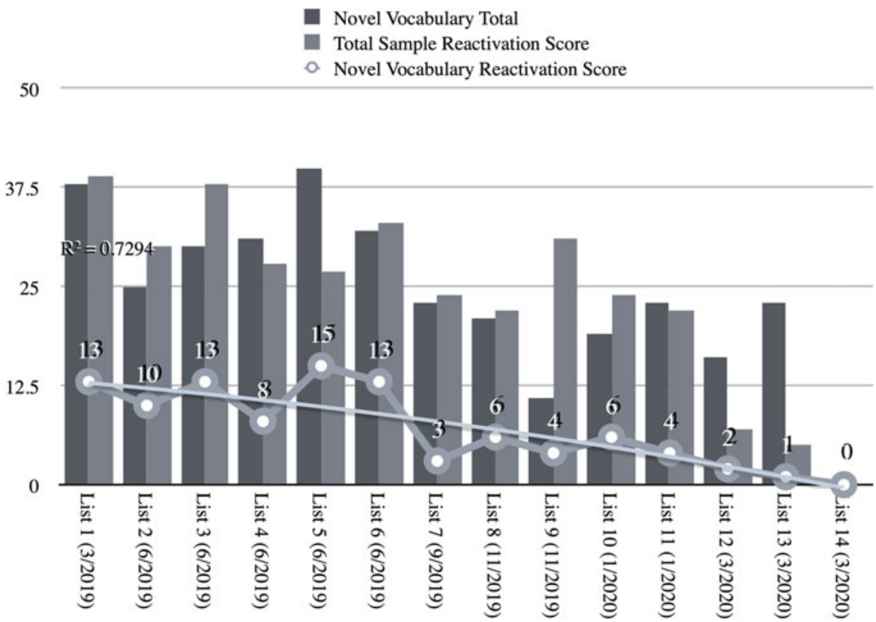

Fig. 9 Reactivation of Novel Vocabulary across the gameplay interval

vocabulary yielded higher levels of retention irrespective of vocabulary type. Two-sample, two-tail *t*-tests comparing reactivated vocabulary scores to non-reactivated vocabulary scores across all 14 delayed vocabulary tests resulted in *p*-values of <0.001 for the total sample aggregates and the Review Vocabulary A sample as well as a *p*-value of less than <0.05 for the Novel Vocabulary sample. With a threshold of <0.05 to reject the null hypothesis, these results indicate that higher average scores for reactivated vocabulary are statistically significant for these categories. Effect sizes were calculated with Cohen's *d* and yielded a *d*-value of 2.98 (large effect size) for the aggregate sample, a *d*-value of 2.78 (large effect size) for the Review Vocabulary A sample, and a *d*-value of 0.79 (moderately large effect size) for the Novel Vocabulary sample. However, the *t*-test conducted on the Review Vocabulary B sample resulted in a *p*-value of 0.41 indicating that the higher average score for reactivated vocabulary is not statistically significant for this category. This non-significant result is more likely to have been caused by the small sample size of the Review Vocabulary B category rather than this specific vocabulary type being immune to the reactivation benefit observed in the remainder of the samples. The *d*-value for the Review Vocabulary B sample was 0.32 indicating a small effect. Finally, the average scores for duplicate vocabulary were higher than the reactivated sample, but this was to be expected as the duplicate vocabulary were tested within a short time frame and thus benefited from the corrective feedback of the scoring process.

Figure 11 illustrates the delayed vocabulary test results across all 14 test dates for all 14 vocabulary lists. The corrective feedback from the testing procedure can be observed in the increase in test scores for the duplicate vocabulary type. Duplicate vocabulary were therefore removed from the other categories so as to not affect the analysis procedure. There is an observable decay in test scores from vocabulary list 1 to vocabulary list 14 in both the reactivated vocabulary (R^2 value of 0.62 indicating a high correlation) as well as in the non-reactivated vocabulary (R^2 value of 0.20 indicating a weak correlation). These results suggest that reactivated vocabulary had better long-term retention the longer the period of engagement with the game. This trend was not as pronounced for the non-reactivated vocabulary. This decay trend of reactivated vocabulary demonstrating better long-term retention for longer delays in

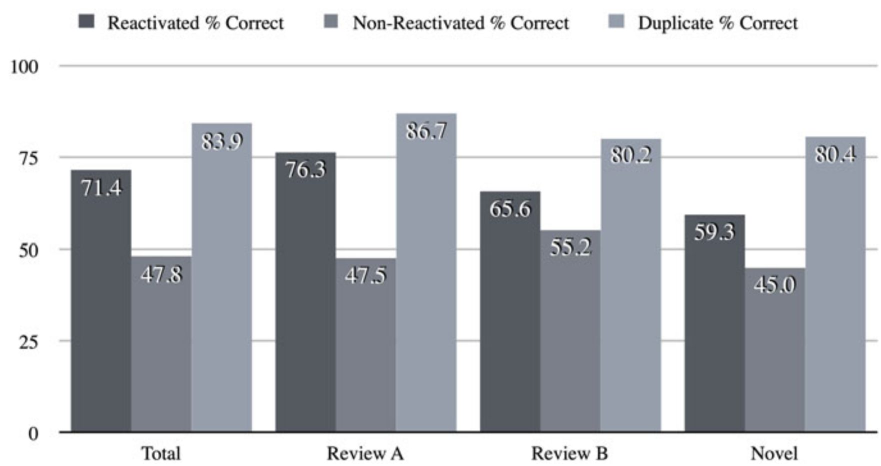

Fig. 10 Delayed vocabulary test results by category

testing is also observable in the Review Vocabulary A sample in Fig. 12. The R^2 value of 0.60 observed in the decay rate of the reactivated vocabulary test scores indicates a strong correlation relative to the gameplay interval. In contrast to the aggregate sample, however, when analyzing the decay rate of vocabulary test scores in the non-reactivated vocabulary from the Review Vocabulary A sample shown in Fig. 12, an R^2 value of 0.53 indicates a moderate correlation to a model indicating better long-term retention for a longer period of gameplay. The test results from the Review Vocabulary B sample illustrated in Fig. 13 however do not display the same trend. The non-reactivated Review Vocabulary B sample displays the same decay trend (with an R^2 value of 0.55 indicating a moderate correlation) observed in the aggregate and Review Vocabulary A samples, however the reactivated Review Vocabulary B sample displays the opposite trend (i.e., vocabulary test scores increasing as elapsed time decreases between the study's end and the delayed tests), albeit with a weak correlation relative to the gameplay interval with an R^2 value of 0.36. Finally, the Novel Vocabulary sample (Fig. 14) further complicates the picture by again displaying a decay rate for the reactivated Novel Vocabulary sample suggesting better long-term retention with longer periods of gameplay, however the reverse trend is observed with the non-reactivated Novel Vocabulary sample. Both trends, however, have weakly correlated R^2 values with 0.36 for the reactivated Novel Vocabulary sample and 0.38 for the non-reactivated Novel Vocabulary sample. To arrive at a clearer understanding about the relationship between the delayed vocabulary test results and period of active gameplay, the statistical significance of correlation coefficients between these two factors were measured.

Figure 15 illustrates the entire range of delayed vocabulary tests from 24 to 12 months post study conclusion plotted with the results of all 14 delayed vocabulary tests for the aggregate sample. The correlation between a longer delay and better long-term retention is again more pronounced in the reactivated vocabulary sample. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the reactivated aggregate vocabulary sample yielded an r-value of 0.75 demonstrating a strong linear correlation between longer delays and better long-term retention, and a p-value of <0.01 from a two-tailed t-distribution sample indicates that this correlation is significant. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the non-reactivated aggregate vocabulary sample yielded an r-value of 0.36 demonstrating a weak linear correlation between longer delays and better long-term retention, and a p-value of 0.19 from a two-tailed t-distribution sample indicates that this correlation is non-significant.

Figure 16 illustrates the vocabulary test results of all 14 delayed vocabulary tests for the Review Vocabulary A sample plotted with the range of post-test delays. The correlation between a longer delay and better long-term retention is again more pronounced in the reactivated vocabulary sample, but the correlation between longer delays and better long-term retention is much more pronounced in the non-reactivated Review Vocabulary A sample than in the non-reactivated aggregate vocabulary sample. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the reactivated Review Vocabulary A sample yielded an r-value of 0.79 demonstrating a strong linear correlation between longer

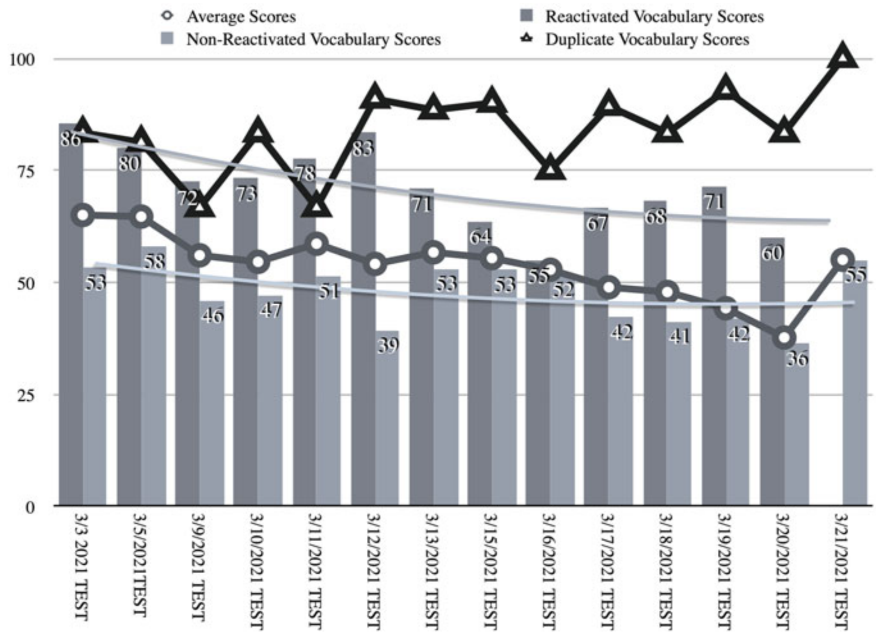

Fig. 11 Delayed vocabulary test results of aggregate sample across the gameplay interval

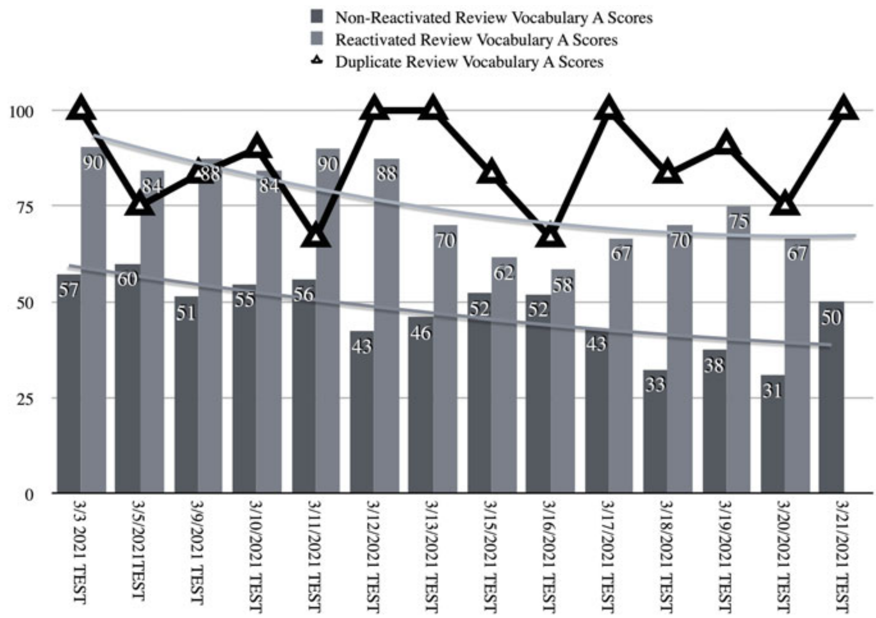

Fig. 12 Delayed vocabulary test results of Review Vocabulary A across the gameplay interval

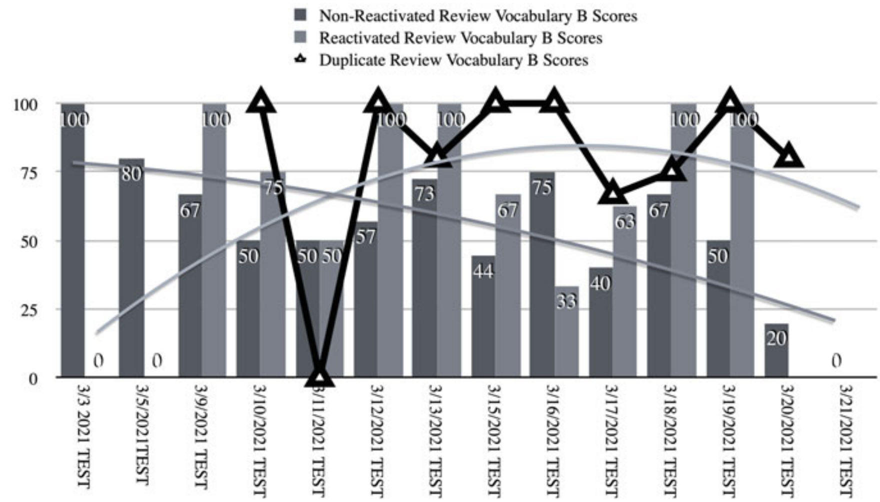

Fig. 13 Delayed vocabulary test results of Review Vocabulary B across the gameplay interval

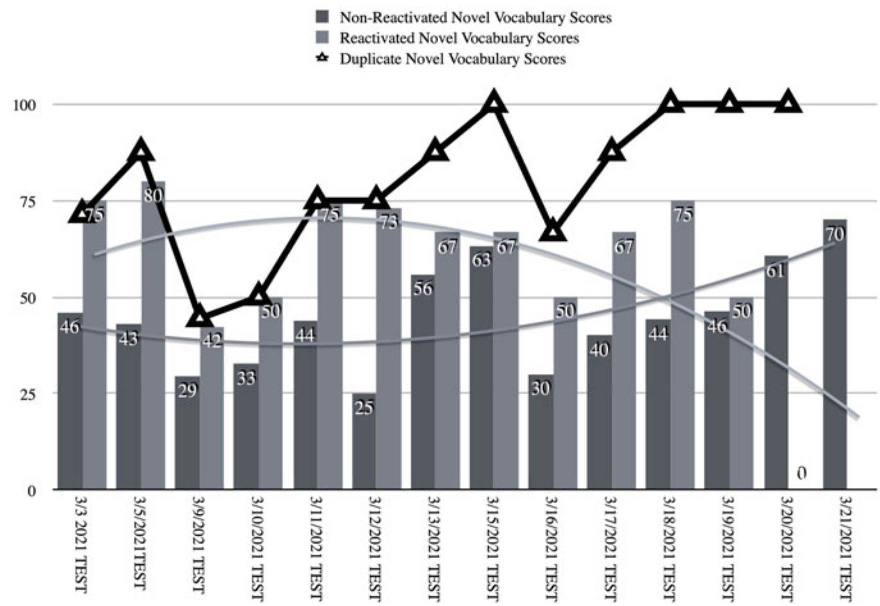

Fig. 14 Delayed vocabulary test results of Novel Vocabulary across the gameplay interval

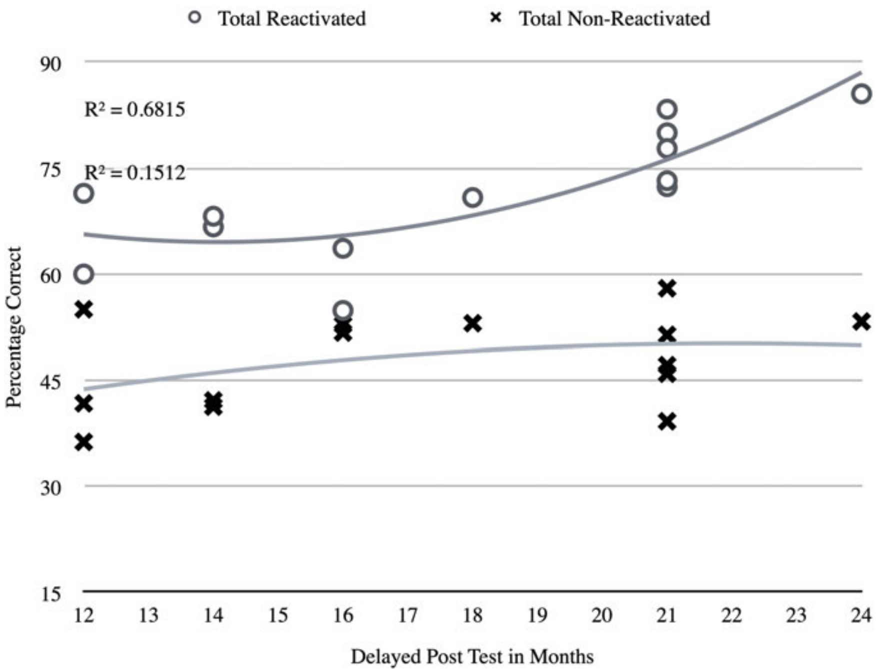

Fig. 15 Correlation between delayed vocabulary test results and delay interval for aggregate sample

delays and better long-term retention, and a *p*-value of <0.001 from a two-tailed *t*-distribution sample indicates that this correlation is significant. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the non-reactivated Review Vocabulary A sample yielded an *r*-value of 0.69 demonstrating a strong linear correlation between longer delays and better long-term retention, and a *p*-value of <0.01 from a two-tailed *t*-distribution sample indicates that this correlation is significant.

Figure 17 illustrates the vocabulary test results of all 14 delayed vocabulary tests for the Review Vocabulary B sample plotted with the range of post-test delays. The correlation between a longer delay and better long-term retention is reversed in the reactivated Review Vocabulary B sample demonstrating worse long-term retention with longer delay intervals in testing, but the correlation between longer delays and better long-term retention is again evident in the non-reactivated Review Vocabulary B sample demonstrating better long-term retention with longer delay intervals in testing. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the reactivated Review Vocabulary B sample yielded an *r*-value of −0.44 demonstrating a weak negative linear correlation between longer delays and better long-term retention, but a *p*-value of 0.11 from a two-tailed *t*-distribution sample indicates that this correlation is non-significant. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in

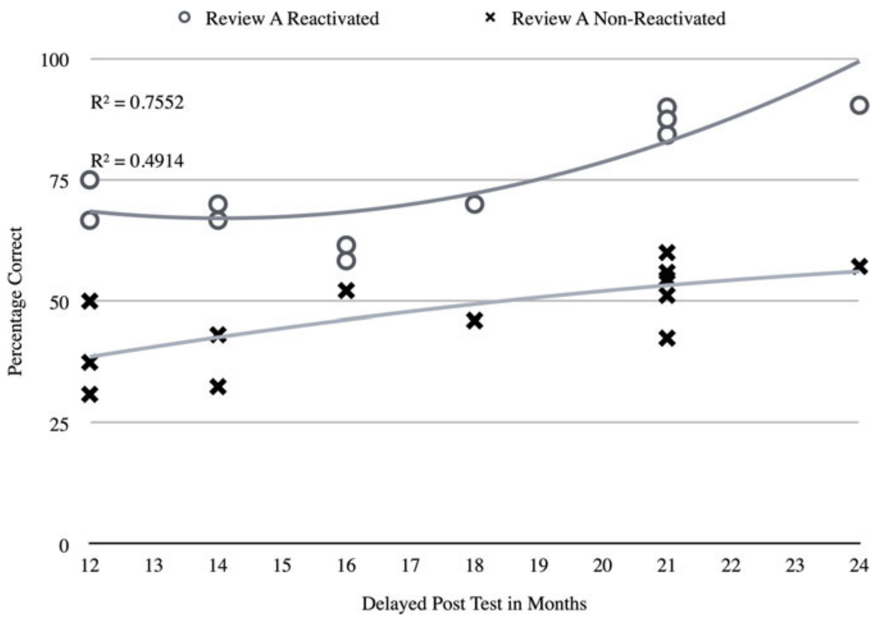

Fig. 16 Correlation between delayed vocabulary test results and delay interval for Review Vocabulary A

the non-reactivated Review Vocabulary B sample yielded an r-value of 0.67 demonstrating a strong linear correlation between longer delays and better long-term retention, and a p-value of <0.01 from a two-tailed t-distribution sample indicates that this correlation is significant.

Figure 18 illustrates the vocabulary test results of all 14 delayed vocabulary tests for the Novel Vocabulary sample plotted with the range of post-test delays. The correlation between a longer delay and better long-term retention is again evident in the reactivated Novel Vocabulary sample demonstrating better long-term retention with longer delay intervals in testing, but the correlation between longer delays and better long-term retention is again reversed in the non-reactivated Novel Vocabulary demonstrating worse long-term retention with longer delay intervals in testing. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the reactivated Novel Vocabulary sample yielded an r-value of 0.47 demonstrating a moderate linear correlation between longer delays and better long-term retention, but a p-value of 0.08 from a two-tailed t-distribution sample, although approaching significance at the threshold of <0.05 to reject the null hypothesis, nevertheless indicates that this correlation is non-significant. The correlation coefficient between the post-test delay intervals and vocabulary test score factors in the non-reactivated Novel Vocabulary sample yielded an r-value of -0.52 demonstrating a moderate negative linear correlation between longer delays and better long-term retention, and a p-value of 0.05 from a two-tailed t-distribution sample indicates that this correlation may be borderline statistically significant.

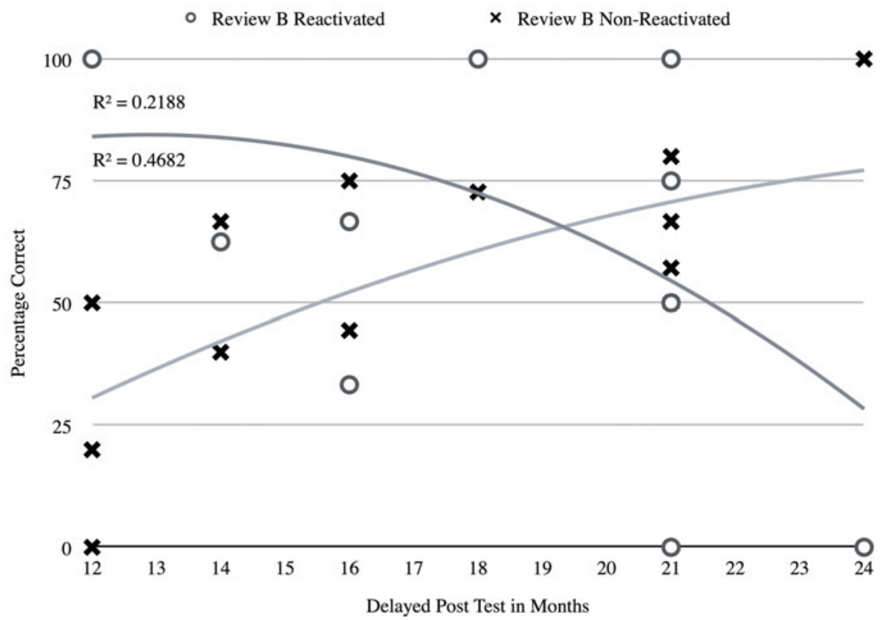

Fig. 17 Correlation between delayed vocabulary test results and delay interval for Review Vocabulary B

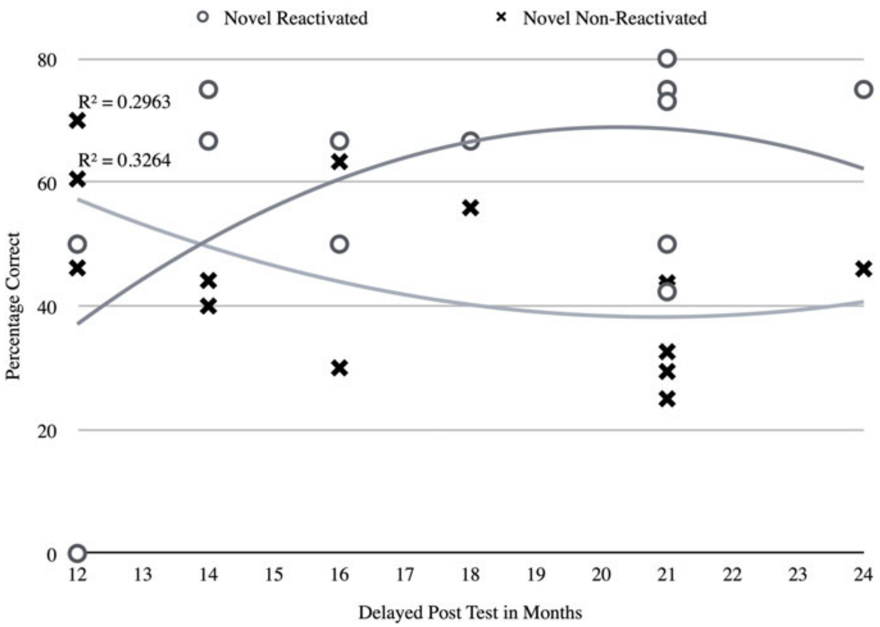

Fig. 18 Correlation between delayed vocabulary test results and delay interval for Novel Vocabulary

3 Discussion

3.1 Limitations

Before discussing how the reactivation of TL vocabulary acquired through gameplay may have helped elicit long-term retention of vocabulary acquired from playing FFIX, it is first necessary to point out that much of the vocabulary reactivation that occurred during this study was the direct result of the author's TL skill level and living situation within the TL culture. Therefore, the results presented in this study are highly subjective. Considering this limitation, this study makes no claims on the generalizability of the findings discussed in the following sections.

3.2 Observations on Vocabulary Acquisition and Reactivation

The following discussion provides insight into which factors influenced vocabulary acquisition and retention within the scope of the data analyses presented above. Furthermore, the established research questions will frame the discussion of the results: (1) What type of vocabulary will be learned from playing FFIX? (2) How will vocabulary acquired from playing FFIX be recalled? and (3) What factors will influence the retention of vocabulary acquired from playing FFIX? Regarding what type of vocabulary was learned from playing FFIX, the overwhelming majority (74.29% as indicated in Fig. 2) of the vocabulary were review vocabulary. By actively studying the TL while engaging with FFIX, more opportunities for in-game reactivation of those words were arguably created, which, as was suggested in the results of the delayed tests (see Fig. 10), may have influenced long-term retention of review-type vocabulary focused on meaning (i.e., Review Vocabulary A). The retention of these review-type vocabulary may have also strengthened and consolidated over time as the author continued to interact with the gameplay mechanics (see Figs. 11 and 12). These results are in line with vocabulary acquisition research which has shown that the frequency of encountering vocabulary is correlated with better vocabulary recall (Cobb & Horst, 2011; Zahar et al., 2001). If the goal of studying TL vocabulary is to be able to recall them after the cessation of study, then the interaction with FFIX over a long gameplay interval while linguistically engaging with study of the TL may have helped the author achieve that goal more effectively.

Regarding how vocabulary was recalled, one factor that may have particularly influenced the long-term retention of acquired vocabulary were the instances of out-of-game vocabulary reactivation. The total percentage of in-game reactivation was significantly smaller at 30.3% (after factoring out duplicate vocabulary entries) than out-of-game reactivation at 69.7%. While some of the out-of-game reactivation subcategories presented in this study were heavily influenced by the author living and working in the TL culture (e.g., native contact at 8.79%, personal use at 18.79%, and in-class/professional at 9.09%), by far the largest out-of-game reactivation subcategory was TL media consumption which accounted for 26.06% of the total reactivation sample. Indeed, the out-of-game TL media reactivation category was second only to the in-game reactivation category at 30.3%. While it may be argued that access to TL media is aided by residing in the TL culture, the technological shift towards the digital streaming of media content has arguably increased internet access to a wider range of culture-specific media content. It is therefore worth noting that the largest out-of-game reactivation category, as observed in this study, was not necessarily dependent on residing in the TL culture.

Finally, regarding the factors which influenced vocabulary retention, the statistically significant correlations found in the quantitative analyses examining how the gameplay interval relates to long-term vocabulary retention align well with the hypothesized psycholinguistic benefits that time-intensive digital game titles such as single-player RPGs may have on long-term TL vocabulary retention. As is evidenced in Fig. 6 (aggregate sample), Fig. 7 (Review Vocabulary A sample), and Fig. 9 (Novel Vocabulary sample) which plot the reactivation data across the length of the study period, there is a common trend for reactivation, as defined by and scored in this study, to plateau at approximately one third of the total sample size at approximately three to six months while engaging with the game. Although it is beyond the scope of this study to make a judgment about the quality and quantity of reactivation after the cessation of gameplay, the results from the linear regression analyses in the aggregate sample (Figs. 11 and 16) have shown that the strong correlation between longer intervals of gameplay and better long-term retention of vocabulary acquired during gameplay was statistically significant for reactivated vocabulary while the same weak correlation in non-reactivated vocabulary sample remained statistically non-significant. Should there have been significant amounts of vocabulary reactivation after the cessation of gameplay, the argument could be made that the data from the non-reactivated sample should have started to approximate the data from the reactivated sample after three to six months. Since the post-test interval from the final vocabulary list was 12 months, this should have been enough time for the reactivation effect to emerge in the data. The spacing of the gameplay interval past the six-month mark may therefore have helped consolidate the retention rate of reactivated vocabulary as well as provide more time for the vocabulary acquired from gameplay to be reactivated.

3.3 Controlling for Total Gameplay Interval and Acquisition Type

Within the structure of this study, the benefits of longer total gameplay intervals did not stop at the three-to-six-month gameplay interval needed to maximize reactivation rates of vocabulary acquired from gameplay. As is evidenced in the quantitative analyses, the rate of long-term retention started to increase at six months into the gameplay interval and continued to the 12-month mark, where the highest scores were observed in the aggregate sample (Figs. 11 and 16) and Review Vocabulary A sample (Figs. 12 and 17) for reactivated vocabulary. This effect of time-sensitive benefits to vocabulary consolidation was also observed in a study by Cobb and Horst (2011) who used the vocabulary training game *My Word Coach* (Ubisoft, 2007) on the Nintendo DS. In this study, it was found that delayed vocabulary post-test scores were higher for review-type vocabulary two months after gameplay than immediate vocabulary post-test scores. Cobb and Horst (2011) subsequently argued that sustained gameplay past this two-month post-game interval may have increased the observed consolidation effect. In this study, sustained interaction with FFIX past the six-month mark may have contributed to better consolidation and retention of the reactivated vocabulary, particularly Review Vocabulary A words which were acquired early on in the gameplay interval. The motivation to sustain interaction with a TL digital game for longer than six months is obviously not generalizable and is dependent on a number of factors in both the game and the player. However, there are some observations that were made in this study that may shed light as to how a single-player RPG such as FFIX was conducive to accomplishing the task of sustained, long-term gameplay for the author.

Some digital games require the comprehension of linguistic input more than others (e.g., following the instructions of in-game characters to accomplish specific tasks), whereas other digital games are much more focused on the execution of gameplay mechanics that do not require the comprehension of linguistic input (e.g., navigating an in-game object through a specific location). The argument could be made that TL digital games which require the comprehension of linguistic input would be more beneficial to a language learner using said games for SLA. However, the argument which this paper presents is that TL digital games which incorporate both of these gameplay mechanics may help increase the gameplay interval, which in turn may have benefits to long-term vocabulary retention. Many story-oriented RPGs, whether single-player or otherwise, feature such gameplay mechanics which balance periods of linguistic up-regulation and down-regulation throughout the game. Throughout the study period for this paper, FFIX was played at regular intervals, however the time required to reach the 100-term vocabulary list limit varied greatly depending on how much language was being used in the game (see Fig. 1). Although FFIX is a heavily story-driven game, a considerable amount of the game requires absolutely no language whatsoever. While it may seem counterintuitive to argue for the benefit to TL vocabulary acquisition through the incorporation of linguistically down-regulated intervals within a TL digital game being used for SLA, these intervals of linguistic down-regulation may have helped extend the total gameplay time to 94 h, the total gameplay interval to one year, and arguably improved the long-term retention of review-type vocabulary acquired from the linguistically up-regulated segments of the game.

3.4 Future Considerations

The trends observed in this study of improved long-term retention of reactivated vocabulary relative to the total gameplay interval, while far from generalizable, open up a number of avenues for future research endeavors to further elaborate upon these findings. The first of these considerations for future research involve dictionary use (or lack thereof) and how the cataloguing process of TL vocabulary (or lack thereof) may affect the reactivation process. In this study, every time an unknown word was encountered during gameplay, the meaning/phonetic Kanji reading of the item was immediately looked up in a dictionary and catalogued for future analysis. In so doing, these dictionary-reinforced and catalogued vocabulary may have been primed for reactivation more so than if they had not been looked up in a dictionary and subsequently catalogued. A study by Hulstijn et al. (1996) identified a reinforcing effect on TL vocabulary acquisition through the use of marginal glosses and dictionaries relative to a control group without TL vocabulary reinforcement when Dutch advanced students of French read French short stories. Despite the aforementioned study being focused on dictionary use while reading traditional literature, a similar effect in story-oriented, single-player RPGs is arguably at play in the findings presented in this study. Future research therefore should focus on addressing whether this potential priming effect by dictionary use and cataloguing is more effective at eliciting more instances of reactivation than a control sample that does not use dictionaries or cataloging before recording instances of reactivation of TL vocabulary acquired through TL digital games.

Another consideration that should be made for future research concerns the upper limit of the time-sensitive consolidation effect on the long-term retention of reactivated vocabulary. Whereas the vocabulary reactivation curve observed in this research (see Fig. 6) plateaued at around three to six months, there was an abrupt increase in the consolidation effect on the long-term retention of reactivated vocabulary (see Fig. 15) starting at six months from the end of the gameplay interval (or at 18 months post-game cessation). These gains in long-term retention continued to increase until 12 months from the end of the gameplay interval (or until 24 months post-game cessation). It remains unclear, however, where the upper limit of this time-sensitive consolidation effect is, and future research should therefore endeavor to identify whether this time-sensitive consolidation effect on the long-term retention of reactivated vocabulary continues to increase past 12 months of sustained gameplay and whether there is a decay rate at some point past this 12-month mark. To investigate this consideration for future research, more precise measurements of time need to be considered by recording the date of vocabulary acquisition from each gaming session instead of the method used in this research which assigned the date of vocabulary acquisition to the completed vocabulary lists of 100 words each. Finally, a larger sample size of participants playing the same game for over a year needs to be employed so as to make the findings more generalizable to a range of language learners at different levels of TL development.

3.5 Conclusion

Incorporating digital games in formal educational environments which have to meet standardized pedagogical goals has not only been identified as being uniquely challenging (Godwin-Jones, 2014), it may also be largely impractical considering the large amounts of time necessary to achieve desirable pedagogical outcomes in which deep learning can take place (Gee, 2003; Squire, 2005). The impracticality for many digital game titles to serve as standardized pedagogical tools should not however discredit the remarkable pedagogical potential that digital games represent, especially outside of formal educational environments. In particular, the ludic element of digital games that is a necessary component of commercially successful games can inspire players to engage with the medium over long stretches of time, during which these instances of deep learning can take place. As Gee (2003) states, "For humans, real learning is always associated with pleasure, is ultimately a form of play - a principle almost always dismissed by schools." For the author, the most effective gameplay interval for TL vocabulary acquisition was between six and twelve months. Playing FFIX for this study for a total of 94 h over the course of a year was also not arduous in the least. Indeed, it was a pleasure. Continual engagement with TL media for SLA purposes over substantial periods of time, be it a textbook, a novel, a comic book, a television series, or a digital game, requires motivation, and certainly one factor that influenced the author's motivation to continue interacting with the game was how much pleasure was derived therefrom. Considering that TL media used for SLA outside of formal educational environments arguably exert much less externalized motivation on the language learner (e.g., getting a good grade in a class), the entertainment value of the TL medium of SLA instruction becomes all the more important.

Ultimately the experience of playing FFIX for the purpose of SLA has been a subjective experience dependent on the author's motivation (heavily nostalgic for this particular game title), schedule to play the game (very little time to devote more than 30 min a day to gameplay), linguistic level (highly proficient in Japanese), living situation (living in Japan at the time of this study), work (actively using Japanese in a professional environment), and personal connections to the TL culture (married to a Japanese national). It also goes without saying that a sample size of one does not lead to generalizable results. Bearing these limitations in mind, the experience of playing FFIX for SLA purposes has led to fruitful observations which may inform future research regarding vocabulary acquisition and retention from single-player RPGs. From the author's perspective, a number of gameplaying habits emerged which were identified as being potentially beneficial for SLA. First, cataloging words/using a dictionary to look words up as they emerge during gameplay may have aided in the reactivation of vocabulary, which in turn may have led to better long-term retention. However, there is no data from this study that can say one way or another that not using a dictionary or cataloging vocabulary acquired through gameplay will not lead to similar results of reactivation and long-term retention. Second, choosing a story-oriented digital game title (i.e., a linguistically up-regulated game) which also employs sufficient amounts of gameplay time devoted to linguistically down-regulated activities may have ultimately prolonged the total gameplay interval and benefited long-term retention of vocabulary acquired through gameplay. It may seem counterintuitive to “waste time” engaging with gameplay mechanics that do not make use of the TL when the player is using the game specifically for SLA purposes, however the data presented in this research as well as throughout the field of memory research points to the benefits of prolonging the learning interval (Appleton-Knapp et al., 2005; Bjork, 2014; Janiszewski et al., 2003). Third, continuing to play the game after the three-to-six-month mark may have maximized reactivation rates with the consolidation of reactivated vocabulary increasing after the six-month mark. Furthermore, actively engaging with TL media and culture as much as possible and studying the TL during the gameplay interval may have also increased levels of review-type vocabulary acquisition (which had better long-term retention rates relative to novel vocabulary) and out-of-game reactivation. Finally, choosing a game that is personally enjoyable and level-appropriate with respect to the player’s linguistic ability in the target language went a long way in encouraging sustained engagement with the game. The process of language acquisition through digital gaming is not automatic and requires careful consideration of the most appropriate gaming style, most beneficial game title, and most productive linguistic habits when not engaging with the game, but the potential benefits to language development are vast.

References

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research, 32*(2), 266–276.

Bjork, R. A. (2014). Forgetting as a friend of learning. In *Remembering* (pp. 39–52). Psychology Press.

Breiner-Sanders, K. E., Lowe, P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines—Speaking (Revised 1999). *Foreign Language Annals, 33*(1), 13.

Bryant, T. (2006). *Using World of Warcraft and other MMORPGs to foster a targeted, social, and cooperative approach toward language learning*. Dickinson College.

Cobb, T., & Horst, M. (2011). Does word coach coach words. *CALICO Journal, 28*(3), 639–661.

Cornillie, F., Thorne, S. L., & Desmet, P. (2012). ReCALL special issue: Digital games for language learning: Challenges and opportunities. *ReCALL: The Journal of EUROCALL, 24*(3), 243.

Final Fantasy IX. (2000). Square Enix.

Gass, S. (2010). Interactionist perspectives on second language acquisition. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 217–231). Oxford University Press.

Gee, J. P. (2003). *Learning about learning from a video game: Rise of nations* (Unpublished manuscript). University of Wisconsin-Madison. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.543.3194&rep=rep1&type=pdf

Godwin-Jones, R. (2014). Games in language learning: Opportunities and challenges. *Language Learning & Technology, 18*(2), 9–19.

HowLongToBeat. (2021). *How long is final fantasy IX?* Retrieved November 28, 2021, https://howlongtobeat.com/game?id=3505

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.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*(1), 138–149.

Kinginger, C. (2010). American students abroad: Negotiation of difference? *Language Teaching, 43*(2), 216–227.

Lai, C., Ni, R., & Zhao, Y. (2012). Digital games and language learning. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary computer-assisted language learning.* A&C Black.

Lantolf, J. P. (1994). Sociocultural theory and second language learning: Introduction to the special issue. *The Modern Language Journal, 78*(4), 418–420.

Mackey, A., Abbuhl, R., & Gass, S. M. (2013). Interactionist approach. In *The Routledge handbook of second language acquisition* (pp. 25–41). Routledge.

My Word Coach. (2007). Ubisoft.

Nietfeld, J. L. (2017). The role of self-regulated learning in digital games. In *Handbook of self-regulation of learning and performance* (pp. 271–284). Routledge.

Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly, 31*(3), 409–429.

Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning, 23*(5), 429–439.

Peterson, M. (2012). Language learner interaction in a massively multiplayer online role-playing game. In *Digital games in language learning and teaching* (pp. 70–92). Palgrave Macmillan.

Peterson, M., White, J., Mirzaei, M. S., & Wang, Q. (2022). A review of research on the application of digital games in foreign language education. In *Research anthology on developments in gamification and game-based learning* (pp. 1948–1971). IGI Global.

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.

Rankin, Y. A., Morrison, D., McNeal, M., Gooch, B., & Shute, M. W. (2009, April). Time will tell: In-game social interactions that facilitate second language acquisition. In *Proceedings of the 4th international conference on foundations of digital games* (pp. 161–168). Association for Computing Machinery.

Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review, 68*(2), 138–163.

Sørensen, B. H., & Meyer, B. (2007). Serious games in language learning and teaching-a theoretical perspective. In *Proceedings of the 2007 DiGRA conference* (pp. 559–566). Digital Games Research Association.

Squire, K. (2005). Changing the game: What happens when video games enter the classroom? *Innovate: Journal of Online Education, 1*(6), Article 5.

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of platform. *Language Learning and Technology, 14*(2), 95–110.

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

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541–572.

Kevin Reay Wrobetz, Ph.D., is a faculty member of the Department of Business Administration at Kobe Gakuin University in Japan. His main area of research is foreign language vocabulary acquisition in the field of digital game-based learning.

Vocabulary Learning from Video Viewing

YouTube for Incidental Vocabulary Learning

Duygu Candarli

Abstract This empirical study examines the potential of YouTube worldwide trending videos for incidental vocabulary learning for the purposes of academic lectures and seminars at universities in English-speaking countries. YouTube is a popular social media platform that hosts videos of a wide range of genres, including vlogs (video blogs), songs, and news. Although previous studies have examined the potential of movies and TV programmes for incidental vocabulary learning, movies and TV programmes do not capture the wide range of genres that people are exposed to. To date, very little attention has been paid to the potential of YouTube for incidental vocabulary learning for academic listening. In this study, the lexical demands of a bespoke corpus of the transcriptions of YouTube worldwide trending videos were identified. Then, encounters with academic spoken word families with different frequency cut-offs were explored in the corpus of YouTube worldwide trending videos. The results reveal that YouTube trending videos have great potential for incidental vocabulary learning, creating a path from informal learning to academic language development. The present study has significant implications for incidental vocabulary learning through watching YouTube trending videos for adolescent and adult English language learners at the global level. It is recommended that YouTube trending videos are incorporated into EFL/ESL classes for adolescent and adult learners and that L2 learners watch these videos as part of their out-of-class learning activities to enhance their academic listening.

Keywords YouTube · Incidental vocabulary learning · Lexical coverage · Corpus linguistics

D. Candarli (✉)
University of Southampton, Southampton, UK
e-mail: d.candarli@soton.ac.uk

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_8

1 Introduction

YouTube, which is a social media platform that hosts a number of different genres, including video blogs (vlogs), songs, and adverts, provides authentic audio-visual input that can be enriched with captions for L2 learners of English although videos may not necessarily be targeted at L2 learners specifically. In addition to providing audio-visual input, YouTube also offers an informal learning environment, contributing to students' digital literacy skills and enhancing their ability to analyse and evaluate content (Tan, 2013). YouTube is globally the most popular platform for online content among children and young people as part of their entertainment and/or leisure activities (Folkvord et al., 2019); however, its potential for incidental vocabulary learning, which is "learning words without deliberate decision to commit information to memory" (Laufer & Hulstijn, 2001, p. 11), is largely an unchartered territory. In an experimental study, Arndt and Woore (2018) found that vocabulary gains from watching video blogs were approximately equal to those from reading blog posts, suggesting the effectiveness of video blogs for incidental vocabulary learning. However, the authors used pseudo-words and created video blogs for their study rather than selecting authentic input from YouTube.

YouTube's role as an informal learning environment is acknowledged in the literature (e.g., Soyoof et al., 2021); however, it remains unknown whether it has the potential for incidental vocabulary learning for lectures and seminars at universities. One of the most popular video genres on YouTube is vlogs in which creators share their life experiences, review products, and share their perspectives on these products as well as personal and social issues (Munnukka et al., 2019). On the other hand, the functions of academic lectures and seminars are interpreting and evaluating the content and interacting with members of the class (Deroey & Tavernies, 2011). Such functions show some similarities with those of vlogs and other YouTube videos, despite the differences in formality, genre, and topics between spoken academic genres and YouTube videos because both genres share perspectives with their audience. These functional similarities point to the need for empirical investigation of vocabulary in these genres to find whether watching YouTube videos can create a path from informal learning to the development of academic listening. Hence, this study aims to identify lexical coverage of the Youtube trending videos and explore encounters with the academic spoken word families (Dang et al., 2017) in the bespoke corpus of YouTube trending videos and argues for the role of watching YouTube videos as one of the ways to develop academic listening.

2 Literature Review

2.1 Previous Studies of Lexical Coverage of Spoken Discourse

Lexical coverage is concerned with vocabulary profiling of spoken and written texts in order to determine their difficulty of them or to identify vocabulary knowledge required to comprehend them for the purposes of teaching, assessment, and materials development (Nation & Webb, 2011). Empirical studies on lexical coverage of both written and spoken texts have increased since the 1980s, along with the development of corpus tools and methods (see Nurmukhamedov & Webb, 2019 for an extensive overview). In specialised spoken discourse, previous studies identified that L2 learners of English would need knowledge of the most frequent 3,000 word families plus proper nouns and marginal words (e.g. *ah*, *ouch*) to understand the listening sections of the internet-based Test of English as a Foreign Language (TOEFL iBT) (Kaneko, 2015), popular English podcasts (Nurmukhamedov & Sharakhimov, 2021), popular songs (Tegge, 2017), movies (Webb & Rodgers, 2009a) and television programs (Webb & Rodgers, 2009b) for good comprehension that corresponds to 95% lexical coverage. Similarly, van Zeeland and Schmitt (2013) reported that 95% lexical coverage would be necessary for good listening comprehension, whereas this figure was found to be 98% for optimal comprehension. The more lexically demanding spoken genres were found to be academic spoken English (Dang & Webb, 2014) and TED Talks (Nurmukhamedov, 2017) whose 95% lexical coverage would necessitate the knowledge of the most frequent 4,000 word families plus proper nouns and marginal words. Although the lexical coverage of a number of specialised spoken genres has been investigated as seen above, relatively new social media genres, including YouTube trending videos have yet to be examined. Given that young people spend more time on YouTube rather than watching television (Burgess & Green, 2018), and that YouTube "has become the most preferred platform for viewing online content" (Folkvord et al., 2019, p. 2) for young people, it is important to gain insights into the lexical coverage of popular YouTube videos in order to identify lexical demands of YouTube videos for teaching and learning vocabulary, which would, in turn, inform L2 learners' extensive viewing practices and teachers' use of YouTube in English as a second/foreign language classes.

2.2 The Potential of Extensive Viewing for Incidental Vocabulary Learning

Previous research has examined the potential of extensive viewing for incidental vocabulary learning, using corpus methods (e.g., Dang, 2020; Green, 2020; Rodgers & Webb, 2011; Webb, 2010). Empirical research shows that learners can incidentally learn vocabulary from movies and TV programmes (see Nurmukhamedov & Webb, 2019 for an overview). Webb (2010) looked at the potential of learning low-frequency vocabulary by watching movies and concluded that L2 learners would encounter 23% of the word families that were classified as low-frequency by watching 70 movies, which would provide opportunities for incidental vocabulary learning.

More recent studies revealed the potential of extensive viewing for vocabulary learning for student reading and writing at secondary schools through watching movies and TV programs (Green, 2020) and for specialised spoken academic vocabulary in the discipline of medicine through watching medical TV programs (Dang, 2020). These two studies are important in that they give evidence for the academic language development of L2 learners through watching movies and TV programmes, potentially creating a path from informal learning to academic language development (Dang, 2020). A common finding in these empirical studies is that the more episodes or TV programmes that L2 learners would watch, the more likely that L2 learners encounter more word families frequently in a wide range of contexts (dispersion).

The concepts of frequency and dispersion are pertinent to the premises of usage-based accounts of language learning, which posit that the more frequently vocabulary is encountered in a wide range of contexts, the more likely that L2 learners acquire and use vocabulary (e.g., Ellis & Wulff, 2015). Informed by usage-based accounts of language learning (e.g., Ellis & Wulff, 2015) and building on the previous empirical research (Dang, 2020; Green, 2020), this study aims to explore to what extent, if any, the same word families that occur in academic lectures and seminars can be encountered by watching YouTube trending videos. To the knowledge of the author, no study has examined Youtube videos' potential for incidental vocabulary learning for academic lectures and seminars; hence, the current study fills this gap, and the findings will have important implications for designing in-class activities for teachers and planning out-of-class extensive viewing activities for L2 learners of English.

3 The Current Study

This exploratory study aims to address the following research questions:

1. What are the lexical demands of YouTube worldwide trending videos, including video blogs (vlogs), in terms of Nation's (2018) British National Corpus (BNC)/Corpus of Contemporary American English (COCA) word family lists?
2. To what extent can L2 learners encounter words of the Academic Spoken Word List (ASWL), which occur in academic lectures and seminars, in YouTube worldwide trending videos at different levels of encounters?

3.1 The Corpus of YouTube Worldwide Trending Videos

A bespoke corpus of YouTube worldwide trending videos was created for this project. Once the author received ethical approval from the university's (University of Dundee) ethics committee (E2019-114), the transcripts of the top 10 worldwide trending videos, which were publicly available, were saved once every week for 45 weeks in 2020 and 2021. Apart from being in the top 10 worldwide trending videos, there were two main criteria for inclusion of the videos: (1) each video needed to be at least one minute long to allow for a text of at least 100 tokens for corpus analysis; (2) the video had to be in English. Only when these two criteria were not met within the top 10 worldwide trending videos, the videos were selected from further down the list. Out of 450 videos, the transcripts of 62 videos were provided by the creators. For the rest of the videos, Google's automatic speech-to-text service embedded within YouTube was used. As Coats (2020) reported, word error rate (WER) was found to be 5–6%, which was comparable to the error rate of human transcribers, and in another study, WER was only 3% (Proksch et al., 2019) for YouTube videos, achieving a very high accuracy rate of transcription. In the corpus of this study, five randomly selected videos were manually transcribed and compared with the transcriptions that were provided by YouTube. The WER was only 6%, and hence the transcription accuracy of YouTube videos was deemed reliable.

Table 1 shows the characteristics of the corpus of YouTube worldwide trending videos. The mean video duration of YouTube worldwide trending video was 11 min 35 s, as seen in Table 1, and the mean video duration varied across the genres in the corpus. These videos were classified into a number of genres, including vlogs, songs, and news. In order to classify the videos into genres, the descriptions of the videos were read, and a short segment of each video was watched to identify the main communicative function of the video. As shown in Table 1, the majority of YouTube worldwide trending videos were vlogs (video blogs) in which creators shared snippets from their life, talked about personal and social issues or reviewed products (Munnukka et al., 2019), followed by songs. The vlog, a user-generated video, is regarded as a digital genre (see Burgess & Green, 2018). Each vlog was coded based on its subject content to reveal the content of the vlogs in the corpus of YouTube worldwide trending videos in this study. As seen in Table 2, the videos of *personal experiences*, in which video bloggers (vloggers) talked about their own daily life experiences, such as their relationships, break-up and mental health challenges, were the most frequent category. The second most frequent category was *product review* in which vloggers reviewed and promoted products, such as make-up and technological products. *Challenge* videos, the most frequent third category, were videos in which vloggers and their friends did fun and challenging activities, such as eating increasingly spicy food while talking. *Challenge* videos were followed by *do-it-yourself* videos that involved making or building things themselves and encouraging their audience to do the same. The most frequent fifth category was social issues in which vloggers discussed social (in)justice issues, such as activism and racism, without focusing on their own experiences. The vlogs that did not fit into these categories were coded as *other*, as Table 2 shows.

The other genres, including news, and adverts were much fewer in the corpus because representativeness was prioritised over balance in the corpus in order to capture trending genres on YouTube. However, since this relatively small corpus of 828,647 tokens could only represent a snapshot of worldwide trending videos published on YouTube for a short time, the representativeness of the corpus is limited (see McEnery & Hardie, 2012 for representativeness), and the corpus of YouTube worldwide trending videos was used as a proxy for trending videos for a limited time frame.

Table 1 The corpus of YouTube worldwide trending videos

Genre	Number of texts	Mean text length	Mean video duration	Corpus size
Vlog	268	2392	14 min 15 s	641,151
Song	98	440	3 min 46 s	43,113
News	13	1625	10 min 2 s	21,125
Entertainment	12	2482	15 min 39 s	29,779
Advert	11	1621	16 min 45 s	17,828
Sports	7	1303	7 min 59 s	9,123
Trailer	7	1606	2 min 40 s	11,244
Science	7	324	14 min 26 s	2,271
Gaming	6	2197	14 min 16 s	13,179
Magazine	6	2016	12 min 57 s	12,098
Animation	4	1413	10 min	5,653
Cartoon	4	924	6 min 5 s	3,696
Documentary	3	2132	22 min 8 s	6,396
Technology	3	3234	11 min 38 s	9,702
Recipe	1	NA	NA	2,289
Total	450	1841	11 min 35 s	828,647

Table 2 The subject content of the vlogs in the corpus of YouTube worldwide trending videos

Number of vlogs	Subject content of the vlogs
99	Personal experiences
85	Product review
42	Challenge videos
19	Do-it-yourself videos
16	Social issues
7	Other

3.2 Data Analysis Procedures

In this study, the unit of analysis was a word family based on the premise that L2 learners who know one of the members of a word family may recognise or understand other members of the word family (Webb & Nation, 2013). This assumption has been supported in previous empirical research on L2 vocabulary research (e.g. Laufer & Cobb, 2020; Sasao & Webb, 2017). Laufer and Cobb (2020, p. 996) note that for comprehension, "what is required is not knowledge of the entire word families of basewords, but the inflections of basewords and a limited number of the most common affixes that participate in the construction of some derived words".

A freely available corpus tool, AntWordProfiler (Anthony, 2021), was used to analyse lexical coverage of the corpus of YouTube trending videos and to examine the potential of incidental vocabulary learning for academic lectures and seminars through watching YouTube trending videos. In order to address the first research question, Nation's (2018) British National Corpus (BNC)/Corpus of Contemporary American English (COCA) word family lists were used. These lists include 25,000 word families that were created based on the frequency and range of word families in the BNC and the COCA. Nation's (2018) BNC/COCA word family lists are the most recent and largest lists of word families of the general English, and they have been used in a number of lexical coverage studies (e.g. Dang, 2020; Nurmukhamedov & Sharakhimov, 2021). Apart from the 25,000 word families, Nation's lists also include the lists of proper nouns, marginal words (e.g. *wow*, *sh*), compounds, and acronyms. The list of proper nouns was edited to add proper nouns (e.g. *Dua Lipa*) that occurred in the corpus of YouTube trending videos, and several acronyms (e.g. COVID, COVID-19) were added to the list of acronyms. Both 95% and 98% lexical coverage of the corpus of YouTube trending videos was reported since van Zeeland and Schmitt (2013) noted that 95% lexical coverage would be necessary for good listening comprehension. For optimal comprehension, this figure was reported to be 98% (van Zeeland & Schmitt, 2013). In addition to the whole corpus of YouTube worldwide trending videos, the lexical coverage of vlogs, which constituted 60% ($n = 268$ out of 450 videos) of the corpus, was explored separately because it is much more likely for L2 learners to encounter the vlog genre than the other genres when they watch popular YouTube videos (see Burgess & Green, 2018).

In order to answer the second research question, the word families within the first three levels of the Academic Spoken Word list (ASWL) (Dang et al., 2017) were used to examine the number of encounters with the academic spoken word families of L2 learners, by watching YouTube trending videos. The ASWL is the most recent and comprehensive list of 1,741 word families that were created on the basis of frequency, range, and dispersion of word families in lectures, seminars, labs and tutorials at universities in English-speaking countries (see Dang et al., 2017 for the sources of the academic spoken corpora). As for the number of encounters, this study used a number of different cut-offs, including 1 or more, 5 or more, 10 or more, and 15 or more. Previous empirical research suggested that between 6 to 20 encounters were needed to acquire vocabulary incidentally through reading (Webb & Nation, 2013). Nation (2014, p. 3) noted that "the moderately safe goal of 12 repetitions is taken as the minimum" for incidental vocabulary acquisition through reading input. In listening, van Zeeland and Schmitt (2013) reported that at least 15 encounters may be needed to incidentally learn vocabulary. Such frequency thresholds may be arbitrary and context-specific; however, it is clear from recent studies on vocabulary learning from viewing that the more frequently words are encountered, the more likely that they will be acquired (e.g. Peters & Webb, 2018; Peters et al., 2016). Hence, following Dang (2020), the number of encounters with different frequency cut-offs was explored in this study. Also, the number of encounters with words in the ASWL was identified in the whole corpus (450 videos) and a randomly selected 10, 40, and 120 videos out of the whole corpus to explore the degree of exposure to academic spoken words in fewer YouTube trending videos because it may not be possible for L2 learners to watch 450 videos on YouTube. The findings of the first research question on the lexical coverage of YouTube trending videos and vlogs informed the approach to the second question since there was only a slight difference between the whole corpus and vlogs in terms of 98% lexical coverage, which may be due to the smaller size of the vlog subcorpus. Hence, rather than analysing vlogs separately, the analysis of encounters with the words in the ASWL involved the whole corpus and randomly selected 10, 40, and 120 videos.

4 Findings and Discussion

This section first reports the lexical demands of YouTube worldwide trending videos and then explores the potential of YouTube worldwide trending videos for incidental vocabulary learning for academic lectures and seminars.

4.1 Lexical Demands of YouTube Worldwide Trending Videos

Based on the lexical coverage analysis of YouTube worldwide trending videos, the vocabulary size that was needed to reach 95% coverage of YouTube worldwide trending videos was the most frequent 3,000 word families with proper nouns and marginal words, as seen in Table 3. Learners would need knowledge of the most frequent 6,000 word families when proper nouns and marginal words were excluded. In L2 vocabulary research, proper nouns and marginal words are associated with a low-level learning burden, and thus included in the lexical coverage figures (Webb & Rogers, 2009a). Although many learners may understand familiar proper nouns that they recognise through exposure to popular culture, it is possible that learners may not recognise unfamiliar or infrequent proper nouns (see Brown, 2010; Klassen, 2021). Therefore, cumulative coverage figures were provided both including and excluding proper nouns and marginal words, as shown in Table 3. Learners would need to have a grasp of the most frequent 9,000 word families with proper nouns and marginal words to reach 98% coverage of the YouTube worldwide trending videos.

The whole corpus of YouTube worldwide trending videos included a number of genres, including vlogs and songs. Since vlogs constituted the largest genre group of

Table 3 Coverage for the whole corpus of YouTube worldwide trending videos (%)

World list	Coverage at each 1,000 word level	Cumulative coverage without proper nouns & marginal words	Cumulative coverage with proper nouns & marginal words
1,000	86.62	86.62	88.93
2,000	4.55	91.17	93.48
3,000	1.8	92.97	95.28*
4,000	0.98	93.95	96.26
5,000	0.66	94.61	96.92
6,000	0.45	95.06*	97.37
7,000	0.27	95.33	97.64
8,000	0.23	95.56	97.87
9,000	0.2	95.76	98.07**
10,000	0.11	95.87	98.18
11,000	0.1	95.97	98.28
12,000	0.09	96.06	98.37
13,000	0.05	96.11	98.42
14,000	0.04	96.15	98.46
15,000	0.03	96.18	98.49
16,000	0.05	96.23	98.54
17,000	0.04	96.27	98.58
18,000	0.03	96.3	98.61
19,000	0.01	96.31	98.62
20,000	0.02	96.33	98.64
21,000	0.01	96.34	98.65
22,000	0.01	96.35	98.66
23,000	0.01	96.36	98.67
24,000	0	96.36	98.67
25,000	0	96.36	98.67
Proper nouns	1.23		
Marginal words	1.08		
Compounds	0.34		
Acronyms	0.1		
Off list	0.9		

*Reaching 95% coverage, **Reaching 98% coverage the corpus, the lexical coverage of vlogs was examined separately to find whether it was similar to the whole corpus. As Table 4 shows, knowledge of the most frequent 3,000 word families with proper nouns and marginal words would be required to reach 95% coverage of the subcorpus of YouTube worldwide trending vlogs, which was the same figure as that of the whole corpus. In order to reach 98% coverage of vlogs, knowledge of the most frequent 8,000 word families with proper nouns and marginal words would be necessary. This suggests that lexical demands of vlogs were lower than the whole corpus of YouTube worldwide trending videos in terms of 98% lexical coverage. The slightly lower demands of vlogs for 98% coverage may be associated with the smaller size of the subcorpus of vlogs in comparison to the whole corpus as well as the repetition of personal pronouns in the subcorpus of vlogs since the vlogs' content mainly included personal experiences of vloggers, such as their lifestyles and relationships with friends and family members. However, the trending vlogs in the corpus of this study did not just represent personal matters. There were also trending vlogs that focused on social (in)justice issues, including racism and the movement for Black Lives Matter, and do-it-yourself videos (please see Table 2 for the subject content of vlogs).

The lexical profile of YouTube worldwide trending videos showed similarities with that of a number of other spoken genres in previous lexical coverage studies (e.g. Kaneko, 2015; Nurmukhamedov & Sharakhimov, 2021; Tegge, 2017; Webb & Rodgers, 2009a, 2009b). With regard to 95% lexical coverage, the lexical demands of the whole corpus of YouTube trending videos and vlogs were quantitatively the same as those of the listening section of the TOEFL iBT (Kaneko, 2015), popular English podcasts (Nurmukhamedov & Sharakhimov, 2021), popular songs (Tegge, 2017), movies (Webb & Rodgers, 2009a) and television programs (Webb & Rodgers, 2009b). However, the whole corpus of YouTube trending videos and vlogs was lexically more demanding in terms of 98% coverage in comparison to the above-mentioned spoken genres. In order to reach 98% coverage of the whole corpus of YouTube trending videos and vlogs, the knowledge of the most frequent 9,000 and 8,000 word families was needed respectively. In this aspect, the academic spoken English (Dang & Webb, 2014) and TED talks (Nurmukhamedov, 2017), both of which required a grasp of the most frequent 8,000 word families, were more similar to the corpus of YouTube trending videos and vlogs in terms of lexical demands. It should be noted that English as a foreign language (EFL) learners may not gain the knowledge of the most frequent 9,000 word families before or at university since the opportunities to be exposed to English may be more limited in EFL contexts than in English as a second language (ESL) contexts. In ESL contexts, English is one of the official languages while this is not the case in EFL contexts which limit exposure to English for L2 learners. Nation (2014, p. 13) argues that "a mixed written and spoken corpus provides better opportunities" to learn the most frequent 9,000 word families than the written or spoken input alone. This suggests that listening input is as important as reading input.

Table 4 Coverage for the subcorpus of YouTube worldwide trending vlogs (%)

Word list	Coverage at each 1,000-word level	Cumulative coverage without proper nouns & marginal words	Cumulative coverage with proper nouns & marginal words
1,000	86.93	86.93	89.19
2,000	4.53	91.46	93.72
3,000	1.71	93.17	95.43*
4,000	0.98	94.15	96.41
5,000	0.65	94.8	97.06
6,000	0.45	95.25*	97.51
7,000	0.26	95.51	97.77
8,000	0.23	95.74	98**
9,000	0.2	95.94	98.2
10,000	0.11	96.05	98.31
11,000	0.11	96.16	98.42
12,000	0.1	96.26	98.52
13,000	0.05	96.31	98.57
14,000	0.05	96.36	98.62
15,000	0.03	96.39	98.65
16,000	0.06	96.45	98.71
17,000	0.03	96.48	98.74
18,000	0.03	96.51	98.77
19,000	0.01	96.52	98.78
20,000	0.02	96.54	98.8
21,000	0	96.54	98.8
22,000	0.01	96.55	98.81
23,000	0	96.55	98.81
24,000	0	96.55	98.81
25,000	0	96.55	98.81
Proper nouns	1.15		
Marginal words	1.11		
Compounds	0.33		
Acronyms	0.09		
Off list	0.77		

*Reaching 95% coverage, **Reaching 98% coverage

4.2 The Potential of YouTube Worldwide Trending Videos for Incidental Vocabulary Learning for Academic Lectures and Seminars

Table 5 shows the encounters with ASWL level 1 word families in YouTube trending videos with different frequency cut-offs for randomly selected 10, 40, 120 videos and the whole corpus. Accordingly, by watching just 10 videos, it would be possible for L2 learners to encounter 314 ASWL word families out of 830 five times or more. The number of ASWL word families that L2 learners would frequently encounter increases as more videos are watched, providing greater potential for incidental vocabulary learning. As the frequency of vocabulary in the input was found to be associated with learning and later use (e.g. Ellis & Wulff, 2015), it would be unlikely for word families that only occurr between one to four times in YouTube trending videos to be acquired incidentally. On the other hand, when 120 YouTube trending videos are watched, it would be possible to learn the majority of the ASWL level 1 word families incidentally though retention of such vocabulary would depend on a number of other factors, including learner characteristics (see Montero Perez, 2020; Reynolds, 2020).

The learning opportunities for ASWL level 2 word families would be reduced when only 10 or 40 videos are watched since only 23 out of 454 word families are repeated 10 or more times when 40 videos are watched, and such encounters further decrease when only 10 videos are watched. As Table 6 shows, watching 120 videos would provide exposure to more than half the ASWL level 2 word families (n = 268) five times or more. Although this figure would be proportionately less than the ASWL level 1 word families, YouTube trending videos could still be considered as fruitful supplementary learning resources for frequent academic words in lectures and seminars.

YouTube trending videos would provide much less exposure to ASWL level 3 word families, in comparison to the ASWL level 1 and 2 word families. In order to encounter the majority of the ASLW level 3 word families five or more than five times, 450 videos would need to be watched. When 120 videos are watched, L2 learners would only encounter 105 ASWL level 3 word families out of 380 word families five times or more, as Table 7 shows. Nevertheless, for L2 learners who would have the habit of watching YouTube trending videos for extensive viewing in their free time, YouTube trending videos would constitute one of the input resources for academic spoken words. Considering that ASWL level 3 word families are less frequent in general English than the first two levels of word families, at least 15 encounters (see van Zeeland & Schmitt, 2013) might be needed for incidental vocabulary learning.

YouTube trending videos are found to provide relatively frequent encounters to the majority of ASWL level 1 and level 2 words even when 120 videos are watched, creating opportunities for incidental vocabulary learning for L2 learners. This finding is in line with previous research that identified movies and television programmes as vocabulary input sources for student reading and writing at the secondary school

Table 5 Encounters with the ASWL level 1 word families ($n = 830$) in YouTube trending videos

Number of encounters	N of word families encountered	Mean of encounters (frequencies)	SD of encounters (frequencies)
1 or more			
10 videos	612	23.32	77.45
40 videos	792	60.31	232.08
120 videos	824	171.66	659.28
450 videos	830	701.26	2705.5
5 or more			
10 videos	314	43.56	104.23
40 videos	580	81.44	268.16
120 videos	772	183.03	679.65
450 videos	827	703.79	2710.10
10 or more			
10 videos	198	65.24	126.41
40 videos	423	109.16	309.54
120 videos	679	207.15	721.42
450 videos	818	711.46	2723.99
15 or more			
10 videos	138	88.49	145.54
40 videos	322	139.72	349.34
120 videos	596	234.37	766.14
450 videos	803	724.52	2747.66

level (Green, 2020) and medical television programmes as sources of discipline-specific vocabulary in lectures and seminars in the field of medicine (Dang, 2020). The findings of this study are promising given that there are notable differences in the topics and genres between YouTube trending videos and academic lectures and seminars at universities. It should be acknowledged that ASWL level 1 and level 2 words contain general high-frequency words because the creators of ASWL included general high-frequency words that occur frequently in academic spoken texts with a wide range in the ASWL (see Dang et al., 2017). Dang et al. (2017) argue that we should not assume that all L2 learners would know such general high-frequency words before encountering academic words in lectures and seminars (Dang et al., 2017). In the literature, the extent of the academic nature of words in the academic word lists has been a matter of discussion (Brezina & Gablasova, 2015; Coxhead, 2011; Dang et al., 2017; Hyland & Tse, 2007). For example, Hyland and Tse (2007) argue that academic words may have different meanings, functions, and collocational patterns across the subjects. In a similar vein, Brezina and Gablasova (2015) note that some frequently occurring *academic* words could be categorised as general vocabulary rather than as academic vocabulary in the Academic Wordlist (Coxhead,

Table 6 Encounters with the ASWL level 2 word families ($n = 454$) in YouTube trending videos

Number of encounters	N of word families encountered	Mean of encounters (frequencies)	SD of encounters (frequencies)
1 or more			
10 videos	167	2.29	2.34
40 videos	317	4.03	5.03
120 videos	418	9.44	10.42
450 videos	450	36.91	39.10
5 or more			
10 videos	16	7.94	3.79
40 videos	87	9.20	7.25
120 videos	268	13.41	11.16
450 videos	424	39	39.32
10 or more			
10 videos	5	12.4	3.58
40 videos	23	17.65	9.93
120 videos	142	19.32	12.63
450 videos	369	43.73	40.06
15 or more			
10 videos	1	18	NA
40 videos	14	21.36	11.30
120 videos	73	26.40	14.36
450 videos	325	48.02	40.84

2000). It is very likely that there would be an overlap between general and academic words in academic wordlists (Coxhead, 2011; Dang et al., 2017) when the criteria of *frequency* and *range* in academic spoken or written texts are used when creating wordlists. Therefore, it is important to take into account the operationalisation and criteria of academic wordlists in lexical coverage studies.

The usage-based accounts of language learning emphasise the frequency of occurrence and dispersion in language acquisition, among many other factors (e.g., Ellis & Wulff, 2015). When we take into account the usage-based accounts of language learning and the necessity of 15 or more encounters for incidental vocabulary acquisition through listening (van Zeeland & Schmitt, 2013), the majority of ASWL level 1 word families can potentially be learned by watching 120 YouTube trending videos, as the findings of this study show (see Table 5). In order for learners to encounter most of the ASWL level 2 word families, watching 450 YouTube trending videos would be recommended, as can be inferred from the findings of this study (see Table 6). Hence, it is useful for L2 learners to watch YouTube trending videos as an out-of-class activity that would lead to cumulative encounters of the academic spoken

Table 7 Encounters with the ASWL level 3 word families ($n = 380$) in YouTube trending videos

Number of encounters	N of word families encountered	Mean of encounters (frequencies)	SD of encounters (frequencies)
1 or more			
10 videos	58	1.81	1.46
40 videos	171	2.89	3.66
120 videos	290	5.09	7.58
450 videos	363	17.11	31.35
5 or more			
10 videos	4	6	1.41
40 videos	23	9.43	6.65
120 videos	105	10.59	10.49
450 videos	278	21.63	34.59
10 or more			
10 videos	–	–	–
40 videos	8	15.38	8.63
120 videos	38	17.82	14.89
450 videos	194	28.05	39.73
15 or more			
10 videos	–	–	–
40 videos	2	26.5	13.44
120 videos	17	25.12	20.16
450 videos	134	35.19	46.09

word families in the long run, which would potentially result in academic spoken vocabulary learning.

5 Conclusion and Implications

This study provided new insights into YouTube worldwide trending videos' lexical coverage and their potential as sources of incidental vocabulary learning for L2 learners. First, YouTube trending videos required knowledge of the most frequent 3,000 word families to reach 95% coverage. Second, YouTube trending videos constituted potential sources of incidental vocabulary learning for academic lectures and seminars at universities in English-speaking countries, providing relatively frequent encounters to the level 1 and level 2 word families for L2 learners. This is an encouraging finding given that YouTube trending videos and academic lectures and seminars are different in terms of formality, topics, and genre.

The findings of this study have significant implications for incidental vocabulary learning for L2 adolescent and adult learners globally. The lexical demands of YouTube trending videos were similar to those of other spoken genres, including popular English podcasts (Nurmukhamedov & Sharakhimov, 2021), movies (Webb & Rodgers, 2009a) and television programs (Webb & Rodgers, 2009a) for good comprehension (95% coverage). Therefore, YouTube videos could be good alternatives to these genres to use for vocabulary activities, since YouTube trending videos potentially attract students' interest more due to their popularity among adolescents and adults, increasing their motivation for watching them. In order for students to better comprehend YouTube trending videos, frequent and/or key proper nouns and marginal words could be introduced first before watching the videos (Brown, 2010; Klassen, 2021) since such words may be unfamiliar to EFL learners. YouTube also provides additional affordances, such as captions and playback speed. For lower-proficiency learners, teachers could use captions and reduce playback speed to enhance learners' comprehension though the use of captions may not always be beneficial for lower-proficiency L2 learners due to their limited working memory capacity (see Moreno Perez, 2020). Hence, YouTube trending videos may arguably be more useful for intermediate and advanced learners of English, English for academic or specific purposes learners who prepare for their university studies.

It is probable that some YouTube trending videos may not be suitable to use for teaching English to adolescent learners or any learners in some contexts due to their inappropriate content or the presence of taboo words in such videos. In this case, teachers can use a free third party application, such as Edpuzzle (see Mischel, 2019) to edit a video in order to make it more appropriate for their learners. Teachers may also incorporate intentional vocabulary learning activities, such as focus-on-form and meaning-focused activities for discourse markers (e.g., *well*) and multi-word sequences (e.g., *keep in mind*), vocabulary journals for newly encountered words for their learners to supplement potential incidental vocabulary gains. In addition to reading, it is useful for the class to watch a diverse range of genres of YouTube videos regularly since cumulative encounters with ASWL word families will increase as the number of videos are watched, and a diverse range of genres is more likely to include a wider range of word families than a collection of videos of one genre (see Nation, 2014). It may be more feasible for L2 learners of English to watch YouTube trending videos in an out-of-classroom context rather than in ESL/EFL classes.

The present study also has important implications for L2 learners for their out-of-class learning activities. Given that YouTube trending videos provide relatively frequent encounters to academic spoken word families used in academic lectures and seminars at universities, it is recommended that ESL/EFL learners watch YouTube trending videos weekly as part of their preparation for university classes. When L2 learners watch 120 trending videos, it would be possible to encounter the majority of the ASWL word families five or more than five times, providing opportunities for incidental learning. L2 learners could be encouraged to pay attention to lexical realisations of sharing perspectives on personal and social issues, interpreting content, and interacting with audience members since such functions also exist in university lectures and seminars (see Deroey & Taverniers, 2011). In this way, L2 learners, who watch YouTube videos as part of their entertainment or leisure activities, are likely to be on the path from informal language learning to the development of academic listening.

Lexical coverage analysis of YouTube videos may not reveal the true nature of spoken language since spoken language includes multi-word sequences to a greater extent than written language (e.g. Biber et al., 2004). Puimège and Peters (2020) found that watching a documentary (audio-visual input) was effective to acquire multi-word sequences incidentally for L2 learners of English who had upper-intermediate proficiency in English. This suggests that it is possible to learn multi-word sequences incidentally by watching YouTube videos though empirical evidence is needed for learners of different proficiency levels. Although it can be argued that many ASWL word families can be learned through reading as well, comparing the encounters of ASWL word families through reading versus listening is beyond the focus of this paper.

Finally, several limitations to this study need to be acknowledged in terms of methodology and analysis. The analysis in this study was limited to lexical coverage of YouTube trending videos and their potential for incidental vocabulary learning for academic lectures and seminars. It is possible that the word error rate (6%) might have influenced the results of lexical coverage of YouTube trending videos. Although this study found that it is possible to learn academic spoken vocabulary incidentally from YouTube trending videos, it does not necessarily argue that this would be the case for all L2 learners. Incidental vocabulary acquisition was found to be context-specific, depending on a number of L2 learner characteristics, including working memory and L2 proficiency and other external factors (see Montero Perez, 2020; Reynolds, 2020). Additionally, the corpus of YouTube trending videos was small, representing a certain period of time in 2020 and 2021. It should also be acknowledged that YouTube is blocked in several countries, including China although users may still access the site using a virtual private network service.

This corpus-based analysis did not address the polysemous nature of vocabulary (see Deignan et al., 2022), and it remains unknown to what extent multiple meanings of vocabulary can potentially be acquired from extensive viewing. Future studies should test to what extent L2 learners can incidentally learn vocabulary by watching YouTube trending videos within a longitudinal experimental design with learners of different proficiency levels. Further research is necessary to examine to what extent YouTube trending videos can provide encounters with multi-word sequences in empirically derived lists (e.g., a phrasal expression list in Martinez & Schmitt, 2012; an academic formulas list in Simpson-Vlach & Ellis, 2010; and the academic collocation list in Ackermann & Chen, 2013) and to what extent learners can incidentally acquire such multi-word sequences through exposure to audio-visual input. This will be a promising direction given the highly phrasal nature of spoken language.

References

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL)—A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Anthony, L. (2021). AntWordProfiler (Version 1.5.1) [Computer Software]. https://www.laurenceanthony.net/software

Arndt, H. L., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning and Technology*, *22*(3), 124–142. http://hdl.handle.net/10125/44660

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at … : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Brown, D. (2010). An improper assumption? The treatment of proper nouns in text coverage counts. *Reading in a Foreign Language, 22*(2), 355–361. https://scholarspace.manoa.hawaii.edu/bitstream/10125/66842/1/22_2_10125_66842_brown.pdf

Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture* (2nd ed.). John Wiley & Sons.

Coats, S. (2020). Articulation rate in American English in a corpus of YouTube videos. *Language and Speech*, *63*(4), 799–831. https://doi.org/10.1177/002383091989720

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213. https://doi.org/10.2307/3587951

Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly, 45*(2), 355–362. https://doi.org/10.5054/tq.2011.254528

Dang, T. N. Y. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related tv programs: Insights from medical corpora. *TESOL Quarterly, 54*(2), 436–459. https://doi.org/10.1002/tesq.552

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning, 67*(4), 959–997. https://doi.org/10.1111/lang.12253

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*(1), 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Deignan, A., Candarli, D., & Oxley, F. (2022). *The linguistic challenge of the transition to secondary school: A corpus study of academic language*. Routledge.

Deroey, K. L. B., & Taverniers, M. (2011). A corpus-based study of lecture functions. *Moderna Sprak, 105*(2), 1–22. https://ojs.ub.gu.se/index.php/modernasprak/article/view/793

Ellis, N., & Wulff, S. (2015). Usage-based approaches in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75–93). Routledge. https://doi.org/10. 4324/9781410615299

Folkvord, F., Bevelander, K. E., Rozendaal, E., & Hermans, R. (2019). Children's bonding with popular YouTube vloggers and their attitudes toward brand and product endorsements in vlogs: An explorative study. *Young Consumers, 20*(2), 1–15. https://doi.org/10.1108/YC-12-2018-0896

Green, C. (2020). Extensive reading and viewing as input for academic vocabulary: A large-scale vocabulary profile coverage study of students' reading and writing across multiple secondary school subjects. *Lingua, 239*, 1–18. https://doi.org/10.1016/j.lingua.2020.102838

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Kaneko, M. (2015). Vocabulary size required for the TOEFL iBT listening section. *The Language Teacher*, *39*(1), 9–14. https://jalt-publications.org/files/pdf-article/39.1tlt_art2.pdf

Klassen, K. (2021). Proper name theory and implications for second language reading. *Language Teaching*, 1–7. https://doi.org/10.1017/S026144482100015X

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

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics, 41*(6), 971–998. https://doi.org/10.1093/applin/amz051

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. https://doi.org/10.1093/applin/ams010

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Mischel, L. J. (2019). Watch and learn? Using EDpuzzle to enhance the use of online videos. *Management Teaching Review, 4*(3), 283–289.

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video. *Studies in Second Language Acquisition, 42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Munnukka, J., Maity, D., Reinikainen, H., & Luoma-aho, V. (2019). "Thanks for watching". The effectiveness of YouTube vlog endorsements. *Computers in Human Behavior, 93*, 226–234. https://doi.org/10.1016/j.chb.2018.12.014

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language, 26*(2), 1–16. https://nflrc.hawaii.edu/rfl/item/303

Nation, I. S. P. (2018). *The BNC/COCA word family lists*. Retrieved June 1, 2021 from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Cengage Learning.

Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal, 8*(4), 768–790. https://doi.org/10.1002/tesj.323

Nurmukhamedov, U., & Sharakhimov, S. (2021). Corpus-based vocabulary analysis of English podcasts. *RELC Journal*. https://doi.org/10.1177/0033688220979315

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching, 52*(2), 188–200. https://doi.org/10.1017/S0261444819000028

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*, 551–577. https://doi.org/10.1017/S0272263117000407

Proksch, S. O., Wratil, C., & Wäckerle, J. (2019). Testing the validity of automatic speech recognition for political text analysis. *Political Analysis, 27*(3), 339–359. https://doi.org/10.1017/pan.2018.62

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

Reynolds, B. L. (2020). Situated incidental vocabulary acquisition: The effects of in-class and out-of-class novel reading. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2019-0059

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research, 21*(1), 12–30. https://doi.org/10.1177/1362168815586083

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Soyoof, A., Reynolds, B. L., Vazquez-Calvo, B., & Mclay, K. (2021). Informal digital learning of English (IDLE): A scoping review of what has been done and a look towards what is to come. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2021.1936562

Tan, E. (2013). Informal learning on YouTube: Exploring digital literacy in independent online learning. *Learning, Media and Technology, 38*(4), 463–477. https://doi.org/10.1080/17439884.2013.783594

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System, 67*, 87–98. https://doi.org/10.1016/j.system.2017.04.016

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, *15*(4), 497–519. https://doi.org/10.1075/ijcl.15.4.03web

Webb, S., & Nation, I. S. P. (2013). Computer-assisted vocabulary load analysis. In C. Chappelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 1–10). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0179

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning, 59*(2), 235–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Duygu Candarli (Ph.D.) is currently Lecturer in TESOL at the University of Southampton. She specialises in second language writing, writing assessment, corpus linguistics, and academic discourse. She has published in international journals such as *Reading and Writing, Corpora*, and *Journal of English for Academic Purposes*. She is the co-author of *The Linguistic Challenge of the Transition to Secondary School: A Corpus Study of Academic Language* (Routledge, 2022).

TED Ed for Incidental L2 Academic Vocabulary Learning: A Corpus-Driven Study

Chi-Duc Nguyen

Abstract This corpus-driven study examined the potential benefit of 360 TED-Ed videos with a total of 235,553 running words for incidental L2 academic vocabulary learning. This benefit was operationalized by the frequency that word families in Coxhead's (2000) Academic Word List and Dang et al. (2017) Academic Spoken Word List occurred within and across the selected videos. Using the RANGE program (Nation & Heatley, 2002) for data analysis, this study found that the videos did not pose a huge lexical demand to L2 learners to reach 95% (4,000 word families) and 98% (7,000 word families) lexical coverage, compared to other academic spoken discourse such as lectures, seminars or TED Talks. Of all word families in one video, 5.48% were found to be academic written vocabulary and 82.69% academic spoken vocabulary. Only 2.95% and 6.26% of all academic written and spoken word families in one video occurred eight times or over. These figures, however, soared to 29.90% and 49.90% when L2 learners viewed 30 videos on the same theme. These findings suggest that TED Ed videos are indeed a potential source for incidental L2 academic vocabulary learning, especially when L2 learners carry out extensive or narrow listening/viewing in their self-study program.

Keywords TED Ed videos · Academic vocabulary · Incidental learning · Lexical load · Word repetition

1 Introduction

One potential benefit of foreign/second language (L2) meaning-focused listening/viewing activities is that they provide opportunities for incidental vocabulary acquisition or, in other words, lexical uptake as a byproduct when students' focal attention is placed on comprehending the input content (Ellis, 1999; Nation & Webb, 2011). This benefit is perhaps of particular importance in the context of extensive listening/viewing where L2 students carry out a regular practice of listening/viewing

C.-D. Nguyen (✉)
VNU University of Languages and International Studies, Vietnam National University, Hanoi, Vietnam
e-mail: ducnc@vnu.edu.vn

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_9

for pleasure on the topics of their own interests at home (Webb & Chang, 2015). Thanks to the advances of modern technology L2 students now have relatively easy access to a wide range of authentic listening/viewing materials outside of their language classroom, which might be a great source for their vocabulary learning in the wild, including academic vocabulary learning. TED Ed videos might be a fine example of materials as such.

TED Ed, in nature, is a virtual platform where teachers and students worldwide share their knowledge and skills in the form of a video clip. In the official website https://ed.ted.com/, there now exists about 1,464 videos (as of February 01, 2020) being grouped under twelve broad themes. From the perspectives of incidental L2 academic vocabulary acquisition, these videos can offer various learning affordances both in terms of their discourse features and recording formats. However, no research to date has examined the merits of these videos for the above purpose. The present study aims to address this research gap. To this end, it examines the potential benefit of a collection of TED Ed videos for incidental L2 academic vocabulary learning by looking at the frequency that academic vocabulary—both spoken and written—occurs and reoccurs within and across the selected videos. The assumption behind this study is that the more often L2 students encounter academic words in the videos, the more likely they are able to pick up the knowledge related to these lexical items during their listening/viewing process. The findings of this study are expected to inform L2 students in their search for suitable materials to incorporate in their extensive listening/viewing in the wild that can foster both content and academic vocabulary gain.

2 Background to This Study

2.1 TED Ed Videos

As an extension of the original TED Talks, TED Ed is created to reinforce a long-held belief of TED founders that ideas have the power to change attitudes, lives, and even the world, and thus they should be widespread. TED Ed is a virtual platform where teachers around the globe can share their lessons in the form of a video clip with textual (i.e., subtitles or auto-translation), audial (i.e., soundtrack), and visual (i.e., animated or video-recorded images) support. The target audience of these videos are all students from the primary to the tertiary education level. This platform also has a built-in discussion board where viewers can exchange their ideas about the viewing content and far beyond. At the time data for the present study were collected there were 1,464 TED Ed videos in total, which were categorized under twelve broad themes, namely Arts (104 videos), Business and Economics (52 videos), Technology, Engineering and Design (218 videos), Health (310 videos), Language and Literature (199 videos), Mathematics (154 videos), Philosophy and Religion (81 videos), Psychology (216 videos), Science and Technology (626 videos), Social Studies

(391 videos), Teaching and Education (30 videos), and Thinking and Learning (238 videos). As these themes were overlapping to some extent, some videos were placed under more than one theme.

As far as incidental L2 academic vocabulary acquisition is concerned, these videos can offer various learning affordances. First of all, unlike their precedential TED Talks videos that target a lay and a semi-lay audience, TED Ed videos target teachers and students worldwide and thus these videos often disseminate academic and even technical topics. Therefore, they might give L2 students a valuable opportunity to encounter a fair amount of academic vocabulary in authentic spoken language use. Second, as these videos are categorized according to themes and even subtle topics that belong to the same theme, L2 students can carry out the practice of narrow viewing, which in turn can help them to encounter a smaller number of new words, compared to the case of wide viewing and can also increase the chance that they reencounter unfamiliar academic words in a wider range of contexts (Webb & Rodgers, 2009). From my initial analysis of these videos and their written transcripts, I find that speakers in these videos often give a clear definition and/or explanation whenever it comes to a new and important concept with a view to fostering their listeners' content comprehension. Such a definition and/or explanation is often found to support L2 students in their word form-meaning mapping process (Sawada, 2009; Vidal, 2003). Further, the presence of both textual and visual support in these videos are also expected to facilitate L2 listening comprehension (Sueyoshi & Hardison, 2005; Vandergrift, 2004) as well as L2 lexical segmentation and recognition (van Zeeland, 2013), which together can enhance the size of incidental (academic) vocabulary gain (Brown et al., 2008; Webb & Chang, 2015).

With regards to the practicality, these videos are all free to access and download from the official website https://ed.ted.com/ or from the popular website https://www.youtube.com/. The speed of speech delivery in these videos can be adjusted to eight levels from 0.25 to 2.00 times of the normal speed that therefore caters to different levels of L2 proficiency. These videos are about a wide range of topics and thus they can meet the general diversity in L2 students' interests. The ready-made comprehension and follow-up discussion questions that accompany each video can also facilitate students' self-study at home to a great extent. The transcripts of these videos are all freely available in the above websites so L2 students can use these transcripts to facilitate their L2 listening process. The potential benefits above altogether can make TED Ed videos a potential source of listening/viewing materials for incidental L2 academic vocabulary learning in the wild. However, to the best of my knowledge, there is no research that has examined this potential.

2.2 Academic Vocabulary

Academic vocabulary can be defined as a group of lexical items that frequently occur in academic discourse. However, they are not high-frequency words that often dominate both academic and non-academic discourse like those described in West's

(1953) General Service List. Neither are they technical words that are specific to a particular discipline. Put differently, they are a shared asset in academic discourse across various subject areas. This vocabulary is generally deemed to play a crucial role in fostering L2 students' academic studies at universities.

To inform the practice of English for Academic Purposes instruction, assessment as well as course and materials development, various attempts have been made during the last five decades to construct academic word lists (e.g., Coxhead, 2000; Dang et al., 2017; Gardner & Davies, 2014; Ghadessy, 1979; Lynn, 1973; Simpson-Vlach & Ellis, 2010; Xue & Nation, 1984). The two well-established, well-validated and thus most widely-used ones to date, however, are Coxhead's (2000) Academic Word List and Dang et al. (2017) Academic Spoken Word List (henceforth referred to as the AWL and the ASWL for short). In the present study, these two word lists were also employed to explore the benefits of TED Ed videos for incidental L2 academic vocabulary learning.

The AWL was created based on an academic written corpus that covered as many as 28 different subject areas. The number of running words in that corpus was 3,513,330. The counting unit of this word list was a word family which included a base word and all of its derived and inflected forms that were understood by learners without having to learn each separately (Bauer & Nation, 1993). A word family was qualified for inclusion in the AWL if its members occurred at least 100 times throughout the whole corpus and at least 10 times across 15 or more out of the 28 subject areas. The final AWL included a total number of 570 word families, which altogether accounted for about 10% of all word families in the above corpus.

Regarding the ASWL, the list was built upon an academic spoken corpus, which featured four typical genres of academic spoken discourse including lectures, seminars, labs and tutorials, and covered 24 different subject areas. This corpus included a total number of 13 million running words. To deserve a place in the list, a word family needed to occur at least 350 times in the above corpus, ranged across at least 12 subject areas, and had Juilland and Chang-Rodriguez' (1964) dispersion of at least 0.6 across all subject areas above. The final ASWL included 1,741 word families, which were further divided into four subtle frequency-based lists. With the support of these word families, L2 students could reach from 92 to 96% lexical coverage of academic spoken discourse, depending the genres that they were exposed to (Dang et al., 2017).

3 Literature Review

3.1 Incidental L2 Vocabulary Learning Through Academic Listening

In the academic context, previous research consistently shows that L2 students are able to acquire new vocabulary knowledge, including academic vocabulary knowledge through their meaning-focused viewing/listening to academic spoken discourse. Vidal (2003), for example, examined incidental L2 vocabulary acquisition through viewing three video-recorded lectures. In this study, 116 EFL university students in Spain were asked to view the above videos over a span of four weeks. Vocabulary gain was measured by a modified version of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993). The results showed that the research participants indeed made a noticeable vocabulary gain after viewing and this gain appeared to be relatively resistant to decay over time. One interesting feature of this study is that Vidal (2003) purposefully included a number of academic words as the target words in her experiment and the students did pick up the knowledge of these lexical items after viewing the videos.

Sawada (2009) also investigated incidental L2 vocabulary learning in an academic context, but through listening to nine TOEFL passages. In this study, 116 EFL university students in Japan were required to listen to these passages either with or without the support of L1 translation or L2 explanation of novel words occurring in the input materials. Vocabulary gain was gauged by means of a word-meaning recall and a word-meaning recognition task. The results showed that the academic listening brought about a considerable vocabulary gain, especially among those who received L1 translation and L2 explanation of the target words during their listening process.

Nguyen and Boers (2019) compared the size of vocabulary gain between (a) viewing a TED Talks video twice and (b) viewing the same video twice, but with a content retelling activity inserted in between. In their intact classes, 64 EFL university students in Vietnam were assigned to one of the two viewing conditions. Vocabulary gain was measured by a word-meaning recall task. The results showed that both viewing conditions led to lexical uptake. However, the retelling activity helped the students in the second viewing condition to make a significantly larger vocabulary gain.

It is clear from the above review that L2 students are able to pick up new lexical knowledge, including academic vocabulary knowledge through meaning-focused viewing/listening to academic spoken discourse. However, research that directly examines academic vocabulary learning from this source remains relatively scarce. Thus, more empirical studies in this area are still welcome. In addition, most of the previous research often gauged the size of vocabulary gain through viewing/listening to a single long text or a couple of short texts. A pertinent question here is whether the vocabulary gain above will be enlarged if L2 students carry out an extensive listening/viewing program or ideally a narrow listening/viewing program with a larger number of academic speeches over a longer period of time. To lay the foundation for such a study, the affordances that available academic listening materials may offer for incidental L2 academic vocabulary learning should be explored. In what follows, studies that give these learning affordances a closer look are briefly reviewed.

3.2 Potentials of Academic Listening Materials for Incidental L2 Academic Vocabulary Learning

Coxhead and Walls (2012) were among the pioneers that examined the potential benefits of academic listening materials for incidental L2 academic vocabulary learning. In their study, they built up a corpus of 60 TED Talks videos that covered six broad subject areas, including Technology, Entertainment, Design, Business, Science and Global Issues. The total number of running words in this corpus was 43,656. The counting unit of this corpus was a word family which comprised a headword and all of its derived as well as inflected forms (Bauer & Nation, 1993). Using the RANGE program for data analysis, they found that word families in Coxhead's (2000) AWL accounted for 3.90% of all word families in this corpus. Nurmukhamedov (2017) further unpacked the potential benefits of these videos by using a larger corpus of 400 TED Talks videos with 997,479 tokens in total. In this corpus, 3.79% of all word families were found to belong to the AWL.

Dang and Webb (2014) also examined the lexical features of academic spoken discourse, but with a special focus on lectures and seminars. In this study, they created a large corpus with a total of 1,691,997 running words. Using the RANGE program for data analysis, they found that word families in the AWL provided a lexical coverage of 4.41% in this corpus, far surpassing what previous research often found in their corpora of TED Talks videos. In their attempt to generate an academic spoken word list as already reviewed above, Dang et al. (2017) created two large corpora of lectures, seminars, labs, and tutorials. One included a total of 13,029,661 running words, while the other consisted of 12,740,619 tokens. Word families in Coxhead's AWL were found to account for 4.17% and 4.03% of all word families in these two corpora, respectively. Meanwhile, academic spoken vocabulary covered 90.13% of all word families in the first corpus and 89.59% of those in the second one.

Together, the studies above clearly demonstrate that L2 students do have the opportunity to encounter academic vocabulary—both spoken and written—when they view/listen to such academic spoken discourse as lectures, seminars, labs, tutorials or conference presentations. However, the quantity of this vocabulary in the above materials, especially in the case of academic written vocabulary is relatively modest, ranging only from 3.73 to 4.41% of all word families in the selected texts. Therefore, we should continue our search for other learning materials that can offer L2 students a more frequent exposure to academic vocabulary. In addition, most of previous research in this area has often been concerned with the presence of academic written vocabulary, but not academic spoken vocabulary in the target materials even though these materials are presented in the spoken form. Given the difference between the spoken and the written academic discourse as well as the difference between the AWL and the ASWL (please refer to Dang et al. (2017) for a more detailed account), such a study should balance their focus on both academic spoken and written vocabulary.

3.3 *Two Key Factors that Influence Incidental L2 Vocabulary Learning from Audio/audio-Visual Materials*

3.3.1 Lexical Coverage

Research into incidental L2 vocabulary learning from audio and audio-visual materials needs to take into account the lexical coverage or the percentage of word families that L2 students already know in a particular input text (Nation & Webb, 2011). It is because this factor is often found to foster L2 listening comprehension (van Zeeland & Schmitt, 2012), which, in turn, helps L2 students to enhance the success rate of their lexical inferencing (Sawada, 2009; Vidal, 2003).

Stæhr (2009), for example, examined the relationship between the breadth and the depth of vocabulary knowledge on the one hand and L2 listening comprehension on the other hand. In that study, 115 EFL students were invited as the research participants. Their L2 listening ability was measured by a CPE listening test and their breadth and depth of vocabulary knowledge was gauged by the Vocabulary Levels Test (Schmitt et al., 2001) and a modified version of the Word Associate Test (Read, 1993). Stæhr found a strong correlation between their L2 listening comprehension and the breadth ($r = 0.70$, $p < 0.01$) and the depth of their vocabulary knowledge ($r = 0.65$, $p < 0.01$). That study also found L2 students who already knew 90% of all running words in the input text could cultivate a reasonable level of text comprehension which was operationalized by having at least 50% of all comprehension questions answered correctly. This listening comprehension level, however, increased significantly only when they already knew 98% of all running words in the input text.

Van Zeeland and Schmitt (2012) also studied the association between lexical coverage and listening comprehension among 40 ESL students. Learning materials in that study were four stories in which 0, 2, 5 and 10% of lexical items were replaced with non-words and thus were deemed to be unknown to the research participants. Their text comprehension was measured by a multiple choice task. Lexical coverage was also found to moderate L2 listening comprehension ($X^2(3) = 52.4$, $p < 0.01$) and to have 50% of the comprehension questions answered correctly, the participants needed to reach 95% lexical coverage of all running words in the input texts.

Therefore, there is now a common assumption in the field of incidental L2 vocabulary learning through listening/viewing that students need to have a minimum lexical coverage of 95% and an optimal lexical coverage of 98% to facilitate their text comprehension and thus their incidental vocabulary acquisition through listening/viewing (Nation, 2006; van Zeeland & Schmitt, 2012). This assumption is often referred to elsewhere as the Coverage Comprehension Model (McLean, 2021). However, caution should be taken when this model as well as the two thresholds above are operationalized in both instruction and research practice. On the one hand, scholars in this field all agree that there exists no lexical coverage threshold which can guarantee a reasonable level of text comprehension as well as incidental lexical uptake from this input processing (e.g., Laufer & Ravenhorst-Kalovski, 2010; McLean, 2021; Schmitt et al., 2017; van Zeeland, 2013). It is because the difficulty level of a listening/viewing text can be influenced by, among many others, specialized content, culture-specific reference, unfamiliar accent or simply the speed of speech delivery (Lynch, 2009). On the other hand, it should also be acknowledged that most of previous research in this area often aims to set a plausible lexical coverage level which can foster text comprehension and incidental vocabulary learning through listening/viewing, but ignores the inherent properties of the remaining 2 or 5% vocabulary which is supposed to be unknown to L2 students. If these unknown lexical items are all crucial for interpreting the input content, they can actually create a heavy learning burden for these students (Nguyen, 2020).

Against the backdrop described above, though the thresholds of 95 and 98% lexical coverage cannot ensure sufficient understanding of a given listening/viewing text nor incidental lexical uptake from this listening/viewing process, they, as clearly shown in prior research (e.g., Nation, 2006; van Zeeland, 2013), do not impede learning. In addition, these thresholds do reflect the current state-of-the-art in our field. Therefore, it is still plausible to use these thresholds as the basis to explore the learning affordances that a particular listening/viewing material might have (Nurmukhamedov & Webb, 2019). In fact, several studies have used the two thresholds above to examine the lexical loads of different academic spoken discourse. In the case of TED Talks, for example, Coxhead and Walls (2012) found that L2 students needed to have a receptive knowledge of the most frequent 4,000 and 9,000 word families in English plus that of marginal words and proper nouns to reach the 95 and 98% lexical coverage of these talks, respectively. Such required knowledge reported in the study by Nurmukhamedov (2017) for a larger corpus of TED Talks was, also in that sequence, 4,000 and 8,000 most frequent word families. With regards to lectures and seminars, Dang and Webb (2014) found that a receptive knowledge of the most frequent 3,000 to 5,000 word families plus marginal words and proper nouns provided 95% lexical coverage for these discourse types and that of 5,000 to 13,000 word families could help L2 students to reach 98% lexical coverage. To make a cross-comparison between the findings of the present study and those of the above research, the two thresholds—95% and 98% lexical coverage—were also used in the present study to examine the lexical demands of TED Ed videos.

3.3.2 Word Repetition

Another variable that is also consistently found to influence incidental L2 vocabulary learning from both audio and audio-visual input materials is word repetition. Brown et al. (2008), for instance, examined the effect of word repetition on incidental L2 vocabulary acquisition from three graded readers in three different conditions—reading only, reading-while-listening, and listening only. A set of 84 novel words that occurred between 2 and 3 times, between 7 and 9 times, between 10 and 13 times and between 15 and 20 times in those graded readers were selected as the target words. Vocabulary gain was measured by a word-meaning recall and a word-meaning recognition task. Word repetition was found to affect the size of vocabulary gain on both measures. Brown et al. (2008) suggested that L2 students might need to encounter a novel word at least 20 times in a listening passage for their incidental vocabulary acquisition to be sustained.

Van Zeeland and Schmitt (2013) also investigated the relationship between word repetition and incidental L2 vocabulary learning through listening. In their study, word repetition was again found to moderate the size of vocabulary gain. Unlike Brown et al. (2008), van Zeeland and Schmitt (2013), however, believed that incidental vocabulary learning might already have happened somewhere between the third and the seventh time that L2 students reencountered the target words in their listening passages.

The influence of word repetition on incidental L2 vocabulary learning is also found in the context of academic listening. In the study by Vidal (2003) as already reviewed above, when all other variables were controlled, word repetition alone accounted for 11% of the variance in the vocabulary gain that the L2 research participants in this study made from their viewing procedure. Likewise, in another study by Nguyen (2017), word repetition was also found to moderately correlate with both word-form gain ($r = 0.58$, $p < 0.01$) and the word-meaning gain ($r = 0.48$, $p < 0.01$) after viewing a TED Talk video twice.

The review above clearly shows that word repetition indeed influences incidental L2 vocabulary learning, including academic vocabulary learning through listening and viewing. Two pertinent questions here are whether there exists any particular threshold of word repetition at which incidental L2 vocabulary learning through listening/viewing can be ensured and if yes, what the threshold might be. For example, as already reported above, van Zeeland and Schmitt (2013) suggested that incidental vocabulary uptake might occur somewhere between three and seven times that L2 students reencountered the target word in their listening text or Brown et al. (2008), in contrast, believed that L2 students might need more than 20 times of word encounters to acquire the knowledge related to these lexical items through listening. A threshold of word repetition as such is especially meaningful in the regard that it can help L2 students and teachers in their process of evaluating and/or selecting listening/viewing materials for their L2 listening courses/programs. Unfortunately, such a threshold does not exist due to the inconclusive findings as reported above even though the positive correlation between word repetition and incidental L2 vocabulary gain from listening/viewing input has been consistently substantiated (Webb, 2014). However, what previous research in this area has found can still work as a basis for us to unpack the potential benefits of available listening/viewing materials for L2 students (Nurmukhamedov & Webb, 2019). Therefore, in the present study, I shall give a closer look at the academic word families that occur between 3 and 7 times and from 8 times upwards, as informed by the studies by van Zeeland and Schmitt (2013) and Brown et al. (2008).

4 Present Study

4.1 Research Aim and Research Questions

The primary goal of this study was to unpack the potential benefit of TED Ed videos for incidental L2 academic vocabulary learning. This benefit was operationalized by the frequency that word families in the AWL and the ASWL occurred and reoccurred in these videos. Such an operationalization was based on the assumption that the more often L2 students encountered an academic word family in input materials, the more likely they were able to pick up knowledge related to this lexical item. However, to foster incidental L2 academic vocabulary learning from TED Ed videos, these learning materials should not pose a huge lexical demand to L2 students; otherwise, L2 students might have to struggle with so many new words in the input texts that they fail to comprehend the input content and also fail to pick up new academic vocabulary knowledge from this source. The lexical items that are supposed to be new to L2 students should also reoccur a sufficient number of times within or across the videos to foster their retention of these words. Put differently, this study sought the answers to the following research questions:

1. *How much vocabulary do L2 students need to reach the 95% and the 98% coverage of all word families in TED Ed videos?*
2. *To what extent is academic vocabulary encountered in these videos?*
3. *To what extent is academic vocabulary encountered between 3 and 7 times and from 8 times upwards within and across these videos?*

4.2 Corpus Design

To create equally-sized sub-corpora across all themes, a collection of 30 videos was randomly selected from each theme (see Table 1). Subsequently, a manual check was carried out to ensure that there was not any video in this corpus occurring in more than one theme, because this was actually the case in the present theme-based organization of TED Ed videos in the official website. The final corpus consisted of 360 videos with a total running time of 1,881.36 min or 31.36 h and a total number

Table 1 Total running words and word families

Themes	Running words	Word families	Running time (in minutes)
Arts	40,211	3289	163.6
Business and Economics	42,618	2621	157.4
Technology, Engineering and Design	39,953	2526	153.5
Health	39,272	2539	154.1
Language and Literature	39,117	2898	160.1
Mathematics	40,443	2785	158.6
Philosophy and Religion	40,028	2432	154.5
Psychology	40,187	2501	153.0
Science and Technology	42,693	3130	163.1
Social Studies	40,503	2668	158.2
Teaching and Education	38,373	2470	153.0
Thinking and Learning	39,373	2671	152.3

of 235,553 running words or 7,675 word families. The breakdown of these running words and the word families for each theme is provided in Table 1.

The primary source of data for the present study was the written transcripts of the above videos. Before the data were analyzed, they underwent a process of initial data cleaning. First of all, only audible words were retained and thus their inaudible counterparts such as stage commands, storyline and speakers' names were all removed from the original transcripts. Second, contractions, connected speech, and hyphenated words were changed to conform with spellings used in the BNC word lists. As this study also focused on the general lexical load of TED Ed videos, both marginal words and proper nouns were not excluded from the original transcripts and their total quantity were taken into consideration in answering the first research question.

4.3 Data Analysis

The transcripts were analyzed using the RANGE software (Nation & Heatley, 2002). This program was written to detect the presence of a particular word family together with its frequency in a written text. In the present study, three different analyses using the RANGE program were carried out. One was based on BNC-COCA's 25 frequency-based 1,000-word lists, another on West's General Service List (1953) and Coxhead's (2000) AWL, and the final one on Dang et al. (2017) ASWL. As this study aimed to compare the frequency of academic vocabulary in TED Ed videos against those in other learning materials, word families, but not running words were used as a counting unit to be consistent with the common practice in most of previous research in this area (e.g., Coxhead & Walls, 2012; Dang & Webb, 2014; Nurmukhamedov, 2017). Using word families as a counting unit implied that if L2 students knew a head word, they also knew all other members of this word family. This view was generally deemed to be true especially for the case of receptive vocabulary knowledge that was necessitated for both L2 reading and listening comprehension (Schmitt & Zimmerman, 2002). The outcome of the RANGE analyses also included marginal words (e.g., “oops”, “oh” or “hey”) and proper nouns (e.g., Egypt or Mariana). As these lexical items, especially the proper nouns were found to be a learning burden for of L2 students in the context of listening/viewing (Kobeleva, 2012), their quantity was also counted in the lexical demand of the selected videos. In this study, the RANGE analysis was first carried out for each video and then for a collection of 30 videos in each theme and finally for the entire corpus.

5 Findings

5.1 Required Vocabulary Knowledge for 95% and 98% Lexical Coverage

Table 2 shows the amount of general vocabulary knowledge that L2 students needed in order to reach the 95% and the 98% lexical coverage of each video, each theme and the whole corpus.

Apart from the knowledge of marginal words and proper nouns, to reach the 95% lexical coverage in a TED Ed video, L2 students on average needed a receptive knowledge of the most frequent 4,000 to 6,000 word families, depending on the video that they watched. For the 98% lexical coverage, a receptive knowledge of the most frequent 6,000 to 8,000 word families was required. However, when L2 students watched a collection of 30 videos on the same theme, this required knowledge got reduced to somewhere between 3,000 and 5,000 most frequent word families for the 95% lexical coverage and between 5,000 and 7,000 most frequent word families for the 98% lexical coverage. If they viewed the full collection of 360 videos in the whole corpus above, they only needed a receptive knowledge of 4,000 and 7,000 most frequent word families for the 95% and the 98% lexical coverage, respectively.

Table 2 The 95% and 98% lexical coverage

	95% lexical coverage	98% lexical coverage
For each video	4,000–6,000	6,000–8,000
For each theme	3,000–5,000	5,000–7,000
For entire corpus	4,000	7,000

Table 3 Coverage of academic vocabulary

	Academic spoken vocabulary		Academic written vocabulary	
	Range	Mean	Range	Mean
For each video	72.28–90.06%	82.60%	1.02–11.09%	5.48%
For each theme	77.28–88.94%	83.67%	4.06–7.48%	5.46%
For entire corpus	Total: 83.70%		Total: 5.50%	

5.2 Academic Vocabulary Coverage

Table 3 presents the coverage of academic vocabulary out of all word families in each video, each theme and the entire corpus.

As we can see from the above table, the coverage of academic written vocabulary out of all word families in each video widely varied from a low of 1.02% to a high of 11.09%. On average, however, such a coverage was relatively high at 5.48%. When L2 students viewed a collection of 30 videos on the same theme, the average coverage across all twelve collections also stayed high at 5.46%. For the entire corpus, this coverage became even higher at 5.50% or at roughly 422 word families in the real term. The three themes that provided the largest coverage of academic written vocabulary were Business and Economics, Philosophy and Religion, and Technology, Engineering and Design. Meanwhile, the three themes that offered the smallest amount of academic written vocabulary included Arts, Health, and Language and Literature.

The coverage of academic spoken vocabulary also varied from one video to another with a wide range from 72.28% to 90.06%. On average, this coverage was relatively high at 82.60%. When L2 students viewed 30 videos on the same theme, the average coverage across all twelve themes stood at 83.67%. For the entire corpus, this coverage was slightly higher at 83.70% or nearly 1,424 word families in the real term. The three themes that offered the highest coverage of academic spoken words were Language and Literature, Thinking and Learning, and Teaching and Education. Science and Technology, Technology, Engineering and Design, and Arts, however, were found to provide the lowest coverage of academic spoken vocabulary.

Taken altogether, TED Ed videos indeed offered L2 students a rich opportunity to encounter academic vocabulary—both spoken and written—in a single video, in a collection of 30 videos on the same theme, and especially in the entire collection of 360 videos.

5.3 Frequency of Academic Vocabulary

To give a closer look at the frequency that academic vocabulary occurred in a single video, a theme was randomly picked up from the pool of the twelve themes above. Table 4 presents the quantity of academic spoken and written word families that occurred between 1 and 2 times, between 3 and 7 times and from 8 times upwards within a single video.

With regards to academic written vocabulary, the chance for L2 students to encounter these word families from 3 to 7 times in a video widely varied from 0 to 33.09% of all academic written word families occurring in that video, but with a fairly high mean of 24.26%. The chance for L2 students to encounter these word families 8 times or over in a video was far slimmer with a range from 0% to 8.42% and with an average of 2.95%. As far as academic spoken vocabulary is concerned, L2 students might encounter from 4.53 to 20.30% of all academic spoken word families in a single video from 3 to 7 times with an average of 12.47%. They might also encounter from 4.82 to 7.96% of these word families at least 8 times in a single video with a mean of 6.26%.

The above relatively high chance for L2 students to reencounter an academic word family—both spoken and written—in a single video might become more pronounced if they watched 30 videos on the same theme. To examine the plausibility of this view, the frequency of these word families in all twelve collections of 30 videos for all themes was also scrutinized. Table 5 summarizes the outcome of such an analysis.

As shown in Table 5, the chance that L2 students would encounter an academic written word family between 3 and 7 times indeed increased to a great extent when they viewed 30 videos on the same theme. On average, such an encounter was found for 34.44% of all academic written word families in each collection, which was equal to 50 word families. This figure was also high at 29.90% for the word families that L2 students encountered 8 times and more in a collection, which was equivalent to 44 word families. When it comes to the case of academic spoken vocabulary, L2 students on average could encounter 27.39% of these word families in a 30-video collection from 3 to 7 times and 49.90% from 8 times upwards, which were equal to 623 and 1,128 word families, respectively. All in all, L2 students are offered a high chance of reencountering both academic spoken and written vocabulary in each video and specially in each theme.

6 Discussion and Pedagogical Implications

In answer to the first research question, the general vocabulary knowledge that L2 students needed to reach the 95% and the 98% lexical coverage in a single TED Ed video was between 4,000 and 6,000 and between 6,000 and 8,000 word families, respectively. However, these figures noticeably reduced to between 3,000 and 5,000 word families for the 95% lexical coverage and between 5,000 and 7,000 word families for the 98% lexical coverage when they viewed 30 videos on the same theme. If they viewed the whole collection of 360 videos, the above figures even shrunk to 4,000 and 7,000 word families. Table 6 compares the size of general vocabulary knowledge that L2 students might need for the 95% and the 98% lexical coverage across different academic spoken discourse.

Table 4 Frequency of academic vocabulary occurrences in a video

	1–2		3–7		≥8	
	AWV	ASV	AWV	ASV	AWV	ASV
V1	100.00	83.60	**0.00**	10.58	**0.00**	5.82
V2	96.55	89.96	3.45	5.02	**0.00**	5.02
V3	91.70	71.55	8.30	11.88	**0.00**	6.63
V4	88.17	77.02	11.83	20.29	**0.00**	6.00
V5	72.67	89.91	24.22	5.26	3.11	**4.82**
V6	72.22	80.59	24.69	12.13	3.09	7.28
V7	69.36	77.05	27.75	16.57	2.89	6.39
V8	67.90	74.08	29.40	19.92	2.70	6.00
V9	77.24	84.58	22.76	7.46	0.00	**7.96**
V10	71.79	89.71	25.00	**4.53**	3.21	5.76
V11	67.64	81.21	29.47	12.72	2.89	6.07
V12	62.86	89.96	33.06	5.02	4.08	5.02
V13	63.50	81.49	32.14	11.88	4.36	6.63
V14	63.42	79.30	31.91	14.51	4.67	6.19
V15	62.79	72.15	31.95	20.19	5.26	7.66
V16	61.70	84.24	**33.09**	7.88	5.21	7.88
V17	60.93	89.71	32.87	**4.53**	6.20	5.76
V18	88.47	73.96	11.43	19.44	0.10	6.60
V19	86.92	84.65	13.00	7.43	0.08	7.92
V20	83.92	89.71	16.00	4.53	0.08	5.76
V21	79.84	81.21	18.61	12.72	1.55	6.07
V22	60.27	81.21	32.06	12.72	7.67	6.07
V23	60.60	78.90	30.98	15.19	**8.42**	5.91
V24	72.57	73.43	24.32	**20.30**	3.11	6.26
V25	67.02	78.90	30.32	15.19	2.66	5.91
V26	67.63	73.43	29.47	**20.30**	2.90	6.26
V27	72.50	83.98	24.38	9.94	3.12	6.08
V28	66.35	78.90	29.91	15.19	3.74	5.91
V29	64.10	73.43	32.40	**20.30**	3.50	6.26
V30	63.29	83.60	32.91	10.58	3.80	5.82
M	72.80	81.05	24.26	12.47	2.95	6.26

AWV/ASV = Academic written/spoken vocabulary
V = Video; M = Mean; Figures in percentage
Bolded figures = the lowest and the highest means

Table 5 Frequency of academic vocabulary occurrences in a theme

Theme	Number of occurrences	Academic written vocabulary	Academic spoken vocabulary
Arts	**1–2**	36.31	39.29
	3–7	38.61	34.67
	≥ 8	25.08	26.04
Business and Economics	**1–2**	30.77	11.32
	3–7	38.46	30.00
	≥8	30.77	58.68
Technology, Engineering and Design	**1–2**	31.66	20.89
	3–7	31.67	25.73
	≥8	36.67	53.38
Health	**1–2**	38.19	27.98
	3–7	38.19	32.14
	≥8	23.62	39.88
Language and Literature	**1–2**	39.49	22.54
	3–7	31.51	25.31
	≥8	29.00	52.15
Mathematics	**1–2**	36.90	21.50
	3–7	38.49	29.63
	≥8	24.61	48.87
Philosophy and Religion	**1–2**	32.36	21.90
	3–7	30.10	24.05
	≥8	37.54	54.05
Psychology	**1–2**	35.93	21.32
	3–7	32.59	22.60
	≥8	31.48	56.08
Science and Technology	**1–2**	37.16	20.06
	3–7	36.25	27.46
	≥8	26.59	52.48
Social Studies	**1–2**	32.53	21.48
	3–7	28.77	24.50
	≥8	38.70	54.02
Teaching and Education	**1–2**	40.55	22.20
	3–7	34.06	27.29
	≥8	25.39	50.51

(continued)

Table 5 (continued)

Theme	Number of occurrences	Academic written vocabulary	Academic spoken vocabulary
Thinking and Learning	**1–2**	36.11	22.00
	3–7	34.58	25.29
	≥8	29.31	52.71
M	**1–2**	***35.66***	***22.71***
	3–7	***34.44***	***27.39***
	≥8	***29.90***	***49.90***

AWV/ASV = Academic written/spoken vocabulary
M = Mean; Figures in percentage

Table 6 Lexical coverage of academic spoken discourse

	95%	98%
TED Talks (Coxhead & Walls, 2012)	4,000	9,000
TED Talks (Nurmukhamedov, 2017)	4,000	8,000
Lectures and seminars (Dang & Webb, 2014)	3,000–5,000	5,000–13,000
TED Ed (present study)	4,000	7,000

As we can see from Table 6, TED Ed videos in fact have a roughly equal or even a smaller lexical demand for L2 students to reach the 95% and 98% lexical coverage, compared to other academic spoken discourse such as lectures, seminars or even its precedent TED Talks. When L2 students do not have to struggle with too many new words in an input text, they are expected to have more mental resources to interpret both the input content and the meanings of novel words occurring in that text (van Zeeland, 2013). Thus, the chance for their incidental academic vocabulary learning from viewing/listening to TED Ed videos appears to be higher on this premise.

As for the second research question, TED Ed videos also provide a rich opportunity for L2 students to encounter academic vocabulary—both spoken and written. Regarding the spoken academic vocabulary, on average, this vocabulary accounted for 77.28 to 88.94% of all word families in each video. When L2 students viewed 30 videos on the same theme, these figures were between 79.30 and 89.71%. If they watched all 360 videos in the corpus, 83.70% of all word families they encountered or nearly 1,424 word families were found to be academic spoken vocabulary. Although this coverage was relatively high, it was still lower than what previous research found in lectures, seminars, labs and tutorials. In the study by Dang et al. (2017), for example, academic spoken vocabulary was found to cover 90.13% and 89.59% of all word families in their two academic spoken corpora. One plausible explanation for this difference is that most of the speeches in Dang et al. (2017) corpora were recorded live from natural academic speaking events and therefore academic spoken vocabulary figured higher in their lexical profile. Meanwhile, a majority of TED Ed speeches were pre-scripted, read aloud and then audio-recorded. It was therefore not very surprising that the amount of academic spoken vocabulary was much lower in the case of TED Ed videos.

With regards to the written academic vocabulary, the present research found that on average, from 4.06 to 7.48% of all word families in a TED Ed video were academic written vocabulary. When L2 students watched 30 videos on the same theme, this coverage was also relatively high, ranging from 4.40 to 6.50%. This coverage for the entire corpus of 300 selected videos stood at 5.50% or about 422 word families in the real term. Obviously, TED Ed videos offered a much higher coverage of academic written vocabulary than its precedent—TED Talks videos. Both Coxhead and Walls (2012) and Nurmukhamedov (2017) reported such a coverage in their corpora of TED Talks videos at merely 3.90 and 3.79%, respectively. The coverage of academic written vocabulary in TED ED videos even outnumbered that of lectures and seminars by a noticeable gap of 1.09% (Dang & Webb, 2014) and that of lectures, seminars, labs and tutorials with a discrepancy of 1.33% (Dang et al., 2017). This therefore suggests that TED Ed videos offer L2 students a much higher chance of encountering academic written vocabulary than other academic spoken discourse such as TED Talks, lectures, seminars, labs and tutorials. This also entails a higher chance for incidental academic vocabulary learning to occur from this source.

As far as the final research question is concerned, L2 students, on average, had the opportunity to encounter 12.27% of all academic spoken word families in a TED Ed video from 3 to 7 times and 6.26% from 8 times upwards in a single video. These figures both surged to 27.39% (i.e., 50 word families) and 49.90% (i.e., 44 word families) respectively when they watched 30 videos on the same theme. This also holds true for the case of academic written vocabulary. In a single video, 24.26% of all academic written word families reoccurred 3 to 7 times and 2.95% reoccurred at least 8 times. In a collection of 30 videos on the same theme, the former climbed to 34.44% (i.e., 623 word families) and the latter even soared by over tenfold to 29.90% (i.e., 1128 word families). As already mentioned, there is no certain threshold of word encounters that guarantees incidental vocabulary learning through (academic) listening/viewing to occur (Webb, 2014). However, researchers in this area generally believe that the more often a word is encountered, the more likely it is acquired (Brown et al., 2008; van Zeeland & Schmitt, 2012; Vidal, 2003; Webb, 2014). In this regard, TED Ed videos are indeed a rich source for academic vocabulary learning as they offer L2 students multiple encounters to the same word family both in a single video and especially in a collection of 30 videos on the same theme. For example, the academic spoken word family "*equal*" occurs 19 times in a single video about Mathematics or the academic written word family "*create*" occurs as many as 25 times in another video about Arts. Such a high frequency of word reencounters is expected to create a favorable condition for incidental academic vocabulary learning through meaning-focused listening/viewing to occur. The chance for this incidental learning is even higher as the speakers in these videos often give clear L2 definitions and explanations to novel and key words which, in turn, helps foster L2 students' lexical form-meaning mapping.

Taken altogether, as shown in this study, TED Ed videos, on the one hand, do not have a huge lexical demand to L2 students. On the other hand, they offer L2 students a high chance of encountering and reencountering academic vocabulary both in a single video and in a collection of 30 videos on the same theme. These factors together create a favorable condition for incidental L2 academic vocabulary learning to occur, especially when L2 students carry out the practice of extensive or narrow listening/viewing of a large number of TED Ed videos in their self-study. If L2 students aim to further develop their academic spoken vocabulary, some themes such as Language and Literature, Thinking and Learning, and Teaching and Education may top their list of choices. However, if they place more emphasis on academic written vocabulary in their courses, they might opt for such themes as Business and Economics, Philosophy and Religion, and Technology, Engineering and Design.

This study also found that a single video already provides L2 students with multiple encounters to the same academic word family. However, such a high frequency of word reencounters even becomes more pronounced when they view 30 videos on the same theme. Therefore, narrow viewing (i.e., viewing videos on the same theme) is highly recommended when they decide to use TED Ed as their listening/viewing materials. In a narrow viewing procedure, the video in which the meanings of unfamiliar academic words are the most easily inferred from the surrounding context should be placed first. When students have already succeeded in interpreting the meanings of these words, this will help them to reduce the number of unfamiliar academic words in the following videos. In other words, they can increase their lexical coverage when they move on from one video to another.

7 Conclusion

This study started with a question as to whether TED Ed videos were a potential source for incidental L2 academic vocabulary learning in the wild. The answer to this question was yes. In fact, these videos provide various benefits for incidental L2 academic vocabulary learning. First of all, these learning materials do not require a large size of general vocabulary knowledge on the part of L2 students to sustain their text comprehension. When L2 students do not have to struggle to comprehend an input text, they stand in a better position to process and hopefully pick up new academic vocabulary knowledge from that text. Second, as shown in this study, TED Ed talks offer L2 students a high chance of encountering academic vocabulary, especially in the case of academic written vocabulary, compared to other academic spoken discourse such as lectures, seminars or even its precedents TED Talks. Last but not the least, the chance for L2 students to pick up academic vocabulary knowledge from viewing/listening to TED Ed videos can be even more robust that their counterparts. It is because there is a relatively large amount of academic vocabulary that occurs in these videos more than eight times, especially when L2 students view 30 videos on the same theme. Thus, if L2 students aim to foster academic vocabulary learning in their listening/viewing in the wild, TED Ed videos are clearly a good option.

References

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253–279.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In B. Laufer & P. Bogaards (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). John Benjamins.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238.

Coxhead, A. J., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *The TESOLANDZ Journal, 20*, 55–58.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Dang, T. N. Y., Coxhead, A. J., & Webb, S. (2017). The academic spoken word list. *Language Learning, 67*(4), 959–997.

Ellis, R. (1999). *Learning a second language through interaction.* John Benjamins Publishing Co.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*, 305–327. https://doi.org/10.1093/applin/amt015

Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum, 17*, 24–27.

Juilland, A. G., & Chang-Rodr´ıguez, E. (1964). *Frequency dictionary of Spanish words.* Mouton.

Kobeleva, P. (2012). Second language listening and unfamiliar proper nouns: Comprehension barrier? *RELC Journal, 43*(1), 83–98.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Lawrence, J. F., Hagen, A. M., Hwang, J. K., Lin, G., & Lervag, A. (2019). Academic vocabulary and reading comprehension: Exploring the relationship across measures of vocabulary. *Reading and Writing, 32*, 285–306.

Lynch, T. (2009). *Teaching second language listening*. Oxford University Press.

Lynn, R. W. (1973). Preparing word lists: A suggested method. *RELC Journal, 4*(1), 25–32.

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to students' academic achievement. *Journal of English for Academic Purposes, 31*, 44–57.

McLean, S. (2021). The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized. *Reading in a Foreign Language, 33*(1), 126–140.

Nation, P. (2006). How large a vocabulary is needed to reading and listening? *The Canadian Modern Language Review, 63*, 59–82.

Nation, P., & Heatley, A. (2002). *Range: A program for the analysis of vocabulary in texts*. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.

Nguyen, C. D. (2020). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal* (Online view first). https://doi.org/10.1177/0033688219895045

Nguyen, C. D. (2017). *Fostering incidental vocabulary uptake from audio-visual materials: The role of text comprehension* (PhD Thesis). Victoria University of Wellington, New Zealand.

Nguyen, C. D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly, 53*(1), 5–29.

Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal, 8*(4), 768–790.

Nurmukhamedov, U., & Webb, S. (2019). Research timeline: Lexical coverage and profiling. *Language Teaching, 52*, 188–200.

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 09–29.

Read, J. (1993). The development of a measure of vocabulary knowledge. *Language Testing, 10*(3), 355–371.

Sawada, K. (2009). *Vocabulary acquisition through listening and its relation to learning channel preferences* (Unpublished doctoral dissertation). Temple University, USA.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching, 50*(2), 212–226.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do students know? *TESOL Quarterly, 36*, 145–171. https://doi.org/10.2307/3588328

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*, 487–512. https://doi.org/10.1093/applin/amp058

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661–699.

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English students: An intervention study. *Reading and Writing, 22*, 993–1019. https://doi.org/10.1007/s11145-008-9141-y

van Zeeland, H. (2013). Vocabulary and listening. In C. A. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 1–6). Wiley-Blackwell.

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*, 457–479.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3–25.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56–89.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*. https://doi.org/10.1177/1362168814559800

Webb, S., & Rodgers, M. (2009). The lexical coverage of movies. *Applied Linguistics, 30*, 407–427.

Webb, S. (2014). Repetition in incidental vocabulary learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). John Wiley and Sons.

West, M. (1953). *A general service list of English words*. Longman, Green.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215–229.

Chi-Duc Nguyen (BA Honor in TESOL, Vietnam National University; MA in TESOL and PhD in Applied Linguistics, Victoria University of Wellington, New Zealand) has been an EFL/ESL teacher and an EFL/ESL teacher educator for over 15 years. His research interests include teaching and learning vocabulary in another language, language testing and assessment, and teacher identity. His latest publications can be found in *TESOL Quarterly, Language Teaching Research* and *RELC Journal*.

Vocabulary Learning from Subtitled Input After Minimal Exposure

Imma Miralpeix, Ferran Gesa, and Maria-del-Mar Suárez

Abstract Despite previous research on adult L2 acquisition after minimal exposure (Rast, 2010a) and on effects of watching subtitled videos for language learning (Bisson et al., 2014; Vanderplank, 2016), little is known about how adults extract information from multimodal input when first exposed to a new language (Gullberg et al., 2012). The present study explores whether intentional vocabulary learning from a subtitled video in an unknown language takes place after minimal exposure and whether aptitude plays a role in this endeavour. For this purpose, 46 Catalan/Spanish bilingual learners of English took the LLAMA B aptitude test (Meara, 2005) and were asked to watch an advert in English but with Polish subtitles, a language they had no knowledge of. After the second viewing, they completed a timed meaning recognition test consisting of 15 Polish words and expressions from the advert. Results from the vocabulary test showed that learning occurred and participants were making word form-meaning connections. It was also observed that high aptitude was related to more learning. The study demonstrates how digital technology and multimodal input can help vocabulary acquisition during first exposure to a novel target language and how individual differences can influence word learning at these very first stages.

Keywords Aptitude · Captioning · First exposure · Minimal input · Vocabulary learning

I. Miralpeix (✉)
English Studies Department, Universitat de Barcelona, Barcelona, Spain
e-mail: imiralpeix@ub.edu

F. Gesa · M.-M. Suárez
Department of Linguistic and Literary Education, Universitat de Barcelona, Barcelona, Spain
e-mail: ferran.gesa@ub.edu

M.-M. Suárez
e-mail: mmsuarez@ub.edu

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_10

1 Introduction

Studies on the age factor have shown that even if young learners can achieve greater second language (L2) proficiency in the long run given the adequate contextual conditions, older learners tend to exhibit quicker development rates at the first stages of learning (see, for example, Muñoz & Singleton, 2019). Although adult L2 acquisition is thought to be slow and laborious when compared to first language (L1) acquisition, it has also been shown that minimal L2 instruction in this population can produce rapid change (McLaughlin et al., 2004). Gullberg et al., (2012, p. 259) have also suggested that "the adult learning mechanism appears to be a great deal more powerful than typically assumed in the L2 acquisition literature".

Very few studies have concentrated so far on the earliest stages of adult L2 acquisition although it has been claimed that this work should be more central to L2 research (e.g., Han & Rast, 2014; VanPatten, 2014). Any second language acquisition (SLA) theory should give account of all phases in the process so as not to be incomplete, including also very initial stages in adult learners (Rast, 2008). We refer to the studies in this line of research as *first exposure* (FE) studies, where "data are collected from the very first moment of contact with the target language (TL) and within the first seconds, minutes and hours of subsequent exposure, and which all TL input is controlled" (Rast, 2008, p. 29). Therefore, we talk in this case of *ab initio* learners (i.e., without any previous knowledge of the TL) and this is different from *en route* or *al fine* learners (i.e., learners who are already familiar with the TL).

An overview of the literature from the last 15 years shows that not much research has been devoted to exploring the first stages of adult L2 acquisition. We find one special issue edited by Gullberg and Indefrey in *Language Learning* (2010), one by Carroll in *Second Language Research* (2013) and two monographs by Rast (2008) and Han and Rast (2014). Gullberg and Indefrey (2010) is a comprehensive volume on FE research offering an innovative collection of papers written by researchers from different fields and approaches to the study of language (psychology, neurology, SLA, etc.). Han and Rast (2014) presents a state of the art on input processing at FE drawing from a variety of theoretical frameworks, mainly from SLA researchers. Carroll (2013) specifically focuses on FE and word learning and Rast (2008) on initial processing in an instructional setting. There are also several interesting one-off studies we will be referring to, although much remains to be studied. In addition to being relevant for theoretical purposes, as we have seen above, FE studies can also provide information on how adult learners can make the most from the input when initially exposed to a new language. Nation (1996), for example, already commented on the importance of having a *Little Language* as soon as possible, i.e., a system that, even if formed by a small number of words, can be sufficient to the learners to make themselves understood and allow them to perform basic tasks.

2 Literature Review

2.1 First Exposure Studies in Second Language Acquisition

As Rast (2008) identified, there are three main areas of investigation in FE studies and they have to do with (1) the pre-existing linguistic knowledge brought by the individual to the acquisition task, (2) the individual differences of the language learners and (3) the influence of the linguistic environment (TL input) on processing and acquisition. Although some studies concentrate on just one of these areas, most tend to combine or put in relation at least two of them.

2.1.1 Pre-existing Linguistic Knowledge at the Outset of L2 Learning

First of all, researchers have been interested in identifying the starting point of adult L2 acquisition (e.g., Carroll, 2005; Rast, 2010a; Sagarra, 2014). Therefore, they have also enquired about any possible source of pre-existing knowledge, that is, any previously acquired languages, such as the L1 or any L2s, or any kind of implicit or explicit knowledge that individuals may bring to the acquisition of the new language. This can help determine which factors may support or hinder learning at the outset of the process. In relation to possible L1 effects, Carroll (2005) found that Anglophone adults were sensitive to phonological, morphological and semantic cues when provided with French auditory stimuli, and that prior linguistic knowledge affected the processing of the new language (e.g., pattern detection when listening to speech was biased to what they already knew from the L1). However, lack of recourse to the L1 has also been found in beginner learners when it comes to morphology. An eye-tracking study by Sagarra (2014) asked whether previous experience with either morphologically rich or poor languages modulated the learning of a morphologically rich L2. Results showed that it was not the case, suggesting that beginning learners may be immune to morphological transfer effects, i.e., that transfer requires a certain level of L2 knowledge. Regarding the effects of other previously learned languages, there is evidence so far that even minimal knowledge of different L2s can be the source of cross-linguistic influence for *ab initio* learners, without necessarily having to recur to the L1. For instance, Rast (2010a) observed an influence of other L2s in a translation task of words in the new language (phonetic and orthographic similarity between new items and those learned in previous languages helped identifying words' meaning). It is also important to consider, though, that what is understood by *outset of L2 learning* can vary in different studies to different degrees. For instance, Rast's (2010a) participants were only exposed to the new language through the task they performed for the study. In Carroll's (2005), 4.5% of the sample claimed to have some knowledge of French (even if just some words). Sagarra's (2014) participants were enrolled in a basic course of the new language, which led researchers to question whether it was an FE study (VanPatten, 2014).

2.1.2 Learners' Individual Differences: The Role of Aptitude

The second area under investigation has to do with learners' individual characteristics. These differences do not tend to be the focus of research in FE studies, but they are taken as variables that are controlled in order not to interfere with results. For instance, in Sagarra (2014) groups of participants tested on morphology are comparable in terms of working memory and inhibitory control. In this way, a possible influence of these learners' characteristics on the findings can be disregarded. Rarely have individual characteristics been studied in conjunction with other factors in FE studies although we may find an exception in Rast (2010b), who studied how psychotypology (i.e., the perceived distance between different languages by each individual) affected acquisition at initial stages.

Other individual variables that seem to be related with language learning at low proficiency levels have not been analysed in FE studies. One of them is language learning aptitude, which can be defined as a specific talent (or set of abilities) that predicts capacity, readiness, rate and speed of language learning (Wen et al., 2017). For example, Dahlen and Caldwell-Harris (2013) focused on word recall and recognition in Turkish early learning and saw that those participants with high foreign language aptitude, as measured by the MLAT (Carroll & Sapon, 1959), recalled and recognised more target words (TW) than participants with low foreign language aptitude. Aptitude appears to be highly predictive of rate of progress at early language learning stages (Doughty, 2019), but its influence tends to decrease as L2 proficiency improves and so do other cognitive skills and strategies related to language learning (Serafini & Sanz, 2016; Winke, 2013). Despite aptitude being considered an influential variable when we start learning an L2, and taking into account the results found in previous literature on L2 learning, it is surprising that FE studies (e.g., Gullberg & Indefrey, 2010; Han & Rast, 2014) have often overlooked its possible effects, although individual differences have been taken into account in studies such as Rast et al. (2014).

It has also been acknowledged that aptitude is multicomponential, i.e., made of different subcomponents, such as inductive learning ability–to infer rules from the input– or rote learning ability–to remember large amounts of foreign language material (Carroll, 1993). This might also explain the ease in acquiring different aspects of language knowledge and skills. Therefore, total aptitude scores may not necessarily be strongly associated with a specific skill (Li, 2016). It is for this reason that, when conducting research, an adequate test for the subcomponents we are interested in assessing should be used. This is of relevance for the present study because the possible relationship between aptitude to learn new words and learning outcomes after minimal exposure is being evaluated. As we are assessing *ab initio* learners deliberately trying to learn from minimal input, LLAMA B (Meara, 2005) can be an appropriate test. It has been shown to measure the subcomponent of aptitude we are interested in (rote learning) as well as explicit learning aptitude (Granena, 2013). This subtest has also shown to have good internal validity for research purposes (Bokander & Bylund, 2020).

2.1.3 Target Language Input

Artificial Language Input

TL input is crucial if we are to understand "the natural processes and mechanisms underlying a learner's capacity to break into the 'wild'" (Han & Rast, 2014, p. 1). In FE studies, input tends to be controlled up to different extents, as we need to examine the impact of input properties on students' learning. Therefore, we can find different degrees of manipulation of the TL input. Several studies have been conducted with entirely artificial languages to uncover what learners can do with a system that is completely new to them, irrespective of any previous language(s) they may have acquired. For example, de Diego Balaguer (2007) used ERPs to analyse how words and structural dependencies are extracted from speech, Cunillera et al. (2006) used the same methodology to assess how stress and statistical cues interact to help word segmentation and Folia et al. (2010) explored artificial language learning in adults and children and suggested that the mechanisms involved in this process are shared with those of natural language acquisition. This kind of study has shown to be useful for testing models and hypotheses in early L2 acquisition. For instance, Yang and Givón (1997) examined Givón's Competition Model in which vocabulary and grammar are supposed to compete for memory, attention and processing capacity when starting to learn a new language.

Natural Language Input

Even if studies with artificial input provide valuable information for a better understanding of language processing, research using real languages is necessary not just to confirm or disprove findings but also to enhance ecological validity. Among the studies conducted with natural language as TL input, a special place must be reserved for those in the pioneer VILLA Project (Varieties of Initial Learners in Language Acquisition; Dimroth et al., 2013) as this longitudinal research programme was very comprehensive in nature. In the project, controlled classroom input was provided to French *ab initio* learners of Polish, assessing the linguistic achievements of these learners at three different testing times (at the very beginning, after 7.5 hours of exposure and after 13.5) and putting learners' performance in relation to the input they had received. This research project focused on speech perception and comprehension as well as on morphology, grammar and vocabulary learning (Rast, 2008, 2010a, 2010b; Shoemaker & Rast, 2013).

Multimodal Input

Apart from those conducted with input from real language classes, other studies have used videos to expose *ab initio* learners to a totally unknown language. Dutch participants in Gullberg et al., (2010, 2012) were asked to watch twice a seven-minute weather forecast in Chinese and were later assessed on word and meaning recognition. The video was specially prepared for the experiments, with tightly controlled linguistic features to relate input's characteristics with learners' performance. Findings showed that these untutored learners made use of frequency and syllable structure cues as well as contextual cues such as gestures to effectively deal with very little and complex input. In the absence of a conscious learning effort, the authors emphasise "the effective adaptation of the adult learning mechanism to new input and to remarkably rapid learning" (Gullberg et al., 2012, p. 257). More recently, another study by Bisson et al. (2014) also used audiovisual input (but this time with subtitles) to investigate incidental acquisition of vocabulary. English speakers with no knowledge of Dutch watched 25-minute episodes of *SpongeBob* under three subtitling conditions: intralingual subtitling (Dutch audio with Dutch subtitles), standard subtitling (Dutch audio with English subtitles) and reversed subtitling (English audio with Dutch subtitles). There was also a control condition with Dutch audio without subtitles. Eye-tracking helped to reveal that participants read the subtitles irrespective of the subtitling condition (and more regularly when the soundtrack was also in the new language). However, no evidence of vocabulary acquisition was found when the results of an unannounced auditory vocabulary test were analysed. To our knowledge, no other FE studies have been conducted with multimodal input and *ab initio* learners, as research available with adult learners has been carried out with *en route* learners. For example, Gesa (2019) identifies several studies with beginner university learners who were exposed to videos in subtitled and non-subtitled conditions and results differ depending on the study. For instance, in d'Ydewalle and Pavakanum (1995) and Sydorenko (2010), subtitling enhanced vocabulary learning although other studies such as Raine (2012) found no significant differences in beginners' word learning between watching foreign language videos with or without subtitles.

2.2 Vocabulary Learning in Minimal Input Studies

In relation to what novice learners can extract from the input they are exposed to in unknown languages–natural or not–, research has been conducted on different linguistic abilities, from word extraction (e.g., Richtsmeier, 2011) to the learning of grammar constructions (Carroll & Widjaja, 2013) and grammar rules (e.g., De Diego et al., 2007), with a very interesting debate focusing on the acquisition of vocabulary and morphosyntax. According to the *Primacy of Meaning Principle* (VanPatten, 1996, 2007, 2014), FE learners would prioritise meaning over form in input processing, while for Carroll (2001), meaning-based processing of words would just be possible when learners know word-forms well enough and will therefore take place later. Achieving a sufficient knowledge of word forms is not easy and can be slow. According to Han and Liu (2013), when learners are presented with natural auditory input, they are able to process word forms only minimally.

Research on L2 vocabulary acquisition has often tended to neglect the very early stages of language learning, although several studies (e.g., Carroll, 2014; Gullberg et al., 2010, 2012; Rast, 2010a; Shoemaker & Rast, 2013) have focused on word extraction and word learning in minimal input conditions. Gullberg et al. (2012) found out that learners were able to extract both form and meaning information about words from minimal contact with a new language, and this would be reminiscent of the fast mapping process that occurs in children's L1 vocabulary acquisition. In this case, with the new language being Chinese, both frequency and syllable structure of words affected performance, as those disyllabic items that appeared more often in the video were also more often recognised in a post-test. Gestural cues were found to help acquisition, too. Results from this study suggested that a set of cues need to accumulate to influence learning at these low levels of exposure. Shoemaker and Rast (2013) and Rast (2010a) saw that words that are placed at utterance edges or were similar to other L1 words were more easily extracted from input in a completely new language. However, other factors such as frequency would not be determinant at the very beginning stages–at least until more exposure is provided– and word length would have no effect on performance. Carroll (2014) also found out that cognate stimuli (i.e., similar words in the L1 and the TL) were easy to identify and learn from oral input.

In relation to vocabulary learning and learners' characteristics, as far as we know, only a few studies have examined the relationship between L2 vocabulary acquisition and aptitude and none is an FE study (as mentioned in Sect. 2.1.2). There is only one study focusing on aptitude and word learning from multimodal input at different proficiency levels (Suárez & Gesa, 2019), and it showed that aptitude was more relevant when proficiency levels were low in tasks that were cognitively demanding, such as making form-meaning connections when watching subtitled videos. However, this study deals with extensive viewing in *en route* learners and no research is available on *ab initio* learners with minimal exposure to multimodal input.

3 Aims and Research Questions

In the light of what has been discussed above, the present study aims at complementing the scarce research so far on adult learners' vocabulary learning after minimal exposure to an unknown language. Given the increase of available multimodal input in society nowadays, we think it is pertinent to explore how adults can intentionally make the most of the opportunities that this input may offer, also when the TL is completely unknown to them. In addition, multimodal input would conform to the guidelines that VanPatten (2014) proposes for further research, as this input (1) offers stimuli which are sentential (subtitles) rather than textual and (2) is accompanied by context (visual or others) to aid in interpretation of the stimuli. In this case, there will be image, text and sound. The study will also include a simultaneous focus on meaning and form, as recommended by VanPatten (2014). Further, to contribute to the debate on the role of learner variables in learning from minimal exposure, we will explore whether language learning aptitude may assist novice learners in acquiring new words from minimal multimodal input. The two research questions that we want to answer in this study are the following:

1. Are adult Catalan/Spanish native speakers with no previous knowledge of Polish able to make word form-meaning connections in the new language after minimal exposure to subtitled multimodal input?
2. Does aptitude contribute to vocabulary learning at very initial stages in these conditions?

4 Method

4.1 Participants

50 Catalan/Spanish prospective teachers pursuing studies in early childhood education at the University of Barcelona participated in the study. Four of the participants had to be excluded as the data they provided were not complete (i.e., they did not take at least one of the tests below). Therefore, our final sample consisted of 46 participants (42 females and 4 males). These 46 participants belonged to three groups of the same compulsory course (G1, $n = 13$; G2, $n = 18$; and G3, $n = 15$). Their mean age at testing was 20 ($sd = 0.667$) and they had studied English at primary and secondary school for about nine years. Furthermore, they had no knowledge of Polish and had never been (in)formally exposed to it. This was confirmed by a questionnaire that participants filled in after taking the tests and by direct questions from the researchers to the participants in case of any doubt.

4.2 Instruments

4.2.1 TV Advert

A TV advert of a well-known online retailer was used as input for the study. The advert was 2:58 minutes long and presented the story of a Polish resourceful grandfather preparing a trip to the UK and learning English at home, a language that, as we discover at the end, will be useful to communicate with his very young granddaughter in Britain. The audio of the advert was in English and the subtitles were presented in Polish. The text of the subtitles consisted of 75 tokens and 45 types (i.e., different words). There were a total of 34 subtitles, from one to six words long, with a mean of 2.2 words per subtitle. Regarding the audio, approximately 95% of the words belonged to 1K and 5% to 2K according to *VocabProfile* (Cobb, n.d.), well within the level of competence of the learners, as their receptive vocabulary size was larger than 2,000 words.

(1) chleb:	(a) breakfast	(b) bread	(c) fridge	(d) I don't know
(2) dziękuję ci:	(a) goodbye	(b) you are perfect	(c) thank you	(d) I don't know

Fig. 1 Examples of test items

4.2.2 Polish Vocabulary Test

The online vocabulary test consisted of 15 TWs and expressions appearing in the advert. We mainly focused on content words (nouns), instead of grammatical words, and very short phrases (e.g., *ono to jest*–it is–) as well as idiomatic expressions (e.g., *kocham cię*–I love you–). We avoided including cognates because they have already been shown to have a facilitating effect when learning from minimal input (e.g., Shoemaker & Rast, 2013) and their meanings could have been guessed just from the test. Therefore, we did not include transparent words in Spanish or English as TWs (e.g., *passport*–passport– or *piżama*–pyjamas–). All the TWs appeared once except for *kocham cię* and *ono to jest* (appearing twice), *cześć*–hi– (5 times) and *ja jestem*–I am– (6 times).

This multiple-choice test showed one word at a time and once participants provided an answer to that word, they could not go back and were asked to proceed with the next item. Each TW was presented with three possible meanings plus the option *I don't know* (see examples 1 and 2 in Fig. 1 and Appendix 1). The two distractors were words that had also appeared and/or were mentioned in the advert. When the target item was an expression, some of the options were items formed by more than one word as well (see example 2 in Fig. 1). The order of appearance of the TWs was randomised to avoid any recency effects. Time limit for the whole test was three minutes and a maximum of 15 points could be obtained in this test, one for each correct answer.

4.2.3 LLAMA B Test

LLAMA B (Meara, 2005) is part of a battery of tests aimed at measuring language learning aptitude and it assesses how good participants are at learning new vocabulary. It is an online tool freely available to examine the ability to learn new names for unfamiliar objects. Once test-takers have introduced a unique identification code, the program shows twenty unfamiliar objects. Then, by moving the mouse over the object, the program displays the name for that object. Two minutes are given to study this material. Then, the pictures disappear, and a new screen displays the twenty pictures again but arranged in a different order. The program tests participants by asking them to identify the correct picture for the name that is provided at the bottom of the screen (one at a time). The maximum score that can be obtained is 20.

4.2.4 English Proficiency Tests: Oxford Placement Test and V_YesNo

Oxford Placement Test 1

The *Oxford Placement Test 1* (OPT) (Allan, 2004) was used to assess participants' English proficiency. The test is divided into two different parts: *grammar* (tapping into grammar, vocabulary and reading skills) and *listening* (focusing on reading and listening skills). The highest score that can be obtained in each of the parts is 100 points because one point is awarded for each correct answer.

The *grammar* part consists of a written multiple-choice test that contains 100 sentences, in which learners must choose which of the three possible options given in each sentence is best. When possible, these sentences are contextualised and linked to one another. A sample item of the test is:

Item #19 – Like any top sportsman Ali ***had to*** / *must* / *should* train very hard.

The *listening* part contains another set of 100 original sentences uttered by native speakers, which test takers are asked to listen to and decide which of the two options given in each sentence corresponds to the version in the recording. Both options are semantically and grammatically plausible, so test takers can only rely on their listening skills to choose the answer. Participants could only listen to each sentence once. A sample sentence of the test is:

Item #42 – We've gone through *today's* / ***two days'*** money in less than an hour.

There is a correspondence between OPT scores and proficiency levels, i.e., below 75: beginner, 80–89: false beginner, 90–104: basic level, 105–119: elementary, 120–134: lower intermediate, 135–149: upper intermediate, 150–169: proficient, 170–189: highly proficient, 190–197: professional command, 198–200: functionally bilingual.

Receptive Vocabulary Size Test

V_YesNo v1.0 (Meara & Miralpeix, 2015) is an online test assessing receptive vocabulary size (i.e., the number of words a person knows in English). It does so by presenting a set of words, one at a time, in a context-free environment. Test-takers must decide whether or not they know the meaning of each word and answer accordingly. The test assesses words from the first 10 frequency bands in English (1K to 10K) and produces an estimate of total receptive vocabulary size at the end. So as to control for guessing, it also contains pseudo-words so that false-alarms are used to adjust the final vocabulary size estimation. *V_YesNo* is an adaptation of Yes/No tests proposed by Meara and Buxton (1987) and Meara and Jones (1988), who developed the technique for learners of English as an L2 (see Meara & Miralpeix, 2017, pp. 113–133 for a complete description on how the test works). Receptive vocabulary size has been used as a surrogate for L2 proficiency, as research has shown that this index tends to correlate with learners' linguistic abilities in the L2 (e.g., see Miralpeix & Muñoz, 2018). The numerical report should be interpreted as follows: below 2,000 (beginner), 2,000–3,000 (elementary), 3,000–5,000 (intermediate), 5,000–7,000 (upper-intermediate), 7,000–9,000 (advanced), 9,000–10,000 (highly proficient/native).

4.3 Procedure

Data for the present study were collected in the framework of a larger project in two different sessions. The first one took place at the beginning of the academic year (October 2020) in the computer room at the university, where students took the OPT (average time taken: 60 minutes) and the *V_YesNo* tests (average time taken: 10 minutes). The second session (December 2020) was conducted online due to the Covid-19 sanitary restrictions. Students were first asked to do the LLAMA test, which took about 5 minutes, and they immediately uploaded a screenshot of the results to a task on Moodle. Then, the teacher told them they would watch a short advert with the audio in English but with the subtitles in a language that was unknown to them. They were instructed to try to learn as much as possible of this new language. After the first viewing, they were told that they would watch the advert again for the same purpose, so in total they saw the advert twice.

After this, the teacher asked the participants to take the vocabulary test available on the Moodle Virtual Campus of the course, highlighting that they would just have three minutes to answer all the questions. The reason for this test to be tightly timed is that, as data were being collected online, researchers wanted to discourage participants from checking word meanings in dictionaries, which would have been feasible if a strict time limit had not been set. In addition, spending too much time thinking about the possible meaning of test items would indicate that a form-meaning link was not actually created and participants were just trying to guess the most plausible option among the distractors. It was also recommended that, in case of doubt, the option *I don't know* should be chosen instead of guessing. Scores for this test were automatically recorded on the Virtual Campus of the subject. Therefore, the duration of this second session was approximately 30 minutes in total.

It should be noted that no pre-test was used because of three reasons. First, none of the participants had been previously exposed to Polish; second, Polish language and participants' previously known languages belong to different language families; third, no cognates were included in the test. Therefore, this justifies that test-takers were not expected to know any of the test items or were able to make any inference based on cross-language similarities. Finally, we really wanted to avoid priming effects of any kind and make sure that the items were first encountered in the video and not before.

4.4 Data Analysis

After processing test scores using SPSS v.27, these were screened and statistically analysed. To answer the first research question, we examined how many word meanings were answered correctly in the Polish vocabulary test after participants had seen the advert. We further checked whether some words were more easily understood than others by a majority of the students, so we explored which were correctly identified by at least half of the participants and which were harder to learn.

In the case of research question 2, enquiring about a possible relationship between aptitude and language learning of these *ab initio* learners, Pearson product-moment correlations were performed between the scores on the Polish vocabulary test and the LLAMA scores, as both variables were found to have a normal distribution. In addition, a linear regression analysis was conducted, with LLAMA scores as the predictor variable and Polish vocabulary scores as the dependent variable, to further explore the extent to which aptitude could predict participants' performance on the meaning recognition test. Finally, correlations were drawn between LLAMA scores and participants' scores in those items that were more difficult to learn, i.e., those learned by less than 50% of the sample.

5 Results

Our participants had a mean receptive vocabulary size of 3,983 words ($sd = 941.3$), 95% CI [3647.44–4309] and an A2 proficiency level according to the OPT (mean = 116.67, $sd = 16.98$), 95% CI [112–121.72]. Even if they belonged to three different groups of the same subject at university (as mentioned in the Sect. 4.1), they had been assigned to these groups randomly. In addition, as no significant differences were found in their English level in a one-way ANOVA [OPT mean scores: G1 = 111.69 (*sd* 19.70); G2 = 117.50 (*sd* 14); G3 = 120 (*sd* 17.92)] [$F(2, 43) = 0.863$, $p = 0.429$], they will be treated as one group for the purpose of the present study.

5.1 Vocabulary Learning in the New Language After Minimal Exposure to Subtitled Multimodal Input

After watching the video twice, the mean score obtained by participants in the meaning recognition vocabulary test was 7.72 ($sd = 2.56$), 95%CI [6.96–8.48]. The range was wide (12), with a minimum of 2 and a maximum of 14 points. As an average, then, participants could make form-meaning connections in about half of the target items that were assessed.

Out of the 15 test items, there were 10 whose meaning was correctly identified by 50% or more of the participants (up to 74% of the learners). There were five items that

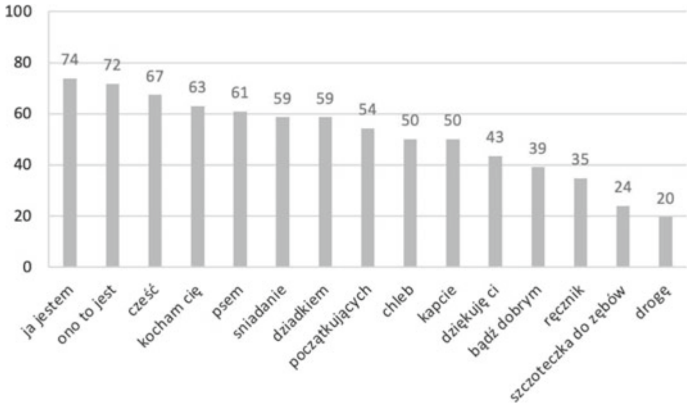

Fig. 2 Percentage of students answering correctly to each of the items in the meaning recognition vocabulary test (*Note.* All items occurred once in the advert except for *ja jestem* (6 times), *cześć* (5 times), *kocham cię* and *ono to jest* [2 times each])

were answered correctly by fewer participants (between 20 and 43% of the sample) as shown in Fig. 2 below. These were *dziękuję ci* ("thank you"), *bądź dobrym* ("be good"), *ręcznik* ("towel"), *szczoteczka do zębów* ("toothbrush") and *drogę* ("way"). An additional medium positive correlation was found between word frequency and the learning percentages displayed in Fig. 2 [$r = 0.575$, $n = 46$, $p = 0.025$].

5.2 *The Role of Aptitude in Vocabulary Learning After Minimal Exposure to Subtitled Multimodal Input*

In relation to LLAMA B scores, participants obtained a mean of 10.28 points ($sd = 4.32$) out of the possible 20. The 95% CI were [9–11.57] and the range was found to be 17, with a minimum score of 1 and a maximum of 18. There was a positive correlation between aptitude and vocabulary scores [$r = 0.510$, $n = 46$, $p < 0.001$], with high aptitude associated with high scores in the meaning-recognition vocabulary test in Polish.

A simple linear regression was calculated to predict participants' Polish vocabulary scores based on their aptitude. The analysis indicated that LLAMA B scores explained up to 24.3% of vocabulary scores. A significant regression equation was found [$F(1, 44) = 15.428$, $p < .001$], with an R^2 of 0.260 (see Tables 1 and 2 below). Participants' predicted vocabulary scores was equal to 4.610 + 0.302 (aptitude) Polish vocabulary scores when aptitude was measured in LLAMA B scores. Participants' average vocabulary scores increased 0.302 for each point in LLAMA B.

The correlation between LLAMA B scores and those obtained in the five most difficult words (i.e., those that less than 50% of participants could answer correctly)

Table 1 Results of linear regression analysis ($N = 46$)

Predictor variable	Dependent variable	R	R Square	Adjusted R Square	Std. Error of the Estimate
Llama B scores	Polish vocabulary scores	0.510	0.260	0.243	2.223

Note Predictor variable: LLAMA B scores. Dependent variable: Vocabulary scores

Table 2 ANOVA results ($N = 46$)

	Sum of Squares	df	Mean Squares	F	Sig
Regression	76.669	1	76.669	15.428	0.001
Residual	218.657	44	4.969		
Total	295.326	45			

Note Predictor variable: LLAMA B scores. Dependent variable: Vocabulary scores

[mean = 1.61, $sd = 1.37$, with a range of 5] was found to be of $r = 0.607$ ($p < 0.001$), showing that those participants with higher aptitude were better able to make form-meaning connections in items where peers with lower aptitude were not able to.

6 Discussion

This study was set up to explore the lexical information that adult participants could deliberately extract when watching a video subtitled in a new language. FE studies have rarely looked at novice learners' performance after being exposed to multimodal input even though the simultaneous presentation of sound, image and text can facilitate learning according to the Dual Coding Theory (Paivio, 2010) and the Cognitive Theory of Multimedia Learning (Mayer, 2014). From a theoretical perspective, these first stages of adult L2 acquisition should not be neglected as they have often been in vocabulary studies. At a practical level, research findings can inform communication and learning practices (e.g., to find the most adequate ways to boost learning of new languages from the very beginning, making use of online available input).

We opted for a simultaneous focus on meaning and form, as we asked our participants to choose the correct meaning for 15 TWs in Polish (the new language). These items were extracted from the three-minute video they had just watched twice. Therefore, according to the four-level typology of intake proposed by Rast (2008), we aimed at assessing level three, that is, when the item is perceived and comprehended, but not (re)produced. Results showed that meaning was correctly identified in about half of the items (mean of 7.72), which indicates that making form-meaning links for new vocabulary is possible with just very minimal information (i.e., after six minutes of exposure). In this respect, our results are in line with Gullberg et al., (2010, 2012), as their participants were also able to map meaning to some word forms when assessed with a sound-to-picture matching task. However, we should highlight that our study tested written form-meaning recognition rather than aural word recognition. Furthermore, the fact that subtitles in the new language were provided and that the audio was in English probably made form-meaning linking easier. Learning in our study may also have been facilitated because the unknown language was Polish and not Chinese, as in Gullberg et al., (2010, 2012) (and Polish was typologically closer to the L1 of the participants than Chinese was for Dutch speakers). In addition, participants in this experiment were explicitly asked to try to learn as much as possible from the new language and intentional learning can lead to high gains (considering the limited number of TWs and the short duration of the video, which participants watched twice). If participants had not consciously focused on learning as much as they could in the new language, chances are that learning would not have occurred (as in Bisson et al., 2014 with novice learners) or that it would have taken place to a much lesser extent. Contrary to the previous two studies, task orientations surely changed here the focus of the participants (VanPatten, 2014).

The findings from the first research question would also support Carroll's claims that "learners may need very little exposure to actually segment sound forms" (2014, p. 133). Even if her previous research was conducted with a different type of input (oral speech) and the present study uses a video where subtitles may aid segmentation, a few minutes of exposure may suffice for the learner. It should also be noted that in this case we are talking about deliberate learning instead of incidental (as in Carroll, 2014). The question would remain whether it is enough for active lexical production to take place, as Gullberg et al. (2012) have also wondered. This is an issue to be explored in further research.

It is also worth noting that, even if in our study TWs in the input were not manipulated for frequency and saliency as in Gullberg et al., (2010, 2012), the words learned by a higher number of participants were those that were repeated more frequently. As shown in Fig. 2, the four TWs in first positions occurred more frequently in the advert. Even if two of them appeared five or six times and two appeared twice, a Pearson correlation further confirmed the relationship between frequency and learning. This finding would coincide with those by Gullberg et al. (2012) in incidental learning, as they found out that Dutch speakers were sensitive to frequency and syllable structure when watching the unsubtitled video in Chinese. However, other studies have not found frequency to be such a relevant factor at the beginning. In Rast (2010a) and Shoemaker and Rast (2013), it became more influential as hours of exposure accumulated. A close observation to our data also suggests that there may be factors other than frequency enhancing learning, such as word relevance (i.e., items needed to understand the story in the advert). For instance, *dziadkiem* (grandfather) is learned by 59% of the sample and *cześć* (hi) by 67%. The learning percentages are similar if we consider that the first appears just once and the latter five times. However, *grandfather* is a very relevant word to understand what is happening in the advert (contrary to *hi*) and, even if word relevance may not play a role at later stages of acquisition when viewing subtitled TV (Peters & Webb, 2018), it can be an important factor when a reduced amount of input is available. Possibly, then, what may induce intake can be a combination of different factors, which deserve further research. Also, as VanPatten (2014, p. 198) points out, the role of frequency can be attenuated by the nature of the word being learned, as "not all words are created equal and some may require much less exposure to be processed and learned compared to others". Therefore, we call for future FE studies to explore how specific input attributes (e.g., word properties) affect processing and learning of new vocabulary from multimodal input.

In relation to the learner-related variable we have analysed, our results confirm that language aptitude for vocabulary learning can help word learning from subtitled input at the outset of the learning process, as the regression analysis suggests that the LLAMA scores would explain up to 24% of the variance in the meaning-recognition vocabulary test. Therefore, our findings support previous studies which have already indicated that aptitude is a significant factor in early learning (Dahlen & Caldwell-Harris, 2013; Doughty, 2019), also in novice learners in the light of the present results. That aptitude assists *ab initio* learners in maximizing learning from multimodal input is further corroborated by the fact that only those with high aptitude scores tended to make form-meaning connections in more challenging items, where those with a lower aptitude score could not.

The close correlation found in our experiment between vocabulary learned and aptitude test scores may also be due to the test used (LLAMA B), as it especially tapped into vocabulary acquisition (rote learning) as a subcomponent of aptitude (Li, 2016). Participants in this study needed to resort to such aptitude subcomponent when learning from the video (even if they were not told they would be tested on vocabulary but they were asked to try to learn as much as possible from the new language while watching the advert). Furthermore, LLAMA B is also believed to measure associative learning (i.e., making links between written forms of words and their referent) and explicit memory (and participants consciously focused on learning from the video), while other aptitude tests (such as LLAMA D) are thought to tap into more implicit types of learning (Granena, 2013). Consequently, the same result may not have been obtained had we used a *general* aptitude score or a test tapping more into implicit learning.

7 Conclusion and Further Research

Advances in digital technology have changed the way we are exposed to new languages. Even if research examining the ways in which *en route* learners could benefit from multimodal input has proliferated, FE studies analysing how *ab initio* learners extract linguistic information from this input modality do not yet abound. Results from the present study show that conscious adult L2 learning from very minimal input is taking place at least at the lexical level, where form-meaning connections are established for new words and expressions in the TL, especially in those that are more frequent or relevant in the input received. It remains to be explored in more detail, though, the particular word features that affect vocabulary learning in these conditions. Findings also evince that learner variables such as aptitude need to be considered in the early stages of vocabulary learning. Further research is needed to determine its role in the case of different language combinations and to identify which other variables (e.g., cognitive style, learning strategies) could facilitate learning from minimal input conditions. As Han and Sun (2014) point out, studying learners' individual differences in conjunction with input attributes is recommended.

Finally, it should be acknowledged that the present study is not without limitations, which can be successfully overcome in future research. For example, the number of TWs is quite limited and the participant sample could be enlarged to obtain more robust conclusions. It should also be noted that some of the tests were administered online. However, the implementation was tightly controlled and monitored. For instance, the vocabulary test was timed (as it also was the learning phase of the aptitude test), the advert could just be seen twice by the whole group, no external help was provided to test-takers and data were automatically saved to make sure that all participants took tests only once and completed all tasks in the assigned time.

Acknowledgements This work was supported by grant PID2019-110594GB-100 from the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033). We thank Neus Frigolé for her help with data collection, Paulina Olender for her revision of the Polish test and Paul M. Meara for his comments on an earlier version of the manuscript. We are also very grateful to the anonymous reviewers and the book editor for their valuable insights and comments.

Appendix: Meaning Recognition Vocabulary Test

Target Word (TW)	Option A	Option B	Option C	Option D
chleb (1)	breakfast	**bread**	fridge	I don't know
dziękuję ci (1)	goodbye	you are perfect	**thank you**	I don't know
dziadkiem (1)	wall	beach	**grandpa**	I don't know
ręcznik (1)	fork	bath	**towel**	I don't know
kapcie (1)	towel	pyjamas	**slippers**	I don't know
ono to jest (2)	**it is**	you are perfect	I love you	I don't know
szczoteczka do zębów (1)	**toothbrush**	I love you	I'm gonna fucking kill you	I don't know
śniadanie (1)	knife	**breakfast**	bread	I don't know
kocham cię (2)	you are perfect	I am gonna fucking kill you	**I love you**	I don't know
początkujących (1)	pyjamas	**beginners**	fork	I don't know
psem (1)	**dog**	beach	grandpa	I don't know

(continued)

(continued)

Target Word (TW)	Option A	Option B	Option C	Option D
drogę (1)	toilet	mirror	**way**	I don't know
bądź dobrym (1)	can you show me	thank you	**be good**	I don't know
cześć (5)	**hi**	thank you	be good	I don't know
ja jestem (6)	**I am**	you are	you can	I don't know

Note Correct answers for each TW appear in bold and TW frequency in the advert within parenthesis

References

Allan, D. (2004). *The Oxford Placement Test 1*. Oxford University Press.

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. https://doi.org/10.1017/S0142716412000434

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA Language Aptitude Tests. *Language Learning, 70*(1), 11–47. https://doi.org/10.1111/lang.12368

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test: MLAT manual*. Psychological Corporation.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. John Benjamins.

Carroll, S. E. (2005). Input and SLA: Adults' sensitivity to different sorts of cues to French gender. *Language Learning, 55*(Suppl1), 79–138. https://doi.org/10.1111/j.0023-8333.2005.00296.x

Carroll, S. E. (2013). Introduction to the special issue: Aspects of word learning on first exposure to a second language. *Second Language Research, 29*(2), 131–144. https://doi.org/10.1177/0267658312463375

Carroll, S. E. (2014). Processing 'words' in early-stage foreign language acquisition: A comparison of first exposure and low proficiency learners. In Z. H. Han & and R. Rast (Eds.), *First exposure to a second language* (pp. 107–138). Cambridge University Press.

Carroll, S. E., & Widjaja, E. (2013). Learning exponents of number on first exposure to an L2. *Second Language Research, 29*(2), 201–229. https://doi.org/10.1177/0267658312473471

Cobb, T. (n.d.). *VocabProfile VP Classic* [Computer program]. Accessed 23 April 2021 at https://www.lextutor.ca/vp/eng/

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research, 1123*(1), 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal, 97*(4), 902–916. https://doi.org/10.1111/j.1540-4781.2013.12045.x

de Diego Balaguer, R., Toro, J. M., Rodríguez-Fornells, A., & Bachoud-Lévi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PloS ONE, 2*(11). https://doi.org/10.1371/journal.pone.0001175

Dimroth, C., Rast, R., Starren, M., & Watorek, M. (2013). Methods for studying the acquisition of a new language under controlled input conditions. *EUROSLA Yearbook, 13*, 109–138. https://doi.org/10.1075/eurosla.13.07dim

Doughty, C. (2019). Cognitive language aptitude. *Language Learning, 69*(1), 101–126. https://doi.org/10.1111/lang.12322

d'Ydewalle, G., & Pavakanum, U. (1995). Acquisition of a second/foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in Europe: The state of the art—perspectives for the future* (pp. 51–64). Westdeutscher Verlag.

Folia, V., Uddén, J., de Vries, M., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in children. *Language Learning*, *60*(Suppl. 2), 188–220. https://doi.org/10.1111/j.1467-9922.2010.00606.x

Gesa, F. (2019). *L1/L2 subtitled TV series and EFL learning: A study on vocabulary acquisition and content comprehension at different proficiency levels.* PhD Thesis, University of Barcelona, Barcelona, Spain. https://www.tesisenred.net/handle/10803/668505

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105–130). John Benjamins Publishing Company.

Gullberg, M., & Indefrey, P. (Eds.) (2010). The earliest stages of language learning. *Language Learning, 60*(Suppl. 2). http://onlinelibrary.wiley.com/doi/10.1111/lang.2010.60.issue-s2/issuetoc

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *IRAL—International Review of Applied Linguistics in Language Teaching*, *50*(4), 239–276. https://doi.org/10.1515/iral-2012-0010

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning, 60*(Suppl. 2), 5–24. https://doi.org/10.1111/j.1467-9922.2010.00598.x

Han, Z. H., & Liu, Z. (2013). Input processing of Chinese by ab initio learners. *Second Language Research, 29*(2), 145–164. https://doi.org/10.1177/0267658313479359

Han, Z. H., & Rast, R. (Eds.), (2014). *First exposure to a second language*. Cambridge University Press.

Han, Z. H., & Sun, Y. A. (2014). First exposure: A replication of Han and Peverly (2007). In Z. H. Han & R. Rast (Eds.), *First exposure to a second language* (pp. 7–39). Cambridge University Press.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.) (pp. 43–71). Cambridge University Press.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*(7), 702–704. https://doi.org/10.1038/nn1264

Meara, P. M. (2005). *LLAMA Language Aptitude Tests.* Lognostics. https://www.lognostics.co.uk/tools/LLAMA_3/index.htm

Meara, P. M., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*(2), 142–154. https://doi.org/10.1177/026553228700400202

Meara, P. M., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied Linguistics in Society* (pp. 80–87). CILT.

Meara, P. M., & Miralpeix, I. (2015). *V_YesNo.* Lognostics. https://www.lognostics.co.uk/tools/V_YesNo/V_YesNo.htm

Meara, P. M., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL skills. *IRAL—International Review of Applied Linguistics in Language Teaching, 56*(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Muñoz, C., & Singleton, D. (2019). Age and multilingualism. In D. Singleton & L. Aronin (Eds.), *Twelve lectures on multilingualism* (pp. 213–230). Multilingual Matters.

Nation, I. S. P. (1996). *Vocabulary lists*. English Language Institute at Victoria University of Wellington, Occasional Publication n.17.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon, 5*(2), 205–230. https://doi.org/10.1075/ml.5.2.04pai

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Raine, P. (2012). *Incidental learning of vocabulary through subtitled authentic videos.* Master's thesis, University of Birmingham, Birmingham, United Kingdom. https://www.birmingham.ac.uk/index.aspx

Rast, R. (2008). *Foreign language input: Initial processing*. Multilingual Matters.

Rast, R. (2010a). The role of linguistic input in the first hours of adult language learning. *Language Learning, 60*(Suppl. 2), 64–84. https://doi.org/10.1111/j.1467-9922.2010.00601.x

Rast, R. (2010b). The use of prior linguistic knowledge in the early stages of L3 acquisition. *IRAL—International Review of Applied Linguistics in Language Teaching*, *48*(2–3), 159–183. https://doi.org/10.1515/iral.2010.008

Rast, R., Watorek, M., Hilton, H., & Shoemaker, E. (2014). Initial processing and use of inflectional markers: Evidence from French adult learners of Polish. In Z. H. Han & R. Rast (Eds.), *First exposure to a second language* (pp. 64–106). Cambridge University Press.

Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology, 2*(1), 157–183. https://doi.org/10.1515/labphon.2011.005

Sagarra, N. (2014). Absence of morphological transfer in beginners: Evidence from eye tracking. In Z. H. Han & R. Rast (Eds.), *First exposure to a second language* (pp. 139–170). Cambridge University Press.

Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*(4), 607–646. https://doi.org/10.1017/S0272263115000327

Shoemaker, E., & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research, 29*(2), 165–183. https://doi.org/10.1177/0267658313479360

Suárez, M. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *The Language Learning Journal, 47*(3), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. https://doi.org/10125/44214

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. Palgrave Macmillan.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research.* Ablex.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 115–136). Lawrence Erlbaum Associates.

VanPatten, B. (2014). Epilogue: Input processing by novices—Issues in the nature of processing and in research methods. In Z. H. Han & R. Rast (Eds.), *First exposure to a second language* (pp. 193–207). Cambridge University Press.

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching, 50*(1), 1–31. https://doi.org/10.1017/S0261444816000276

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal, 97*(1), 109–130. https://doi.org/10.1111/j.1540-4781.2013.01428.x

Yang, L., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck second language learning project. *Studies in Second Language Acquisition, 19*(2), 172–193. https://doi.org/10.1017/S0272263197002039

Imma Miralpeix is an associate professor at the University of Barcelona. Her main research interests include second language vocabulary acquisition, especially lexical development and assessment, and multilingualism. She is the author of several publications in these areas and has taken part in different funded projects on second language learning and teaching.

Ferran Gesa holds a PhD in Applied Linguistics from the Universitat de Barcelona. He is now postdoctoral researcher at this same university and visiting scholar at the Universidad de La Rioja. His main research interests include vocabulary acquisition, extensive viewing, and the role of individual differences in language learning.

Maria-del-Mar Suárez is an associate professor at the Faculty of Education of the University of Barcelona. She has taken part in several funded projects on second language acquisition. Her research focuses on individual differences, multimodal input, and formative assessment. She is also the Editor-in-Chief of *Didacticae: Journal of Research in Specific Didactics.*

Academic Videos for Incidental Vocabulary Learning Among ESL Foundation Students

Jessica McLaughlin, Csaba Szabo, Arathi Jane Reddy, and Mangeet Kaur Khera

Abstract Research in the Malaysian context has shown that English as a second language (ESL) students have an insufficient vocabulary for higher education and thus, must be provided with opportunities to acquire new words. Recently, evidence has emerged that viewing audio-visual input is effective for incidental vocabulary learning (IVL). Yet, there is limited understanding of the effectiveness of academic videos for IVL, an accessible, cost and time-efficient resource that is encouraged to be used for promoting personalised and interactive teaching and learning. Therefore, this research investigated the potential vocabulary gains through a 10-min academic video among 56 ESL foundation students in an English-medium university in Malaysia. Before watching the video, IVL was measured using a modified Vocabulary Size Test, followed by a comprehension test and a post-target words test afterward. Results indicated a significant gain of 0.78 words on average after watching the video, where thirty participants made gains of 44 words in total. However, further research is necessary to investigate the role of various learner-related factors on IVL. These findings provide insight for educators on how teaching practices using academic videos may address the low vocabulary knowledge of ESL learners and develop long-lasting L2 proficiency for successful education.

Keywords Incidental vocabulary learning · Academic videos · Second language acquisition · Learner-related factors · Innovative teaching strategies

1 Introduction

Vocabulary learning plays a vital role in the development of second language (L2) reading, speaking, listening, and writing proficiency (Schmitt, 2010). Since these critical skills impact academic achievement and successful L2 acquisition for higher

J. McLaughlin (✉) · C. Szabo · A. J. Reddy · M. K. Khera
University of Nottingham Malaysia, Semenyih, Malaysia
e-mail: jessicamclaughlin00@gmail.com

C. Szabo
e-mail: csaba.szabo@nottingham.edu.my

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_11

education (Alsager & Milton, 2016; Harrington & Roche, 2014), classroom teaching has emphasised explicit vocabulary teaching. However, the resulting marginal growth in students' vocabulary suggests that explicit teaching may be an inefficient teaching method for their L2 vocabulary acquisition (Bisson, 2013; Montero Perez, 2020; Siyanova-Chanturia & Webb, 2016). Therefore, the integration of in-class and out-of-class activities—such as audio-visual input—as a language learning strategy provides English as a Second Language (ESL) learners with a better opportunity for acquiring new words.

Incidental vocabulary learning (IVL) is an indirect process of learning vocabulary that occurs as a by-product of performing a meaning-focused activity (Webb & Nation, 2017). Previous research has found that reading and listening are effective for IVL (Bisson, 2013; Erlandsson & Wallgren, 2017) and recent evidence has emerged that television shows and movies can lead to IVL (Ashcroft et al., 2018; Montero-Perez, 2020; Peters & Webb, 2018). Similarly, watching TED Talk videos with an assistive oral output task has been shown to enhance vocabulary learning (Nguyen & Boers, 2018). According to a recent lexical coverage analysis of TED Talks, academic videos could be beneficial for vocabulary learning, particularly for ESL students who need supplementary English programmes (Nurmukhamedov, 2017). Yet, little is known about the effectiveness of viewing academic videos for IVL in the context of L2 learners. Academic videos are an accessible, low-cost, and time-efficient resource that can be used in blended learning and flipped classrooms to make learning more personalised and interactive (López-Pérez et al., 2011; Montgomery et al., 2019; Nwosisi et al., 2016). Thus, it is important to explore how watching academic videos can facilitate IVL. This study aimed to investigate the effectiveness of academic videos in promoting vocabulary learning in an ESL context at an English-medium university in Malaysia. The findings could help educators understand how academic videos can be used in in-class and out-of-class activities and practises to address the problem of low vocabulary knowledge in L2 learners.

2 Literature Review

2.1 Incidental Vocabulary Learning through Audio-Visual Input

Two theories contribute to a better understanding of the effectiveness of IVL as a by-product of viewing audio-visual input. The Multimedia Principle proposed by Mayer (2005) suggests that people learn better through both words and pictures as the brain constructs verbal and pictorial schemas by selecting and integrating information from multimedia elements. Furthermore, according to the Dual-Coding Theory (Paivio, 1971), words and meanings are processed in two independent but connected channels: verbal code and nonverbal code. The verbal code processes forms of language like speech, while the nonverbal code processes our knowledge of the world through imagery; the links between the codes allow for the decoding of words (Sadoski, 2005). These theories are supported by empirical evidence where the integration of captions, audio, and images in videos assisted L2 learners comprehend unknown words (Wong & Samudra, 2019) as they used their background knowledge, vocabulary, and comprehension strategies to establish a link between the auditory and visual stimuli (Teng, 2018, 2020). Thus, both theories suggest that multiple and simultaneous input consisting of auditory, written, and pictorial elements shown in videos can improve learning and the retention of vocabulary knowledge (Fig. 1).

Existing evidence shows that vocabulary can be learned incidentally through audio-visual input. The extensive viewing of a full-length, one-hour L2 television programme led to vocabulary gains of approximately four words, or a 14% increase in meaning recognition (Peters & Webb, 2018). Viewing a documentary twice for a total exposure of 45-min also had a significant impact on word learning, with an average of 2.97 words picked up after viewing (Montero Perez, 2020). Likewise, watching the first 60 min of a L2 captioned movie resulted in an average recall of 1.77 (4.2%) words, even when the post-test was taken two-days after viewing the video (Ashcroft et al., 2018). Specific to academic videos, watching a 12-min TED Talk video twice resulted in a modest gain where one group that revised and organised notes gained approximately 5 words and the other group that completed a content summary output activity had a higher gain of approximately 8 words (Nguyen & Boers, 2018). Thus, there is evidence that audio-visual input is effective in incidentally expanding receptive vocabulary knowledge; however, there is limited understanding of IVL through viewing academic videos.

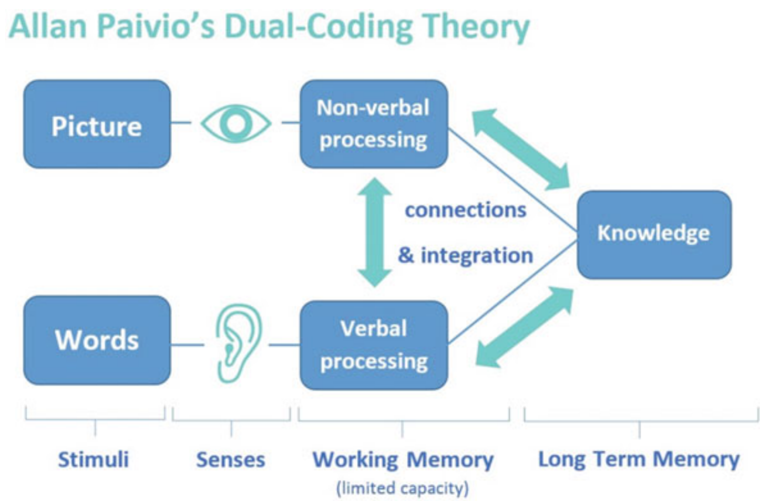

Fig. 1 Diagram of the dual-coding theory (*Note* Public domain graphic adopted from Kirschner and Neelen [2017])

2.2 The Importance of Comprehension in Incidental Vocabulary Learning

Various factors have been recognised to significantly impact IVL for L2 learners, including video input related factors (Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2020) and individual differences in knowledge or abilities (Lin, 2010; Monter-Perez, 2020; Peters & Webb, 2018). Specifically, content comprehension is an important factor that needs to be outlined to understand the extent to which IVL may take place through viewing academic videos.

Content comprehension of input is crucial for IVL (Montero Perez et al., 2014), as the acquisition-learning distinction hypothesis states that the process of IVL involves the learners' ability to guess and acquire new words from contextual clues (Ahmad, 2012) through their "feel for correctness" (Krashen, 1982, p. 10). This suggests that lexical development occurs as learners attempt to derive the meaning of new words through comprehension of context when encountering new information (Bisson, 2013; Kara, 2013; Reynolds & Turek, 2012). Given that academic videos contain visual, written, and spoken input that provide learners with various support to understand the video content, it is evident that reading or listening comprehension is important in the process of IVL.

The importance of reading and listening comprehension in IVL is outlined in the Input Hypothesis (Krashen, 1985), which states that language is acquired when receiving input that is slightly more advanced than the learners' level but still comprehensible. Furthermore, the Matthew Effect (Stanovich, 1986) suggests a cumulative advantage where one's vocabulary knowledge facilitates reading comprehension and reading comprehension skills facilitate vocabulary learning. This reciprocal relationship also applies to L2 learners (Geva & Wiener, 2015), as learners with better reading comprehension are likely to have more vocabulary uptake due to better inference from contextual clues. Likewise, listening comprehension may facilitate greater understanding of context and allow learners to extract information of new words from the example sentences provided (Zhang & Graham, 2020).

As suggested, comprehension is determined to be integral to IVL and retention of new lexical items, as greater levels of comprehension lead to enhanced activation of relevant background knowledge to construct mental representations and successful lexical inferencing (Pulido, 2007). Respectively, a previous study on participants' vocabulary learning and news video comprehension found a moderate relationship between overall video comprehension and vocabulary recognition (Lin, 2010). Thus, the extent to which IVL occurs through watching academic videos may depend on their understanding of the video content and how they can use their reading and listening comprehension to successfully guess the meaning of unknown words.

2.3 The Use of Academic Videos as a Learning Resource

Academic videos, also called supplementary videos, are short videos used as an educational tool by complementing learning objectives (Ljubojevic et al., 2014). Due to their availability, they are commonly used in English classrooms (Kirana, 2016) or embedded in homework tasks (Brame & Perez, 2016) to make learning more interactive and interesting in both in-class and out-of-class activities. Like general audio-visual input, they contain visual elements such as motion graphics which are effective in improving the retention and attention of learners (Kumar & Jamil, 2016). These non-verbal linguistic features are provided with authentic language input, which aids learners' understanding of how the target language is used and acts as a stimulus to activate their background schema of the subject (Kirana, 2016).

However, it is imperative to conduct research explicitly in the context of academic videos as they differ from general audio-visual input such as television shows and movies. Academic videos contain relevant, authentic content and specific academic vocabulary to expand the learners' understanding of the content (Miner & Stefaniak, 2018; Webb & Nation, 2017). This is effective in motivating ESL students to acquire the target culture and language (Bajrami & Ismaili, 2016) necessary for understanding specialist vocabulary, comprehending lectures, and using academic language in higher education (Evans & Morrison, 2011). These distinctions call for additional research to understand the effects of academic videos on IVL for L2 learners.

2.4 Purpose of the Study

Findings converge to suggest that viewing audio-visual input is effective for IVL; however, it has not yet been explored in the context of academic videos. Given that academic videos are often used in higher education (Carmichael et al., 2018), it is crucial to investigate whether they promote IVL. This research may suggest the effectiveness of IVL in higher education and potentially how to address the problem of low vocabulary knowledge in L2 learners.

Consequently, this research aims to fill the gaps identified by addressing the following research question: To what extent does watching academic videos enable incidental vocabulary learning for ESL learners in higher education?

3 Methodology

3.1 Research Design

The research employed a quantitative methodology with a pre-modified vocabulary size test (VST) and post-target words test. A repeated measures design was adopted to study the effect of academic video viewing on learners' vocabulary size before and after the exposure. As participants served as their own controls for factors that cause variability (i.e., vocabulary knowledge), it increased the statistical power of this experimental study (Ellis, 1999).

3.2 Participants

Convenience sampling was used to gather students ($n = 56$) from Foundation in Arts and Education and Sciences from a private international university in Malaysia. The average reported age of the participants was 18.9 years old, as shown in the participant demographics in Table 1. All participants were ESL students with a variety of first languages and a minimum level of English proficiency equivalent to CEFR B2 level or 6.0 in IELTS needed for their university course entry requirements. Although these learners are considered independent users of English at an upper intermediate proficiency level, most students who use English as an L2 in higher education are found to have insufficient vocabulary knowledge (Tan & Goh, 2020). In addition, despite their passing score in the English proficiency test, foundation students may not have acquired specific academic vocabulary due to a lack of experience and learning in their field of study. This suggests that the sample can represent ESL learners in English-medium higher education who have limited vocabulary knowledge.

3.3 Materials

3.3.1 Video

The academic video selected for this research was a 10-min CrashCourse (2014) video on "Social Influence". The video explored the effect of social influence on people's actions, which is relevant to a topic in the foundation students' core module that aims to examine views that affect educational policies. Therefore, the video can be classified as an academic video which complements learning objectives by providing additional academic context to expand learners' understanding (Ljubojevic et al., 2014) and contains both verbal and non-verbal linguistic features to accommodate learners' understanding of how the target language is used in context (Kirana,

Table 1 Participant demographics

Variable	Total Participants ($n = 56$)	%
Age		
Mean (SD)	18.9 (1.20)	
16	1	1.79
17	5	8.93
18	15	26.8
19	17	30.4
20	15	26.8
21	1	1.79
22	2	3.57
First language		
Malay	13	23.3
Mandarin	33	58.9
Tamil	3	5.36
Korean	1	1.79
Japanese	2	3.57
Dhivehi	1	1.79
Spanish	1	1.79
Arabic	1	1.79
Other	1	1.79

2016). The video was approved by the module convenor as it can potentially reflect participants' learning in a lecture and promote engagement.

3.3.2 Target Words

Twenty target words that were relevant for understanding the content of the video were selected from a range of frequency levels and an equal ratio of the parts of speech (nouns, verbs, and adjectives). To select the words, the New Academic Word List (Browne et al., 2013) was used, as it allows for the comparison of the frequency of common words from a range of academic disciplines in an English academic register based on the British National Corpus (BNC) (Hyland & Tse, 2007). However, this list is within the domain of general academic English, which did not have specific vocabulary from the video. Therefore, additional academic words were selected, and frequency levels were identified using the BNC on the Lexical Tutor website (https://www.lextutor.ca/) (Cobb, 2020). Furthermore, participants were considered to have mastered the most frequent 2,000 word families, which are deemed necessary for studying content in English, as they were attending an English-medium university with specific English entry requirements (Roche & Harrington, 2013).

Table 2 Target words with the part of speech, frequency level, and frequency of occurrence in video

Word	Part of speech	Frequency level	Frequency of occurrence
Restraint	Noun	4 K	1
Publicised	Verb	4 K	1
Incompetent	Adjective	4 K	1
Subsequent	Adjective	4 K	1
Comply	Verb	5 K	4
Exert	Verb	5 K	1
Deception	Noun	6 K	1
Prestigious	Adjective	6 K	1
Arouse	Verb	6 K	1
Fiascoes	Noun	7 K	1
Mesh	Verb	7 K	1
Prods	Noun	8 K	2
Increments	Noun	8 K	1
Mimicry	Noun	8 K	2
Lynch	Adjective	9 K	1
Contagious	Adjective	9 K	1
Underscored	Verb	9 K	1
Defiance	Noun	11 K	1
Yelps	Noun	13 K	1
Normative	Adjective	13 K	2

Consequently, to reduce the likelihood of participants' having prior knowledge of the selected words, the target words were selected from mid to low frequency levels of 3,000 and above. The target words with the relevant parts of speech, frequency level, and frequency of occurrence in the video are listed in Table 2.

3.4 Research Instruments

3.4.1 Target Words Test

The target words test was used as a pre-test before and a post-test after watching the video. Each test consisted of the 20 target words in the same four-option multiple-choice format as the VST, with all options from the first to third frequency levels. The pre-test was incorporated into the modified VST which acted as distractors to prevent participants from identifying the target words while watching the video. Cronbach's alpha indicated moderate internal consistency (Brown, 2014) for the pre-test consisting of 80 items ($\alpha = 0.71$) and the post-test consisting of 20 items

1. PRESTIGIOUS: It is a **prestigious** school.
    a. Historic
    b. Recognised
    c. Small
    d. New

Fig. 2 Extract from item in the target words test

($\alpha = 0.75$), which suggests that the items in the test were reliable for measuring the participants' knowledge of the target vocabulary before and after watching the video.

The target words tests were scored binomially, using 0 for an incorrect response and 1 for a correct response. Although relative gains consider the variability in opportunities to learn words for each participant (Webb & Chang, 2015), absolute gains were used to score the tests as it is a more conservative criterion (Shefelbine, 1990). The absolute gains were found by identifying the number of words that were unknown in the pre-test but known in the post-test (Fig. 2).

3.4.2 Vocabulary Size Test

Spoken or written audio-visual input is likely to result in the acquisition of receptive vocabulary knowledge (Schmitt & Zimmerman, 2002), which refers to the ability to recognise word form and retrieve word meaning. In other words, knowing what the words look and sound like without necessarily being able to pronounce or spell them (Nation, 1980; Zhong, 2011). This is because new words are initially only recognised before passing through various phases of partial word knowledge to be produced accurately (Arndt & Woore, 2018). Therefore, the VST was used to measure receptive vocabulary size through form-meaning recognition (Nation & Beglar, 2007; Stoeckel et al., 2016). The VST indicates a reliable estimate of vocabulary size as words are assessed at different frequency levels (e.g. Miralpeix & Muñoz, 2018), consisting of 10 items from the first to fourteenth 1,000 most frequent word families of English according to the BNC (Nation & Beglar, 2007). However, the VST was modified to reduce fatigue or boredom from the extensive use of tests in this study, which may cause unreliable results if the participants guess or do not complete the tests (Lavrakas, 2008). The number of questions was reduced to five randomly selected items from each third to fourteenth frequency level (see Szabo et al., 2020) in a four-option multiple-choice format. To calculate the vocabulary size, the number of correct answers on the VST was added, then multiplied by 200. Consequently, the VST used in this study was a moderately reliable (60 items; $\alpha = 0.86$) instrument to measure estimated vocabulary size.

3.4.3 Comprehension Test

To ensure that IVL occurred as a by-product of a meaning-focused activity, a comprehension test was taken to evaluate the participants' understanding of the video. To reduce the guessing effect and distinguish between low and advanced learners, the comprehension test consisted of 10 questions: 5 multiple-choice questions with four options each and 5 short-answer questions (Brown, 2004; Ko, 2010). All questions were explicit and implicit content questions based on factual information in the video (Freeman, 2014) and were created following specific guidelines in terms of organisation, language, and distractors (see Gay, 2010). The test had a total of 20 points based on a criterion for each question (Lesage et al., 2013), and inter-rater reliability was established through an agreement between three examiners to ensure consistency in scoring (Wang, 2009). Data from three participants who scored less than 40% were removed as the tests were deemed as incomplete or not focused on the content of the video. Cronbach's alpha for the 10 comprehension test items was 0.70, indicating adequate internal consistency for measuring participants' video content comprehension.

3.5 Procedure

The Research Ethics Form was approved, and ethical clearance was obtained as guidelines set by the British Education Research Association (BERA, 2011) and the university were followed. Participants took part in this research as a group through an online platform called Microsoft Teams across two weeks during two sessions of their regular lecture hours for their core module. In the first session, the participants received the consent and an information sheet outlining the procedure and conditions prior to conducting the research to ensure that voluntary consent was obtained. Then they participated in a series of activities outlined in Table 3: taking the modified VST to measure English receptive vocabulary size and prior knowledge of target words, watching the academic video, and then completing the post-target words test and the comprehension test. This order was chosen to reduce the impact of the comprehension test on the post-test, where students may see the target words or gain a better understanding of the video after answering the questions. The participants were not given detailed information on what type of test they would take to create a more natural learning environment. Furthermore, no specific instructions were provided for the tools required to attend the sessions. Participants were encouraged to follow their usual study habits and therefore, were given the freedom to participate through various devices (i.e., computer, phone), use headphones, or take notes while watching the video. All the participants' data were uniquely coded to ensure anonymity and were only accessed by the researchers for confidentiality.

Table 3 Session experimental procedure

Procedure	Duration (minutes)
Consent and information sheet	5
Modified vocabulary size test	30
Break	5
Academic video	10
Post-target words test	10
Comprehension test	10

3.6 Data Analysis

After confirming that the data fit the assumptions of parametric tests by assessing the normality using the Shapiro–Wilk test and histograms, inferential statistics were analysed using R (Rstudio Team, 2020). A paired sample *t*-test was utilised to answer the research question by analysing whether the scores of the target words test taken before and after watching the video differed statistically.

4 Results

Table 4 shows the descriptive statistics for the pre- and post-target words tests. The average gain for participants was 0.78 (3.9%) target words after watching the 10-min academic video once, with the highest gain of 4 words and a reduction of -3 words. Thirty participants (53.6%) made gains and a total of 44 words was gained by all participants.

A paired sample *t*-test was conducted to investigate the difference in pre-test and post-test scores for the same group of participants before and after watching the academic video. Although a Shapiro–Wilk test (pre-test, $p = 0.024$; post-test, $p = 0.0006$) found that the data was not normally distributed ($p < 0.05$), a paired sample *t*-test was used because *t*-tests are robust to violations of normality, especially with larger sample sizes, and can be preferred over non-parametric tests, which exclude data when transforming measurements into ranks (Zumbo & Jennings, 2002). The difference in pre-test and post-test scores was found to be statistically significant ($t(55) = -3.92$, $p < 0.001$), with a medium effect size ($d = -0.79$). Therefore, the null hypothesis was rejected, as the findings suggest that watching academic videos promotes IVL for ESL learners in higher education.

Table 4 Descriptive statistics for pre- and post-target words tests

Test Scores	*n*	*M*	*SD*	*Min*	*Max*
Pre-Test	56	15.3	3.05	6	20
Post-Test	56	16.1	3.05	8	20

5 Discussion

This study aimed to investigate the extent to which IVL occurs from watching academic videos that have auditory, written, and pictorial elements relevant to the participants' course of study. Overall, the students in this study gained target words after watching the video, suggesting that ESL learners in higher education can incidentally learn vocabulary through watching academic videos.

The research question was explored through a comparison of participants' knowledge of form-meaning recognition of the target words before and after watching the academic video. The results indicate that vocabulary was learned after watching the 10-min academic video once, with an average absolute gain of 0.78 (3.9%) target words. Although findings from this study show a relatively lower average gain of words compared to previous research that found a mean increase of approximately 4 words (Peters & Webb, 2018), 2.97 words (Montero Perez, 2020), 1.77 words (Ashcroft et al., 2018), and 5 to 8 words (Nguyen & Boers, 2018), it is consistent in revealing that IVL occurs from viewing audio-visual input. The greater gain in previous studies was likely due to the longer viewing of 60-min videos and the substantial number of 64 target words (Peters & Webb, 2018), 15 pseudowords (Montero Perez, 2020), and 42 target words (Ashcroft et al., 2018) compared to the 10-min video with 20 target words in this study. Furthermore, viewing TED Talk videos twice with subsequent comprehension tasks (Nguyen & Boers, 2018) may have led to greater vocabulary gains compared to this study, where learners viewed the video once and did not have any additional tasks. Nevertheless, the results provide evidence to support the Multimedia Principle (Mayer, 2005) and Dual-Coding Theory (Paivio, 1971) in concurrence with previous studies (Yawiloeng, 2020; Teng, 2018, 2020) that simultaneous auditory, written, and pictorial elements in audio-visual input improve vocabulary learning (Lin, 2010).

Moreover, the relatively low average gain of words may be related to the extent to which participants understood the content of the video. Research indicates that vocabulary is gained as learners attempt to derive the meaning of words through inference from contextual clues (Ahmad, 2012; Reynolds & Turek, 2012). Although the scores from the comprehension test were normally distributed, the average score was 12.7 out of 20, suggesting that the video may have been somewhat difficult for the participants. This may have been due to the relatively high speech rate of the speaker in the video, which has been found to be a factor that affects second language listening comprehension (Chang, 2018). Furthermore, the participants' average vocabulary size ($n = 56$, $M = 8800$) is seen as adequate to comprehend the text without assistance, as research shows that a vocabulary size of approximately 5,000 to 9,000 most frequent word families is sufficient for understanding television programmes and academic videos like TED talks (Nurmukhamedov, 2017; Webb & Rodgers, 2009). This therefore suggests that the participants had appropriate prior vocabulary knowledge, but other factors contributed to the lack of video content comprehension and inadequate assistance for successful lexical inferencing to occur, resulting in a low average of IVL.

Furthermore, the significant vocabulary gain of target words may be related to the frequency of occurrence of target words in the video, as studies propose that a word is more likely to be noticed when occurred more frequently (Restrepo Ramos, 2015; Rodgers & Webb, 2020). Previous research found a medium correlation between frequency of occurrence and vocabulary learning (Rodgers, 2013) and that it was 20% more likely for a correct response when the frequency of occurrence increased (Peters & Webb, 2018). Likewise, the current study suggests that the word that had the highest frequency of occurrence of 4 times (i.e. *comply*) was gained by all six participants who had not known the word on the pre-test (100%), a word that appeared twice (i.e. *mimicry)* was gained by three out of five participants who had previously not known the word (60%), and a word that only appeared once (i.e. *publicised*) was learned by nine of the 20 participants who had not known the word previously (45%), after the viewing. However, results also showed that a word that appeared twice (i.e. *prods*) was learned by 4 out of 24 participants (16.6%) while a word that only appeared once (i.e. *mesh*) was learned by 6 participants out of 15 participants who had not known the word previously (40%). Consequently, no common pattern was observed between the frequency of occurrence and vocabulary gained. This can be supported by studies which found that the frequency of occurrence in spoken input had little effect on vocabulary learning despite the repeated encounters of target words (Brown et al., 2008; van Zeeland & Schmitt, 2013). Yet, these studies explore incidental exposure from listening, suggesting that frequency of occurrence may play a different role in audio-visual input viewing which provides both listening and visual support. Therefore, further research is necessary to understand the relationship between the frequency of occurrence and IVL through audio-visual input.

Essentially, the findings provide an insight into how viewing academic videos may be an effective tool for IVL and propose possible benefits if utilised in in-class and out-of-class activities. With the advancement of technology, there are increased opportunities for students to encounter academic videos, whether through social media, open online courses, or online educational platforms (Ali, 2019). Given that academic videos are accessible online at all hours, cost and time-efficient, and accommodate the individual needs of learners (Carmichael et al., 2018), teachers may consider integrating these tools to supplement learning. For example, to use the academic videos as powerful tools in classroom activities, teachers may employ viewing and listening techniques such as silent, partial, or repetitive viewing (Hadijah & Pd, 2016). Outside of the classroom, academic videos can be embedded into larger homework assignments with additional activities such as guiding questions to keep the students active in the learning process (Brame & Perez, 2016). As academic videos can be viewed at home before or after class, time in class can be strategically utilised for interactive activities such as discussions or quizzes (Tuna et al., 2018) and students can further autonomous learning at home through various videos, including those that present graphic visualisations of experiments and problems, show examples of real-life cases, or summarise the content (Rajadell & Garriga-Garzón, 2017). As shown, academic videos can be used effectively in and out of the classroom to supplement students' content and vocabulary learning; should be encouraged in both teaching and learning.

Overall, the findings indicate that IVL occurs for ESL learners in higher education as vocabulary was learned at the form-meaning recognition level after watching the academic video. Unlike previous studies that explored audio-visual input, such as television shows (Peters & Webb, 2018), documentaries (Montero Perez, 2020), and movies (Ashcroft et al., 2018), the current study explored IVL through short academic videos. This is an innovative approach to incidental vocabulary development that can be easily used in teaching for in-class and out-of-class activities to complement content learning objectives and encourage personalised learning (Ljubojevic et al., 2014; Nwosisi et al., 2016). Therefore, this research coincides with previous research that show that extensive viewing leads to more sizeable vocabulary gains (Montero Perez, 2020; Webb, 2015) and provides insight into how academic videos can be integrated into authentic activities to provide opportunities for learners to acquire vocabulary.

6 Limitations and Recommendations for Further Research

This study has several limitations related to the nature of the participants and materials used. Suggestions to overcome these limitations may provide further insight into the extent to which IVL occurs while watching academic videos.

Firstly, participants in this study are from an English-medium university and their vocabulary size ($n = 56$, $M = 8800$) is deemed adequate to comprehend the L2 academic video where approximately an 8,000 word-family vocabulary, prior knowledge of the base word with its inflections and derivatives, is needed (Nurmukhamedov, 2017). This suggests that the participants may not have been representative of higher education ESL students in Malaysia who have insufficient vocabulary knowledge (Tan & Goh, 2020) and different patterns may be found in lower proficiency learners. This may explain the high results on the pre-target words test, which revealed that the participants ($n = 56$) scored a high mean of 15.3 out of 20, with a minimum of 6 and a maximum of 20, where 54 participants (96%) knew at least half or 10 of the target words prior to watching the video. Therefore, potential vocabulary gains may have been limited due to the ceiling effect, as most of the participants in this study knew most of the target words prior to the experimental condition of watching the video. As the small number of target items may have resulted in low learning gains (Pellicer-Sánchez & Schmitt, 2010), future studies should select more target words from the video or collect data on the acquisition of words other than the target words.

Next, 7 participants scored lower in the post-target words test than the pre-test, suggesting that these participants may have guessed some of the answers on the tests. This does not reflect their ability to derive the meaning of words through comprehension and can even reinforce the learning of words with incorrect meaning, which is a difficult process to fix (Laufer, 1997). Correctly guessing words has been identified as a source of error on objective MCQ tests that are either too short or have too difficult items (Ubulom & Amini, 2012); it inflates estimates of vocabulary knowledge. This limitation is likewise for other vocabulary breath tests, including the Vocabulary Levels Test (Nation, 1983; Schmitt et al., 2001) and Yes/No test (Meara & Jones, 1990). Furthermore, participants' tendency to guess was the same when instructed not to guess by penalising wrong answers as when encouraged to answer every question despite their uncertainty, suggesting that guessing may be dependent on the participants' personalities (Waters, 1967). Thus, eradicating the guessing effect is difficult, but it may be reduced by giving participants clear instructions to avoid confusion: modifying the vocabulary tests to include an "I don't know" option (Zhang, 2013) to improve reliability and discrimination capacity or increasing the number of distractors (Stewart, 2014).

Additionally, form-meaning recognition was measured in this study to assess vocabulary learning. This provides an insight into how vocabulary can be learned through watching academic videos and suggests its effectiveness as a tool for improving L2 learners' vocabulary. However, to assess vocabulary learning more accurately (Nation & Webb, 2011), multiple factors such as captioning (Montero Perez et al., 2018), word-related variables (Nation, 2001; Peters & Webb, 2018), and working memory (Montero Perez, 2020) may be measured. This will provide further understanding of L2 learners' vocabulary acquisition and the effectiveness of academic videos as a tool for IVL.

Lastly, the 10-min academic video selected was only viewed once, with an average of 1.3 occurrences per target word. Furthermore, as only an immediate post-test was conducted, the results may not represent vocabulary learned over time or in long-term memory storage. Thus, this may not reflect authentic flipped classrooms or supplement homework tasks that students can extensively view and participate in, such as note-taking or discussions. Using authentic videos with highly informative contextual clues and relevant tasks could improve comprehension and vocabulary knowledge, as it provides opportunities for learners to use the acquired vocabulary in practise (Restrepo Ramos, 2015). Therefore, further research is necessary to be conducted using a delayed post-test to see if vocabulary is acquired for a longer period, investigate whether the process of IVL is mediated by word characteristics other than the frequency of occurrence in the video or use in general contexts, as well as complementary academic tasks to support learning such as comprehension questions or essay writing.

7 Conclusion

This study provides insight into the limited understanding of the potential effect of viewing an academic video on IVL for ESL learners in higher education. The results show that IVL occurs through watching a 10-min academic video, as participants learned approximately 0.78 words at the level of meaning-recognition. This contributes to previous findings on IVL through audio-visual input by implying that academic videos may be a beneficial tool to address problems of low vocabulary knowledge in higher education L2 learners (Tan & Goh, 2020) and a possible innovative approach to promote lexical development for successful learning in higher education (Webb, 2015). However, this study did not outline the impact of learner-related factors such as working memory, prior vocabulary knowledge, the attitude of participants, or the construct of variables and materials such as the limited number and occurrences of target words that could be significant predictors of vocabulary incidentally gained from academic videos. Further research is therefore warranted to understand IVL through viewing academic videos and explore the factors that affect the extent to which vocabulary is learned to continue to provide TESOL educators with insights into how academic videos can be integrated into classroom practices to improve vocabulary in various ESL students and develop long-lasting L2 proficiency for successful learning and academic achievements in higher education.

References

Ahmad, J. (2012). Intentional vs. incidental vocabulary learning. *International Association of Research in Foreign Language Education and Applied Linguistics ELT Research Journal*, *1*(1), 71–79. https://www.kau.edu.sa/Files/0053226/Researches/65160_36507.pdf

Ali, S. (2019). Impacts of watching videos on academic performance at university level. *European Journal of Education Studies, 6*(3), 114–125. https://doi.org/10.5281/zenodo.3244393

Alsager, R., & Milton, J. (2016). Investigating the relationship between vocabulary knowledge and academic success of Arabic undergraduate learners in SwanseaUniversity. *Language in Focus Journal, 2*(2), 88–124. https://doi.org/10.1515/lifijsal-2016-0010

Arndt, H., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning & Technology*, *22*(3), 124–142. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44660/22_03_arndt_10125-44660.pdf

Ashcroft, R., Garner, J., & Hadingham, O. (2018). Incidental vocabulary learning through watching movies. *Australian Journal of Applied Linguistics*, *1*(3), 135–147. http://doi.org/https://doi.org/10.29140/ajal.v1n3.89

Bajrami, L., & Ismaili, M. (2016). The role of video materials in EFL classrooms. *Procedia - Social and Behavioral Sciences, 232*, 502–506. https://doi.org/10.1016/j.sbspro.2016.10.068

BERA. (2011). *Ethical guidelines for educational research* (1st ed.). London: British Educational Research Association.

Bisson, M. (2013). Incidental acquisition of foreign language vocabulary through multi-modal situations. *Thesis Submitted to the University of Nottingham for the Degree of Doctor of Philosophy*. http://eprints.nottingham.ac.uk/13883/1/thesis_MJB.pdf

Brame, C. J., & Perez, K. E. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE–Life Sciences Education, 15*(4). https://doi.org/10.1187/cbe.16-03-0125

Brown, H. D. (2004). Language assessment. *Principles and Classroom Practice*. Longman.

Brown, J. D. (2014). Classical theory reliability. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1165–1181). Wiley-Blackwell.

Browne, C., Culligan, B., & Phillips, J. (2013). New Academic Word List 1.0. http://www.newgeneralservicelist.org/nawl-new-academic-word-list

Carmichael, M., Reid, A., & Karpicke, J. D. (2018). *Assessing the impact of educational video on student engagement, critical thinking and learning: The current state of play* [White Paper]. SAGE Publishing. https://us.sagepub.com/sites/default/files/hevideolearning.pdf

Chang, A. C. S. (2018). Speech rate in second language listening. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). Wiley Blackwell. https://doi.org/10.1002/9781118784235.eelt0576

Cobb, T. (2020). Range for texts v.3 [computer program]. https://www.lextutor.ca/cgi-bin/range/texts/index.pl

CrashCourse. (2014). *Social influence: Crash course psychology #38* [Video]. YouTube. https://www.youtube.com/watch?v=UGxGDdQnC1Y

Ellis, M. V. (1999). Repeated measures designs. *The Counseling Psychologist, 27*(4), 552–578. https://doi.org/10.1177/0011000099274004

Erlandsson, T., & Wallgren, S. (2017). Incidental vocabulary acquisition through reading: A literature review examining vocabulary acquisition, reading comprehension and their connection. *Linköping University Department of Culture and Communication*, 1–43. https://www.diva-portal.org/smash/get/diva2:1118558/FULLTEXT01.pdf

Evans, S., & Morrison, B. (2011). The student experience of English-medium higher education in Hong Kong. *Language and Education, 25*(2), 147–162. https://doi.org/10.1080/09500782.2011.553287

Freeman, D. (2014). Reading comprehension questions: The distribution of different types in global EFL textbooks. In N. Harwood (Ed.), *English language teaching textbooks: Content, consumption, production* (pp. 205–238). Palgrave Macmillan.

Gay, P. (2010). Multiple-choice testing as an assessment tool. *Welsh Assembly Government*. Caerphilly. https://dera.ioe.ac.uk/1864/1/101006guidanceen.pdf

Geva, E., & Wiener, J. (2015). *Psychological assessment of culturally and linguistically diverse children and adolescents: A practitioner's guide*. Springer Publishing Company.

Hadijah, S., & Pd, M. (2016). Teaching by using video: Ways to make it more meaningful in EFL classrooms. *Proceedings of the Fourth International Seminar on English Language and Teaching (ISELT-4), 4*(2), 307–315. http://ejournal.unp.ac.id/index.php/selt/article/view/6990

Harrington, M., & Roche, T. (2014). Word recognition skill and academic success across disciplines in an ELF university setting. *Papers in Language Testing and Assessment, 3*(2), 76–99. https://www.academia.edu/9856869/Harrington_M_and_Roche_T_2014_Word_recognition_skill_and_academic_success_across_disciplines_in_an_ELF_university_setting_Papers_in_Language_Testing_and_Assessment

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235–253. http://www.jstor.org/stable/40264352

Kara, S. (2013). The relationship between text comprehension and second language vocabulary acquisition: Word-focused tasks. *International Journal on New Trends in Education and their Implications, 4*(1), 39–53. http://ijonte.org/FileUpload/ks63207/File/04.kara.pdf

Kirana, M. (2016). The use of audio visual to improve listening. *English Education Journal, 7*(2), 233–245. https://core.ac.uk/download/pdf/291613903.pdf

Kirschner, P. A., & Neelen, M. (2017). Double-barrelled learning for young & old. https://3starlearningexperiences.wordpress.com/2017/05/30/double-barrelled-learning-for-young-old/

Ko, M. (2010). A comparison of reading comprehension tests: Multiple-choice vs. open-ended. *English Teaching, 65*(1), 137–159. https://doi.org/10.15858/engtea.65.1.201003.137

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Pergamon Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* Longman.

Kumar, M., & Jamil, M. (2016). Enhanced learning using motion graphics in higher education. *Conference: International Technology, Education and Development Conference.* https://doi.org/10.21125/iceri.2016.0157

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 University Press.

Lavrakas, P. J. (2008). Respondent fatigue. *Encyclopedia of Survey Research Methods.* Sage. https://doi.org/10.4135/9781412963947.n480

Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education – Is it still a matter of number right scoring or negative marking? Accepted for publication in *Studies in Educational Evaluation.* https://users.ugent.be/~mvalcke/CV/scoring.pdf

Lin, L. (2010). A video-based CALL program for proficient and less-proficient L2 learners' comprehension ability, incidental vocabulary acquisition. *Educational Media International, 47*(3), 199–216. https://doi.org/10.1080/09523987.2010.518812

Ljubojevic, M., Vaskovic, V., Stankovic, S., & Vaskovic, J. (2014). Using supplementary video in multimedia instruction as a teaching tool to increase efficiency of learning and quality of experience. *International Review of Research in Open and Distance Learning, 15*(3), 275–291. https://doi.org/10.19173/irrodl.v15i3.1825

López-Pérez, M. V., Pérez-López, M. C., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation. *Computers & Education, 56*(3), 818–826. https://doi.org/10.1016/j.compedu.2010.10.023

Mayer, R. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Meara, P., & Jones, G. (1990). *Eurocentres vocabulary size test* (version E1.1/K10,MSDOS). Zurich: Eurocentres Learning Service.

Miner, S., & Stefaniak, J. E. (2018). Learning via Video in Higher Education: An Exploration of Instructor and Student Perceptions. *Journal of University Teaching and Learning Practice, 15*(2), 1–14. https://ro.uow.edu.au/jutlp/vol15/iss2/2

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching, 56*(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition, 42*, 749–773. https://doi.org/10.1017/S0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44357/1/18_01_monteroperezetal.pdf

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*, 1–26. https://doi.org/10.1080/09588221.2017.1375960

Montgomery, A. P., Mousavi, A., Carbonaro, M., Hayward, D. V., & Dunn, W. (2019). Using learning analytics to explore self-regulated learning in flipped blended learning music teacher education. *British Journal of Educational Technology, 50*(1), 114127. https://doi.org/10.1111/bjet.12590

Nation, I. S. P. (1980). Strategies for receptive vocabulary learning. *Guidelines, 3*, 18–23. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/1980-Receptive-vocab.pdf

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*, 12–25.

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

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13. https://www.lextutor.ca/tests/nation_beglar_size_2007.pdf

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

Nguyen, C. D., & Boers, F. (2018). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly*. https://core.ac.uk/download/pdf/215386365.pdf

Nurmukhamedov, U. (2017). Lexical coverage of TED talks: Implications for vocabulary instruction. *TESOL Journal, 768–790*. https://doi.org/10.1002/tesj.323

Nwosisi, C., Ferreira, A., Rosenberg, W., & Walsh, K. (2016). A study of the flipped classroom and its effectiveness in flipping thirty percent of the course content. *International Journal of Information and Education Technology, 6*(5), 348–351. https://doi.org/10.7763/ijiet.2016.v6.712

Paivio, A. (1971). *Imagery and verbal processes.* Holt, Rinehart & Winston.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*, 31–55. http://hdl.handle.net/10125/66652

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 57*(155–199), 469–523. https://doi.org/10.1111/j.1467-9922.2007.00415.x

Rajadell, M., & Garriga-Garzón, F. (2017). Educational videos: After the why, the how. *Intangible Capital, 13*(5), 903–923. https://doi.org/10.3926/ic.1042

Restrepo Ramos, F. D. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *Profile Issues in Teachers' Professional Development*, 17(1), 157–166. https://doi.org/10.15446/profile.v17n1.43957.

Reynolds, M. R., & Turek, J. J. (2012). A dynamic developmental link between verbal comprehension-knowledge (Gc) and reading comprehension: Verbal comprehension knowledge drives positive change in reading comprehension. *Journal of School Psychology, 50*(6), 841–863. https://doi.org/10.1016/j.jsp.2012.07.002

Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in English as a foreign language setting. *Language Testing in Asia, 3*(1), 1–13. https://doi.org/10.1186/2229-0443-3-12

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Doctoral dissertation, Victoria University). Te Herenga Waka—Victoria University of Wellington. https://doi.org/10.26686/wgtn.17005078.v1

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *International Journal of Applied Linguistics, 171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Sadoski, M. (2005). A dual coding view of vocabulary learning. *Reading & Writing Quarterly, 21*(3), 221–238. https://doi.org/10.1080/10573560590949359

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

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145–171. https://doi.org/10.2307/3588328

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22*(1), 71–97. https://doi.org/10.1080/10862969009547695

Siyanova-Chanturia, A., & Webb, S. (2016). Teaching vocabulary in the EFL Context, 227–239. https://doi.org/10.1007/978-3-319-38834-2_16

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition literacy. *Reading Research Quarterly, 21*, 360–406. https://doi.org/10.1598/RRQ.21.4.1

Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST? *Language Assessment Quarterly, 11*(3), 271–282. https://doi.org/10.1080/15434303.2014.922977

Stoeckel, T., Bennett, P., & McLean, S. (2016). Is "I don't know" a viable answer choice on the vocabulary size test? *TESOL Quarterly, 50*(4), 965–975. http://www.jstor.org/stable/44984728
Szabo, C. Z., Stickler, U., & Adinolfi, L. (2020). Predicting the academic achievement of multilingual students of English through vocabulary testing. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2020.1814196
Tan, A. W. L., & Goh, L. H. (2020). Comparing the effectiveness of direct vocabulary instruction and incidental vocabulary learning in improving the academic vocabulary of Malaysian tertiary students. Social Science & Humanities*, 28*(2), 263–279. http://www.pertanika.upm.edu.my/pjssh/browse/special-issue?article=JSSH(S)-1266-20
Teng, F. (2018). Incidental vocabulary learning for primary school students: The effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher, 46*, 113–136. https://doi.org/10.1007/s13384-018-0279-6
Teng, F. (2020). Retention of new words learned incidentally from reading: Word exposure frequency, L1 marginal glosses, and their combination. *Language Teaching Research, 24*(6), 785–812. https://doi.org/10.1177/1362168819829026
Tuna, T., Tuhin, D. E. Y., Subholk, J., & Leasure, J. L. (2018). Video supported flipped classroom. *Journal of Educational Multimedia and Hypermedia, 27*(4), 529–548. http://icsvideos.cs.uh.edu/intro/VideoSupportedClassroom.pdf
Ubulom, W. J., & Amini, C. M. (2012). Determining the effect of guessing on test scores. *Mathematical Theory and Modeling, 2*(12), 16–22. https://core.ac.uk/download/pdf/234679081.pdf
van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*, 609–624. https://doi.org/10.1016/j.system.2013.07.012
Wang, P. (2009). The inter-rater reliability in scoring composition. *English Language Teaching, 2*(3), 39–43. https://files.eric.ed.gov/fulltext/EJ1083076.pdf
Waters, L. K. (1967). Effect of perceived scoring formula on some aspects of test performance. *Educational and Psychological Measurement., 27*, 1005–1010.
Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667–686. https://core.ac.uk/download/pdf/215386265.pdf
Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*, 335–366.
Wong, K. M., & Samudra, P. G. (2019). L2 vocabulary learning from educational media: Extending dual-coding theory to dual-language learners. *Computer Assisted Language Learning, 34*(8), 1182–1204. https://doi.org/10.1080/09588221.2019.1666150
Yawiloeng, R. (2020). Second language vocabulary learning from viewing video in an EFL classroom. *English Language Teaching, 13*(7). https://doi.org/10.5539/elt.v13n7p76
Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning: A Journal of Research in Language Studies, 70*(4), 1017–1053. https://doi.org/10.1111/lang.12411
Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly, 47*(4), 1–36. https://doi.org/10.1002/tesq.98
Zhong, H. (2011). Learning a word: From receptive to productive vocabulary use. *Proceedings of the Asian Conference on Language Learning, Japan*, 116–126. https://www.researchgate.net/publication/319002965_Learning_a_Word_From_Receptive_to_Productive_Vocabulary_Use
Zumbo, B. D., & Jennings, M. J. (2002). The robustness of validity and efficiency of the related samples t-test in the presence of outliers. *Psicológica, 23*, 415–450. https://www.redalyc.org/pdf/169/16923209.pdf

Jessica McLaughlin graduated from the University of Nottingham with a Bachelor of Education with Honours in TESOL and is currently working as a full-time teacher at an international school in Malaysia. Her research interests are in exploring the various factors and strategies for learning English as a second language.

Csaba Zoltan Szabo , Ph.D., holds an MA in Teaching English as Foreign Language from Swansea University, Wales and a PhD in Second Language Acquisition from The Open University, UK. He became Fellow of Higher Education Academy (Advanced HE, UK) in 2020. Currently, he is an assistant professor and the Director for Postgraduate Research Programmes in the School of Education at University of Nottingham Malaysia. He is founding and organising chair of the Alternative Research Lounge. His research interests include vocabulary acquisition and assessment in multilingual contexts, cognate recognition and production, emotional vocabulary knowledge, game-based language learning, and academic achievement in English-medium settings.

Arathi Jane Reddy is an alumni Education student from the University of Nottingham Malaysia and is a practicing Secondary English teacher. Her research interests include using multimedia to facilitate language learning.

Mangeet Kaur Khera is a graduate of Bachelor of Education with Honours in TESOL from the University of Nottingham Malaysia. She is currently teaching at an international school in Malaysia. Her research interests are mainly in exploring the social factors involved in learning English as a second language.

Nonformal Vocabulary Learning

Computer-Assisted Learning of English Formulaic Expressions from YouTube Videos

Phoebe Lin

Abstract The web-based vocabulary-learning tool *IdiomsTube* was launched in November 2018 to facilitate independent formulaic expression (FE) learning through YouTube videos. Using the app's interface, a controlled experiment was conducted to: (1) compare the effectiveness of video versus reading input for second language (L2) FE learning; (2) investigate whether *IdiomsTube's* automatically generated learning tasks (i.e. glossary, glosses, and cloze exercise) may enhance L2 FE learning; and (3) identify factors of successful computer-assisted L2 FE learning. Sixty-seven EFL undergraduate students took part in the experiment. Their knowledge gains were compared within subject along with information about their L2 vocabulary size, and in-app behaviour (e.g. time on task, number of clicks on hyperlinks to dictionaries). The results indicate that video and reading input are equally conducive to L2 FE learning. L2 FE knowledge gain was significantly higher when assisted by *IdiomsTube's* automatically generated learning tasks (i.e. glossary, glosses, and cloze exercise) than when unassisted. Finally, logistic regression results indicate that L2 vocabulary size, time on task and looking up the dictionary are factors of FE learning in a computer-assisted language learning context.

Keywords Formulaic language · Vocabulary learning · YouTube · Application design · CALL effectiveness

1 Introduction

Learners have long shown a preference for using popular media (i.e. television, movies and songs) as a means of learning English outside the classroom. Since the 1980s, researchers have explored the potential of videos as foreign language learning tools and ways of incorporating videos into classroom teaching (Lin, 2014, Lin & Siyanova, 2014; Marshall & Werndly, 2002; Meinhof, 1998; Pujola, 2002; Rodgers & Webb, 2011; Vanderplank, 1988; Webb, 2010; Webb & Rodgers, 2009).

P. Lin (✉)
Department of English and Communication, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
e-mail: phoebe.lin@polyu.edu.hk

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_12

These studies have increased our understanding of the opportunities and challenges facing independent English learning through videos. However, breakthroughs are needed to turn videos into a well-structured independent foreign language learning opportunity. Answers have yet to be found for practical questions which learners often ask (e.g. How do we find videos suitable for our proficiency levels and needs? How often should we watch videos to benefit our vocabulary growth?). Given the lack of personalised guidance on learning through videos, learners with weak independent language learning skills struggle the most because they cannot identify what can be learned from each video and how to put gained knowledge into practice. The level of support and the amount of resources for facilitating English learning through videos are minimal, unlike the support available for reading (e.g. graded readers are readily available). There are no graded reader/video series, pre- or post-activity guided questions or dedicated class time for teachers and students to watch and discuss videos. This lack of support for learning English through videos is unjustifiable.

1.1 IdiomsTube: A Mobile and Personalised Web-based App for FE Learning Through YouTube Videos

The *IdiomsTube* app was launched in November 2018 to meet the demand for personalised guidance for independent L2 vocabulary learning through videos. Its goal is to maximise the educational benefits of online videos by not only offering information and entertainment, but also giving real opportunities for mobile and independent FE learning with each viewing (Lin, 2022, see Lin, 2023, for a review of CALL tools for English teaching and learning). Following an in-depth analysis of learners' needs, the app was designed to conduct three main automatic functions: (1) it generates learning tasks for any English captioned YouTube videos; (2) it assesses the difficulty of a video's language in terms of speech rate and lexical difficulty; and (3) it recommends videos based on the learner's proficiency level. At the time of writing, the *IdiomsTube* app had over 8,000 registered users worldwide.

In its current version, *IdiomsTube* generates three types of FE learning tasks: pre-learning tasks (i.e. the glossary and the dictionary look-up), several consolidation tasks (i.e. the fill-in-the-blanks, the spelling task, the pronunciation task, and the selected scene replay), and a number of revision tasks (i.e. the flashcards, vocabulary bookmarks, exercise histories and video diaries). Figure 1 shows the workflow of learning tasks that learners complete for every video they watch.

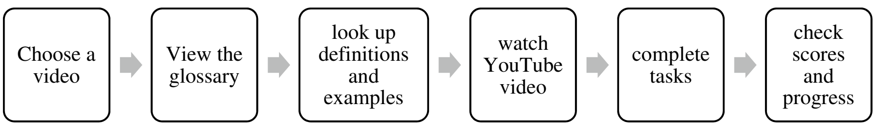

Fig.1 The workflow of learning tasks that learners need to complete on IdiomsTube

The app's design incorporates key features known to enhance the effectiveness of vocabulary learning in context. These features include automatic identification of noteworthy FEs in the English subtitles of any videos posted on YouTube, automatic assessment of the difficulty of the vocabulary, the automatic generation of multimodal concordances of system-identified vocabulary items, the automatic generation of contextual usage, formal recall and verbal repetition tasks, gamification features, a learning progress tracking system for class teachers and so on (see Lin, 2022, for details about the rationale and features of the app).

The *IdiomsTube* project is unique in several ways. First, it is the first-ever computer-assisted language learning (CALL) tool to teach English formulaic expressions (FEs) in context. Other self-learning tools either focus on the learning of single-word vocabulary items (see Nesselhauf & Tschichold, 2002, for a review) or the learning of English FEs through concordances (Chen et al., 2014). Second, the *IdiomsTube* project developed the *technology* to automatically generate various types of English learning activities for any English-captioned video on YouTube rather than to simply provide a platform for disseminating activities developed by teachers. Once developed, this automatic technology should also be applicable to other online streaming platforms. Third, *IdiomsTube* offers a system for teachers to monitor their students' learning, therefore supporting classroom-based as well as independent English learning.

This chapter reports a study recently conducted to: (1) compare the effectiveness of video versus reading input for L2 formulaic language learning; (2) investigate whether *IdiomsTube's* automatically generated learning tasks (i.e. glossary, glosses, and cloze exercise) may enhance L2 FE learning; and (3) identify factors of successful computer-assisted L2 FE learning.

2 Literature Review

2.1 The Optimal Mode of Input for L2 Vocabulary Learning

Reading has long been regarded as 'the ideal medium for L2 vocabulary learning' (Ellis, 1995, p. 106). It is reasoned that, unlike listening input, reading input allows learners to learn L2 vocabulary at their own pace. Learners have more time to process the form and meaning of new words because there is no need to worry about real-time speech segmentation (van Zeeland & Schmitt, 2013; Vidal, 2011). The superiority of reading input is supported by research evidence. In Vidal's (2011) study, 230 EFL learners acquired L2 English vocabulary through reading or listening. Between-subject comparisons of vocabulary knowledge gains indicate that the reading group significantly outperformed the listening group in both the immediate and the one-month delayed posttests. However, the edge of the reading input disappeared among learners of the highest level of L2 proficiency.

This debate over the optimal mode of input for L2 vocabulary learning reopened recently as initial new empirical evidence (Lin, 2021) emerged which suggests that spoken input may be more conducive to L2 FE learning than written input. In Lin's (2021) study, 182 university student EFL learners and 30 L1 English speakers learned novel three-word English phrases presented auditorily versus orthographically. Within-subject comparisons of immediate knowledge gains revealed that both the EFL and L1 English groups recalled significantly more phrases presented auditorily than orthographically, with the advantage of auditory input being significantly greater in magnitude and effect size in the L1 English group than the EFL group.

The conclusions by Vidal (2011) and Lin (2021) may appear contradictory at first because the former suggests that reading input is more conducive to L2 vocabulary learning while the latter suggests the opposite. However, the two studies indeed differ in two fundamental ways. First, the type and nature of vocabulary investigated are different. While Vidal (2011) focused on single-word lexical items, Lin (2021) examined 3-word expressions. As Lin (2021) suggests, word learning involves the memory of new forms and their meanings. FE learning, on the other hand, involves the memory of the syntagmatic relationship between known words. Second, learners in Vidal's (2011) study learned words embedded in passages, whereas those in Lin's (2021) study learned expressions out of context. For a robust comparison of the effect of learning FEs through videos versus reading, a controlled experiment is necessary.

2.2 The Benefits of Learning from Videos

Television and videos have long been learners' preferred strategy for self-directed language learning (Gieve & Clark, 2005). They are very popular, not only because they tend to be entertaining, but also because they offer multi-sensory stimulations. From a linguistic perspective, the availability of visual, audio and textual cues can also facilitate text comprehension and memory. Hence, it may be easier for a learner to comprehend a movie than the novel from which the movie was adapted. Time may be another concern. Learners may be keener to watch a video that they can finish in 10–15 min than to engage in other self-directed language learning activities (such as reading novels, writing to pen pals or attending classes) which require much more time and personal commitment.

There has been considerable discussion among English teaching and learning experts surrounding the value of television for learning English as a foreign language since the 1980s (Meinhof, 1998; Vanderplank, 1988, 1990, 1994). As Vanderplank (1990: 221) states, "no teacher, no classroom can provide the amount, the quality, the variety of language in interesting, meaningful, informative and often amusing contexts that television can". A survey of a one-hour episode of the British television drama *Holby City* (i.e. Series 15 Episode 23) by Lin and Siyanova (2014), for example, found that the episode contained as many as 86 dialogues covering a variety of communicative purposes (e.g. showing concern and care, offering support and advice, asking for a favour, opinion exchanges, confrontation and peace-making) and interlocutor relationships (e.g. father and son, husband and wife, boss and officer, doctor and patient). These television dialogues are often closer to natural authentic speech than the samples created by EFL textbooks or teachers (see also Al-surmi, 2012; Grant, 1996; Quaglio, 2009). As Lin and Siyanova (2014) suggest, internet videos can be an excellent resource, particularly for learning FEs used in everyday spoken English. This is not only because they capture the most frequently used FEs in everyday spoken English, but also because the frequencies at which FEs appear on internet television and in everyday spoken English are comparable (Lin, 2014). This means that extensive exposure to internet television can help learners tackle the difficulty of judging a specific FEs' frequency of use (see also Foster, 2001; Nesselhauf, 2005; Paquot, 2008).

2.3 Conditions for Independent English Learning Through Videos

Although learners are well motivated, learning English through videos remains challenging (Lin, 2022). A range of conditions need to be met before input can be converted into intake, so extensive exposure to English videos alone is insufficient (at least in the case of adult second or foreign language acquisition). First, the input needs to be comprehensible, which, according to the Input Hypothesis (recently renamed *Comprehension Hypothesis*, Krashen, 2009), means that it should be one level above that of the learner's current proficiency. If it is too difficult or too easy, learning will be hindered.

Second, learners need to have a certain level of metalinguistic and conscious awareness of the language in the video to learn from it (see Schmidt, 1990, and related discussions of the Noticing Hypothesis). As the human brain is predisposed to prioritise the decoding of meaning from auditory input rather than remembering words verbatim (due to the limited capacity of the short-term memory), this metalinguistic and conscious awareness of language may be more difficult to achieve than expected.

Third, consolidation of memory is necessary to produce a lasting learning effect. In Szudarski and Conklin's (2014) study, for example, subjects showed significant gains in FE knowledge when tested immediately after treatment. The knowledge gain had disappeared, however, in the 6-week delayed post-test. Interventions are needed to overcome this attrition (see Hansen et al., 2002; Schmitt, 1998; Weltens & Grendel, 1993).

Fourth, learners need repeated exposures and revision to secure learning. The L2 vocabulary acquisition literature (Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003) suggests that learners generally need 8–20 exposures to learn a lexical item. These repeated exposures should be spaced over time rather than massed (e.g. receiving 20 exposures in one session; see Bahrick et al., 1993; Ellis, 1995).

In the internet age, CALL technology is essential for meeting these conditions and facilitating English learning through online videos. For example, key factors of listening comprehension, including the speech rate and the lexical difficulty of a text, can be estimated automatically using information in the video caption. Such pieces of information can provide a basis for finding comprehensible input and recommending videos for any learner. Working with the captions, CALL tools can also automatically detect, gloss and generate exercises for noteworthy features of English usage to enhance learners' conscious awareness and engagement with the language of the video. With the learner's consent, his or her video preferences and performance in the exercises may be stored so that an app may reinforce the learning by repeating the exposures sufficient times through flashcards and exercises. These essential CALL elements have all been incorporated in the *IdiomsTube* app (see Lin, 2022, for details).

2.4 Key Factors Determining the Effectiveness of a CALL Vocabulary Building App

CALL has a long history of developing vocabulary building programs (Lin, 2022; Chen et al., 2014; Shei & Hsieh, 2011; see also Nesselhauf & Tschichold, 2002, for a review). These apps have tremendous potential in CALL, because, as Nesselhauf and Tschichold (2002: 251) say:

> vocabulary is one of the fields that can relatively easily be practised outside the classroom, especially when compared to pronunciation or dialogue practice, where feedback from a qualified teacher is very valuable and can often not be given adequately by a CALL program.

While reviews of vocabulary building apps typically focus on their design and features, the success of such apps is equally dependent on the user's personal characteristics (e.g. L2 proficiency, L2 vocabulary size, and learning styles and preferences) and in-app behaviour. CALL apps typically provide many options and features, such as the default language of the glosses, optional display of dictionary definitions or concordances of vocabulary items, optional links to the external webpage offering cultural notes and so on. It is important to investigate which of these app options and features are most useful for learners.

Three key factors affecting the success of a CALL app, namely the app's design, the learner's personal characteristics and the learner's in-app behaviour, have been individually studied in the context of gloss use in CALL reading comprehension apps. Yoshii (2006), Yanguas (2009), and Poole (2012) explored the effects of different gloss designs. Lee et al. (2018) explored the effect of learners' L2 proficiency as a learner characteristic, finding that higher proficiency L2 learners seemed to benefit more from certain types of computer-generated aids than lower proficiency learners. Leveridge and Yang (2013), Plass et al. (1998), and Pujola (2002) explored the relationship between the cited learning outcomes and learners' in-app behaviour, particularly the frequency of use of specific app features. Unfortunately, in the context of vocabulary building apps, very few studies, if any, have been conducted to examine the extent to which the effectiveness of an app is affected by learner characteristics or learners' in-app activities.

3 The study

This study adopted a pretest-treatment-posttest research design with within-subject comparison. It aimed to address three research questions:

1. Which input mode (video or reading) is more conducive to L2 FE learning?
2. Can L2 FE learning from videos be enhanced by auto-generated awareness-raising tasks (i.e. glossary, dictionary lookup and fill-in-the-blanks exercise)?
3. Which factors explain the success of computer-assisted learning of L2 FEs from videos?

3.1 Subjects

One hundred and forty four undergraduates from an English-as-a-medium-of-instruction university in Hong Kong, who were L1 Cantonese speakers, born and educated in Hong Kong, and who had never studied abroad, were recruited through a campus-wide open call-for-participation. They took a 20-item multiple-choice pretest (see Sect. 4.4.1) and filled in a background questionnaire. Only those who knew two or fewer of the 10 experimental stimuli in the pretest were invited to take part in the treatment a week later. The number of subjects who completed the study was 67 (51 female and 16 male), with a mean age of 21.3 (age range: 19–25). These subjects provided their HKDSE English language paper scores.[1]

3.2 Stimuli and Materials

The stimuli were 10 FEs used spontaneously in two authentic YouTube videos*:* *"Why can't I poop when I travel"* produced by the popular science channel *SciShow* (hereafter *Text 1*) and "*The most devious cheat in Olympic Games history?*" produced by the *Olympic Channel* (hereafter *Text 2*). The videos were chosen from a corpus of 20,000 YouTube videos compiled especially for this research. This representative and balanced corpus was built by sampling the latest 1,000 videos from each of the 21 YouTube video categories (e.g. *science & technology*, *news & politics*, *travel & events*). While Texts 1 and 2 were used directly as input for the video input condition,

[1] HKDSE scores were benchmarked against IELTS scores (see https://www.hkeaa.edu.hk/DocLibrary/MainNews/press_20130430_eng.pdf). A 4 in the HKDSE English paper for example, is equivalent to an IELTS score of 6.31–6.51. Since sitting the IELTS is not a requirement for graduation in Hong Kong, few subjects were able to provide their IELTS scores.

they were adapted into reading passages for the reading input condition. All full texts are available from https://tinyurl.com/35x3aje6.

Four criteria were applied that led to the selection of these two texts from among 20,000 videos: (1) the words in the videos should be easy to ensure text comprehension; (2) the videos should each contain 5 FEs that Hong Kong university students are unlikely to know; (3) the video length should be under 4 min (to keep the data collection session below an hour); and (4) the topics of the videos should be accessible to the wider audience so that specialist subject knowledge is not required to comprehend the texts.

The FEs used spontaneously in the texts varied in their composition, types and length. This variety reflects the actual nature of YouTube videos as authentic input for L2 vocabulary learning. Since ecological validity was considered crucial to this study, no attempt was made to modify the input or the stimuli.

3.3 Tests

3.3.1 Test of L2 FE Knowledge

The subjects' knowledge of L2 English FEs was tested before and after treatment. A multiple-choice test assessed subjects' recognition of the meaning of 20 FEs, half of which appeared in Texts 1 and 2 while the other half were foil items (see Appendix for the test paper). Each target FE was presented out-of-context and in bold. The subject's task was to assess which meaning, of the 5 choices provided, applied to the target FE. To discourage subjects from guessing the answer, three measures were taken: (1) the task instruction warned against guessing; (2) subjects were given "I don't know" as the fifth option in each question; and (3) subjects were instructed that they would lose 1 mark for every incorrect guess. An example of the test items is as follows:

To "fall on deaf ears" means:

a) no one answers one's phone call
b) no one accepts one's offer of help
c) no one listens to one's suggestions or requests
d) no one believes in a story that one tells
e) I don't know

3.3.2 Vocabulary Size Test (VST)

Nation and Beglar's (2007) 140-item Vocabulary Size Test (VST) was used to assess subjects' vocabulary size. The VST is a validated measure of receptive vocabulary size (Beglar, 2010). In this study, subjects were given the Chinese version of the

VST, translated by the author for Cantonese learners of English from Hong Kong.[2] They began with the first 1,000-word band and progressed from the second to the highest (14th) 1,000-word band. A subject's receptive vocabulary size is the number of correct answers out of 140. Each target vocabulary item was presented in a non-defining context and bolded. The subject's task was to choose which of the 4 meanings provided, related to the target word. An example of test item is as follows:

SOLILOQUY: That was an excellent soliloquy!

a) 六重唱曲目
b) 格言
c) 表演
d) 自白

3.4 Procedure

To compare the effectiveness of video versus reading input for L2 FE acquisition, gains in knowledge of the stimuli were compared *within subject*. Assisted by a version of the *IdiomsTube* interface customised for this experimental setup, 54 subjects learned FEs through both video and reading input using Texts 1 or 2. To control for the effects of texts and order of presentation on FE learning, the texts used in the video/reading conditions and the order in which the conditions were presented were all counterbalanced. In other words, the experimental group subjects were randomly assigned to learn FEs in any one of four possible orders (see Table 1 for an outline of the data collection process):

A) read text 1 then view the video of text 2;
B) view video of text 2 then read text 1;
C) view video of text 1 then read text 2; and
D) read text 2 then view the video of text 1.

In the one-hour data collection session, subjects were individually seated in front of a desktop computer equipped with headphones in the computer laboratory and told to follow all on-screen instructions (see Figs. 2–5). Everyone had 4 min to watch the English-captioned video or read the text, but they were not told about the need to complete learning tasks after viewing or reading. To facilitate engagement with the input material, subjects watched the video or read the text twice and had to write a three-bullet point summary after the first viewing or reading. A survey about their viewing or reading experience was launched post treatment, followed by the posttest, and an English learning habit survey.

---

[2] The Chinese version of the VST adapted for the Cantonese EFL learners in this study is available from https://tinyurl.com/35x3aje6.

Table 1 An outline of the data collection procedure

	Experimental				**Control**
	Order A	Order B	Order C	Order D	
	n = 14	n = 13	n = 14	n = 13	n = 13
Pretest					
Part 1	Glossary + dictionary lookup				-
	Text 1 passage	Text 2 video	Text 1 video	Text 2 passage	Text 1 video
	Summary task				
	Text 1 passage	Text 2 video	Text 1 video	Text 2 passage	Text 1 video
	Cloze exercise				Self-reflection
	End-of-part survey				
Break					
Part 2	Glossary + dictionary lookup				-
	Text 2 video	Text 1 passage	Text 2 passage	Text 1 video	Text 2 video
	Summary task				
	Text 2 video	Text 1 passage	Text 2 passage	Text 1 video	Text 2 video
	Cloze exercise				Self-reflection
	End-of-part survey				
Posttest					
Learning habit survey					

ANOVAs were conducted to test if subjects in the experimental and control groups were comparable in term of their L2 English proficiency. The test results confirmed that there was no significant difference in terms of the subjects' reading, writing, speaking, listening and overall HKDSE English test scores (see Table 2, all *p*-values are non-significant).
Note The 5 FE stimuli in Text 1 are: it's happened to the best of us, loading up on caffeine, a perfect recipe for blockage, creep to a halt, out of whack. The 5 FE stimuli in Text 2 are: on the wane, away from prying eyes, smelled a rat, fell on deaf ears, a dressing-down.

4 Findings

4.1 Which Input Mode (Video or Reading) is More Conducive to L2 FE Learning?

To answer this question, knowledge of the stimuli was measured using the 10-item FE knowledge test, which was found to be reliable (Cronbach's $\alpha = 0.82$). Gains were then calculated separately for FEs learned from video versus reading input (see

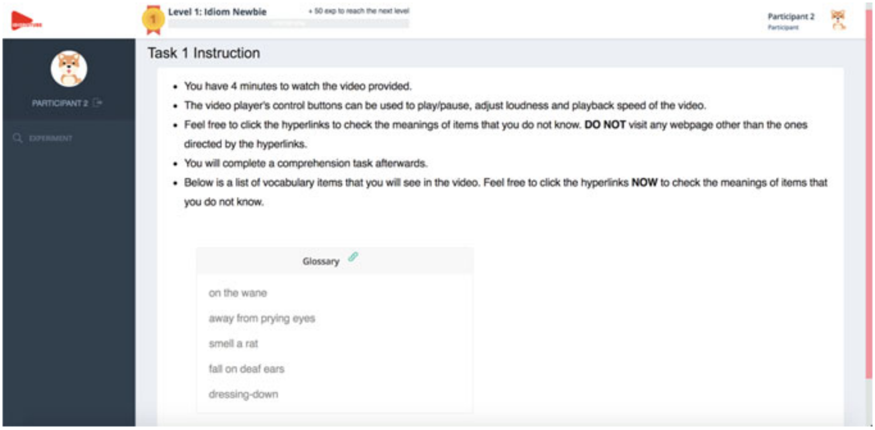

Fig. 2 On-screen instructions for subjects before the video task

Table 2 Subjects' HKDSE English test score means and standard deviations

	Experimental group				**Control group**	*F*	*df*	*p*
	Order A	Order B	Order C	Order D				
Listening	3.71 (1.27)	3.77 (1.17)	3.5 (0.94)	3.46 (0.97)	3.85 (1.46)	0.28	4.00	0.89
Speaking	3.86 (1.35)	4.38 (0.87)	3.86 (0.95)	3.77 (1.01)	4.08 (0.86)	0.88	4.00	0.49
Reading	3.79 (1.42)	4.23 (1.3)	3.93 (1.14)	4.08 (1.32)	4.38 (1.76)	0.32	4.00	0.86
Writing	3.5 (1.02)	4.00 (1.0)	3.36 (0.5)	3.62 (0.96)	3.62 (0.77)	1.13	4.00	0.36
Overall	3.79 (1.05)	4.00 (1.0)	3.64 (0.5)	3.62 (0.87)	3.92 (0.95)	0.51	4.00	0.73

Fig. 6) and analysed using a paired-samples *t*-test.[3] The results showed that video input (mean gain $= 3.30$, $sd = 1.49$, $n = 54$) generated greater gains than reading input (mean gain $= 3$, $sd = 1.66$, $n = 54$).[4] However, the difference between the two conditions was not statistically significant ($t = 1.28$, $df = 53$, $p = 0.21$, *Cohen's d* $= 0.18$).

[3] The paired-samples *t*-test was used because all experimental subjects underwent both the video and the reading conditions (in a counter-balanced manner). Therefore, their performance under both conditions was compared within-subject.

[4] Note that the full score in this study is always 5. This is because each subject was presented 5 FEs under each condition.

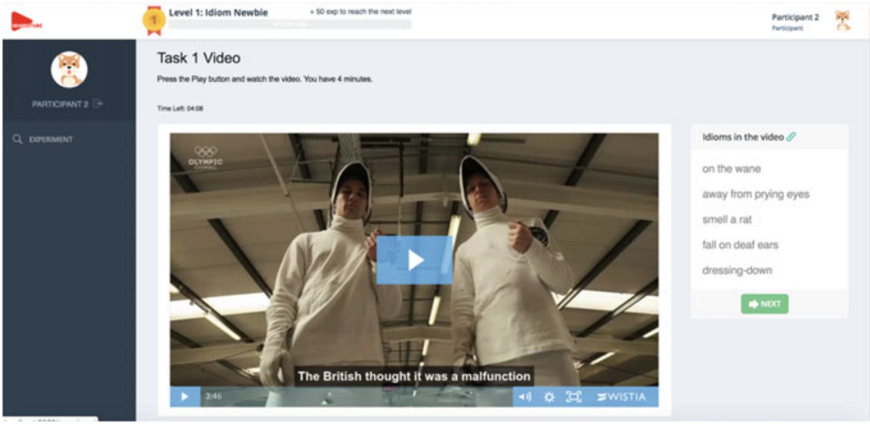

Fig. 3 IdiomsTube customised interface: the video page. The page showed a glossary of noteworthy FEs and hyperlinks to corresponding dictionary entries

Fig. 4 IdiomsTube customised interface: the reading page. The page showed a glossary of noteworthy FEs and hyperlinks to corresponding dictionary entries

4.2 *Can L2 FE Learning from Videos be Enhanced by Auto-generated Awareness-raising Tasks?*

To enhance FE learning from videos, the app *IdiomsTube* auto-generates awareness-raising tasks. A glossary with hyperlinks to online dictionary entries was presented both pre-video (see Fig. 2) and during video (see Fig. 3). The cloze (see Fig. 5), pronunciation and spelling exercises were administered post-video. The rationale and details about these functions are provided in Lin (2022). All these auto-generated

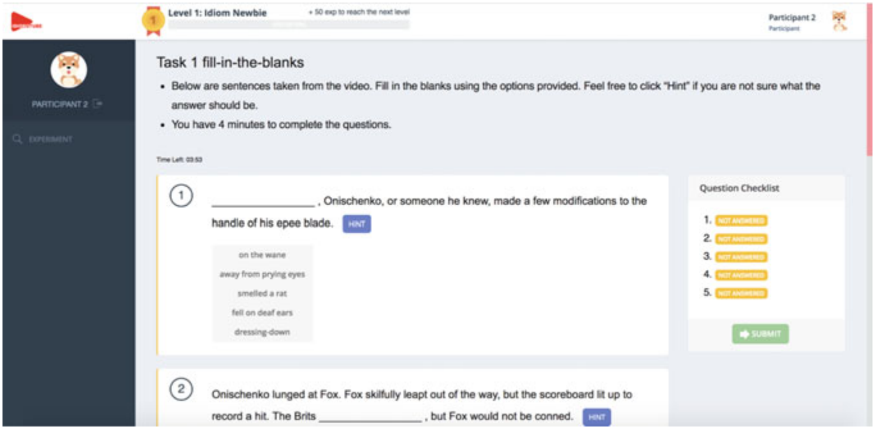

Fig. 5 IdiomsTube customised interface: automatically-generated cloze exercise. Learners could click the Hint button to return to the exact scene in which a noteworthy FE was used. The multiple-choice questions were auto-marked

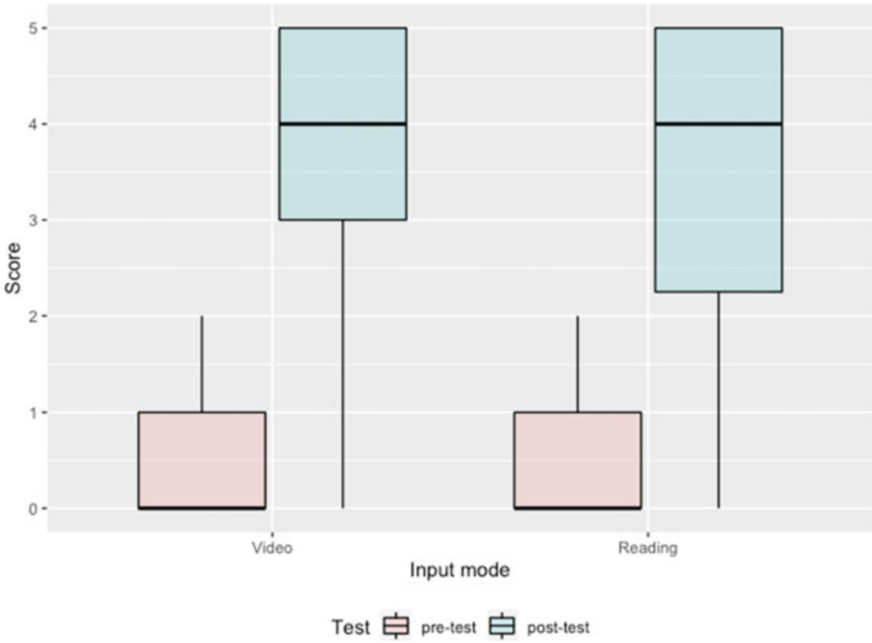

Fig. 6 A boxplot showing changes in knowledge of the target items before and after viewing the video/reading

tasks except the post-video pronunciation and spelling exercises were included in this controlled experiment.

To answer the research question, 13 subjects formed the control group and were required to learn FEs from videos (Texts 1 and 2) independently *without* access to the awareness-raising tasks. Like the 54 subjects in the experimental group, the control group subjects watched the videos twice. After the first viewing, they were given three minutes to write down a three-bullet point summary. After each video, they responded to questions about their viewing experience.

The results indicated that this group, whose learning from videos was unassisted, had much smaller gains in FE knowledge (mean = 1.15, $n = 13$) than the rest of the subjects whose learning from videos was assisted by the three awareness-raising tasks (mean = 3.30, $n = 54$). An independent-samples t-test confirmed that this difference had statistical significance ($t = 4.64$, $p < 0.001$). The effect size (*Cohen's* $d = 2.38$) was large.

While noting the uneven sample sizes of the ± tasks groups, these results clearly show the positive impact of the awareness-raising tasks. Learners who completed the glossary, the dictionary lookup and the cloze tasks learned nearly three times as many FEs as learners who had no access to the tasks.

4.3 Which Factors Explain the Success of Computer-assisted Learning of L2 FEs from Videos?

To further isolate factors that may explain the success of L2 FE acquisition from video-viewing, a logistic regression analysis was conducted with the posttest results (i.e. correct/incorrect) as the binary outcome variable and the vocabulary size test results, records of subjects' in-app behaviour, and the subjects' survey responses as predictor variables. These in-app behaviours include whether and which hyperlinks to online dictionary entries were clicked, time spent on reading the passages/viewing the videos, time spent on writing the summaries following reading/viewing, total word count of the summaries following reading/viewing, time spent completing the cloze exercise following reading/viewing, whether and which Hint buttons were clicked during the cloze exercise following reading/viewing. The survey collected subjects' responses to questions including:

- the extent to which they enjoyed the reading passage/video
- the extent to which they enjoyed each of the awareness-raising activity
- between video and reading, which one they prefer for learning English idioms, helped them learn idioms more effectively, they do more often in their free time, and choose for improving their English proficiency in the future
- the extent to which they enjoyed each of the awareness-raising activity and found each difficult
- the extent to which each awareness-raising activity facilitated their comprehension of the video/reading passage, facilitated their learning of English idioms.

Table. 3 The best-fitting logistic regression model

	Coefficients	S.E	Wald Z	p
Intercept	−0.90	0.49	−1.83	0.0672
withExercise = Yes	1.87	0.42	4.5	< 0.0001
posttestTime	−0.51	0.14	−3.68	0.0002
presentationOrder = second	0.90	0.28	3.21	0.0013
dictionaryLookup = No	−1.12	0.38	−2.98	0.0029
VST	0.32	0.14	2.26	0.0238

Observations concerning knowledge gained from reading and FEs that subjects knew in the pretest were excluded from the logistic regression analysis. Therefore, 359 observations from all 67 subjects were entered into the logistic regression model, 201 of which (56%) were correct answers in the posttest and 158 of which (44%) were incorrect answers in the posttest. Using the R package *rms* (version 5.1–3.1, Harrell, 2019), the best-fitting model was identified (see Table 3).

The model was found to be statistically significant ($\chi^2 = 150.47$, $df = 5$, $p < 0.0001$) and have predictive capacity (C = 0.85, Somers' $D_{xy} = 0.70$, $R^2 = 0.46$). Following Baayen (2008), bootstrap validation was also conducted on the model. The fast backwards elimination algorithm reported that all predictors were retained.

The model indicates three factors that have positive effects on success in computer-assisted FE acquisition from videos, including receiving guidance from automatically generated learning activities (withExercise = Yes), viewing the video after reading (presentationOrder = second) and scaled vocabulary size test scores (VST). It also indicates two factors with negative effects on success, including the time spent completing the posttest (posttestTime) and not looking up the meaning of an FE using *IdiomsTube's* built-in dictionary hyperlinks (dictionaryLookup = No).

Particularly noteworthy is the odds ratio of 1.87 for withExercise = Yes (see Table 3), which indicates that the odds for successful FE learning when assisted by the exercises are 87% higher than the odds for successful learning when unassisted by the exercises. This lends further evidence to support the overall facilitative effect of the automatically generated exercises for FE learning from videos.

5 Discussion

This study was conducted to investigate the effectiveness of video versus reading input for facilitating FE learning, to compare gains assisted versus unassisted by auto-generated awareness-raising tasks, and to explore how the three factors (app design, learners' personal characteristics and learners' in-app behaviour) interact to affect the success of computer-assisted FE learning from videos.

The results above confirmed that video and reading input are equally effective for facilitating contextualised FE learning. Reading was once regarded as the ideal medium for L2 vocabulary learning (Ellis, 1995). This argument was made based on empirical research evidence involving L2 word learning. The present study, however, has shown that video-viewing is as effective for FE learning as reading.

The study also found that learners' mean test scores nearly tripled when assisted by the auto-generated learning tasks. This facilitative effect of the exercises is also evidenced by the logistic regression results. The likelihood of successful FE learning when assisted by the exercises was 87% higher than when unassisted by the exercises.

Finally, the logistic regression model revealed the factors that explain learners' success in FE learning. As predicted, of high significance within this success are factors arising from the app's design (i.e. whether awareness-raising tasks were done), factors concerning learners' personal characteristics (i.e. learners' vocabulary size), and factors arising from learners' in-app behaviour (i.e. not clicking the hyperlinks to online dictionary entries).

5.1 *Optimal Mode of Input for L2 FE Acquisition*

Interest in the effect of mode of input on L2 FE acquisition has continued to grow in recent years. Lin (2021), for example, investigated incidental learning of novel binomials presented out of context. She found that both native speakers and learners of English were significantly better at recalling items learned through aural as opposed to visual input. This finding matches the prediction of Lin (2012) who argued that aurally presented FEs will be learned more readily because of their prosodic salience. Lin (2021) added further explanations for the finding from working memory and evolutionary perspectives. Webb and Chang (2020), on the other hand, compared incidental learning of L2 collocations through reading, listening, and reading-while-listening. They found that reading-while-listening to graded readers generated significantly greater gains in collocation knowledge than reading alone. That reading-while-listening generates significantly greater gains in collocation knowledge than reading alone was also observed in a recent longitudinal study (Vu & Peters, 2021) which investigated incidental learning of English collocations from graded readers. These three studies indicate that the availability of auditory stimulation facilitated learners' memory of the meaning and form of the items.

This study, however, provides an alternative perspective on the question about the mode of input effect on L2 FE acquisition: when awareness-raising tasks are provided to increase the salience of FEs, L2 English learners can learn FEs equally well through reading and videos. This finding is surprising. Given the fact that videos provided multimodal input (i.e., textual, sound and images) and auditory stimulation is known to facilitate learning (see Lin, 2012, 2021; Webb & Chang, 2020), videos should be more conducive to FE learning than reading. That it is not the case in the present study may indicate that the advantage of the awareness-raising tasks was so much greater than the effect of input mode that the latter became overshadowed. In other words, participating in the tasks (particularly the dictionary lookup) facilitated

FE learning much more readily than hearing the sounds or seeing the images in the video could be rendered relatively less important.

5.2 *Automatising Support for L2 FE Acquisition from YouTube Videos*

This study's findings have far-reaching implications for frontline teachers and learners. First, for decades, learners have expressed a clear preference for learning English through videos. This is understandable because videos offer multisensory stimulation and are entertaining. In the age of social media and online streaming, videos are the trend. This preference, however, often seems to be ignored, with many Asian EFL syllabi demonstrating considerable bias toward written input both within the classroom and for home study (Lin, 2012), while scant and passive support is allocated to video-viewing. This study has provided much needed empirical evidence to advocate education reforms so this serious bias related to input mode may be addressed.

Second, this study has shown and validated an innovative means of delivering large scale, low cost support to independent learners wanting to expand their English vocabulary through watching online videos. The technology now exists to generate learning tasks for any English captioned YouTube video, assess the difficulty of the language, and make recommendations based on the proficiency levels of the learners (see Lin, 2022). It should also be transferable to other online streaming platforms to further expand learners' choices. This opens a new chapter in English learning where unprecedented support can be provided to independent learning from online videos. This should be great news for teachers and learners alike. Learners will now have access to an unlimited choice of online videos and personalised exercises as they learn English independently outside the classroom. Since the videos are graded in terms of difficulty, and learners accumulate experience points and badges as they continue, learners can now see a clear path of progression to assist them in managing their independent learning outside the classroom.

Teachers may also find their workload eased. Traditionally, teachers have had to spend substantial time producing worksheets with activities and questions to guide learning from videos. Considering the astonishing rate at which videos are uploaded to the internet and the diversity of learners' video preferences, it has been impossible for teachers to prepare worksheets quickly enough to meet the huge demand for guidance from learners.[5] Furthermore, class teachers now have unprecedented access to information about the learning progress and the video preferences of their

[5] It has been suggested that every minute 400 h of new content is uploaded to YouTube alone (Wojcicki, 2015).

students through the *IdiomsTube* teacher's interface.[6] With such precious information, educational policy makers, teachers and material developers can now update the English syllabi and study materials accordingly to further enhance the appeal of their resources as well as to increase learners' motivation to learn.

5.3 A More Subtle Approach to Evaluating the Effectiveness of CALL Apps

This chapter has demonstrated that the success of a CALL app is seen as the combined result of the app's design, learners' characteristics and their in-app behaviour, not just a quality of the app's design. Evidence for the interaction between these three factors and how they combine to affect this success in computer-assisted language learning is clear from the logistic regression results.

This renewed concept of app effectiveness as the combined result of the three factors is important because it may prompt CALL app developers to consider who the app's target learners are and how the app should be used to produce maximal learning outcomes. In the past, developers' sole focus was often on building the main functions into a CALL app. In the case of an app for FE learning from videos, for example, the main functions may include, finding FEs in the caption and working with the official API. However, the three-pronged model of CALL app effectiveness laid out in this chapter will hopefully draw developers' attention to the importance of learners' personal characteristics (such as their vocabulary size, L2 proficiency and learning styles) and their in-app actions in leading to successful learning. For example, the logistic regression results mentioned above demonstrated that dictionary-lookup was a significant factor in successful FE learning. The odds for learning a FE were 12% lower if learners did not click the hyperlinks to lookup the dictionary. Based on this observation, app developers should introduce measures that encourage learners to click hyperlinks to lookup FEs in a dictionary, such as doubling the experience points offered for every click and/or rewarding every minute spent on reading the dictionary pages. In *IdiomsTube*, substantial effort has been expended to build in various gamification features (such as experience points, a progress bar, level titles and badges and leader boards) as well as interaction features (such as video recommendations from teachers and other users, and the teacher's interface) so that learners' in-app behaviour can be tracked and promptly rewarded. Positively motivated by the prospect of moving up the leader board by points, it is believed that learners may click around the app more often and engage more deeply with the learning tasks.

[6] Further elaboration of the Teacher's interface's functions are accessible at https://www.youtube.com/playlist?list=PLCQbU2C4L6AaigyxyGAyF4ryZgYuoVEEi.

6 Conclusion

In the age of the internet, YouTube offers learners unlimited access to authentic spoken input at the click of a button. They can learn anytime and anywhere from videos on any genres of their choice (e.g. TEDtalks, documentaries, vlogs, music videos). This high level of flexibility and freedom explains why the concept of exploiting YouTube as an L2 English learning resource appeals greatly to today's young learners.

This chapter has presented the world's first computer-assisted FE learning tool *IdiomsTube* and reported on the findings of an experiment conducted to investigate the effectiveness of the tool for facilitating formulaic language learning through viewing YouTube videos. The results show that meaningful and effective FE learning tasks can be automatically generated using modern natural language processing technology and L2 vocabulary acquisition research findings. Importantly, given the same level of automated guidance, video-viewing can produce as much FE knowledge gain as reading.

Acknowledgements This project was funded by the Hong Kong Research Grants Council (Ref numbers: 15602017 and 15600121).

Appendix 1 Test of FE knowledge

Instruction: You will be given 20 multiple-choice questions. Answer the questions based on what you know.

For the sake of accuracy of the data, it is important that you:

- DO NOT try to find answers on the internet
- DO NOT GUESS the answer. If you do not know, just choose "I don't know". You will lose 1 mark for every wrong guess. (Please please resist the temptation to make guesses)

The task should take 5 min. If you cannot spare 5 min now to finish the task in one go, please come back later today.

1. To **"have a champagne taste on a beer budget"** means to:

   A. buy fake luxury items
   B. consume expired food or beverages
   C. like expensive items that one cannot afford
   D. get expensive things with a discount
   E. *I don't know.*

2. **"dressing-down"** is an act of:

   A. speaking angrily to someone because s/he has done wrong
   B. disclosing one's sexual preferences in front of other people
   C. getting changed into casual clothes
   D. relaxing oneself after a day's hard work
   E. *I don't know.*

3. To **"talk the talk, walk the walk"** means to:

   A. be unable to reach an agreement
   B. order someone to follow a set of strict rules
   C. act in a way that matches with the things that you say
   D. take action despite difficulties
   E. *I don't know.*

4. To be **"out of whack"** means something is:

   A. not enough
   B. not important
   C. not worth a lot of money
   D not working properly
   E. *I don't know.*

5. To **"press someone's buttons"** means to:

   A. seek someone's attention
   B. do things that trigger emotional reactions in other people
   C. attack someone's weaknesses
   D. force other people to agreement
   E. *I don't know.*

6. To be **"run-of-the-mill"** means to:

   A. be ordinary
   B. be easily scared
   C. be motivated
   D. be able to react quickly
   E. *I don't know.*

7. To **"fall on deaf ears"** means:

   A. no one answers one's phone call
   B. no one accepts one's offer of help
   C. no one listens to one's suggestions or requests
   D. no one believes in a story that one tells
   E. *I don't know.*

8. To **“smell a rat”** means to:

   A. find new evidence about a mystery
   B. discover hidden secrets
   C. be conscious of risks and danger
   D. suspect that something dishonest is happening
   E. *I don’t know.*

9. To **“load up on <u>something</u>”** means to:

   A. make something happen
   B. gather or store a lot of something
   C. recall a past memory
   D. keep something in mind
   E. *I don’t know.*

10. To **“creep to a halt”** means:

    A. someone is so nervous that s/he feels sick
    B. someone is so exhausted that s/he falls asleep
    C. something is moving slowly until it stops
    D. something is heated up until it boils
    E. *I don’t know.*

11. To be **“full of hot air”** means to:

    A. talk quickly without stopping
    B. talk a lot without saying anything of value
    C. gossip a lot about other people
    D. speak loudly and angrily about unpleasant experiences
    E. *I don’t know.*

12. It is **“a recipe for <u>something</u>”** means it is:

    A. a situation or method that is likely to result in something
    B. a solution that should be able to solve a complicated issue
    C. a set of actions that can be taken to deal with a problem
    D. a standard and official way of doing something
    E. *I don’t know.*

13. To **“take the bull by the horns”** means to:

    A. earn a lot of money in a short period of time
    B. criticise other people
    C. lead a group of people effectively
    D. deal confidently with a difficult problem
    E. *I don’t know.*

14. To **"have a chip on your shoulder"** means you feel:

    A. angry because you think you have been treated unfairly
    B. like you have forgotten to do something
    C. no pressure even in difficult situations
    D. anxious about speaking in front of other people
    E. *I don't know.*

15. We say **"it happens to the best of us"** when:

    A. someone experiences an inevitable change in life
    B. someone experiences something bad or unlucky
    C. someone faces a risky situation
    D. someone feels very tired after a long day
    E. *I don't know.*

16. To be **"away from prying eyes"** means to:

    A. avoid attention
    B. be free from responsibilities
    C. run away from challenges or criticisms
    D. run away from prison, hospitals or similar institutions
    E. *I don't know.*

17. To **"see someone eye to eye"** means to:

    A. have similar hobbies
    B. be honest with each other
    C. agree with each other
    D. be in a romantic relationship with someone
    E. *I don't know.*

18. To **"have the last laugh"** means to:

    A. live a life without any regrets
    B. force a fake laugh at someone's joke
    C. think back to one's happiest memories
    D. succeed after being doubted by other people
    E. *I don't know.*

19. To be **"on the wane"** means to:

    A. remain unchanged
    B. rise and fall rapidly
    C. become stronger or more
    .D. become weaker or less
    E. *I don't know.*

20. We say **"don't sweat the small stuff"** to encourage people not to:

    A. always put other people's needs first
    B. worry about unimportant things
    C. be careless at work
    D. fight over limited resources
    E. *I don't know.*

References

Al-surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL QuArterly, 46*(4), 671–694.

Baayen, H. R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S. & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science.*

Beglar, D. (2010). A rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101–118. https://doi.org/10.1177/0265532209340194

Chen, M. -H., Huang, C. -C., Huang, S. -T., Chang, J. S., & Liou, H. -C. (2014). An automatic reference aid for improving EFL learners' formulaic expressions in productive language use. *IEEE Transactions on Learning Technologies, 7*(1), 57–68.

Ellis, N. C. (1995). Psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*(2), 103–112.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75–94). Longman.

Gieve, S., & Clark, R. (2005). The Chinese approach to learning: Cultural trait or situated response? the case of a self-directed learning programme. *System, 33*(2), 261–276.

Grant, L. E. (1996). Teaching conversation using a television soap. *Prospect, 11*(3), 60–71.

Hansen, L., Umeda, Y., & McKinney, M. (2002). Savings in the relearning of second language vocabulary: The effects of time and proficiency. *Language Learning, 52*(4), 653–678.

Harrell, F. E. Jr. (2019). Rms: Regression modeling strategies. *R Package Version 5*, 5.1–3.1. https://cran.rhub.io/web/packages/rms/rms.pdf

Krashen, S. (2009). The Comprehension Hypothesis Extended. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 81–94). Multilingual Matters.

Lee, H., Warschauer, M., & Lee, J. H. (2018). Advancing CALL research via data-mining techniques: Unearthing hidden groups of learners in a corpus-based L2 vocabulary learning experiment. *ReCALL, 31*(2), 1–15.

Leveridge, A. N., & Yang, J. C. (2013). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL, 25*(2), 199–214.

Lin, P. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics, 33*(3), 342–347.

Lin, P. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System, 42*(1), 164–176.

Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly, 55*(3), 1011–1023.

Lin, P. (2022). Developing an intelligent tool for computer-assisted formulaic language learning from YouTube videos. *ReCALL, 34*(2), 185–200.

Lin, P. (2023). Why corpus linguistics matters to L2 teaching and learning. In K. Sands, M. J. Petray, G. D. Clements, & L. Santelmann (Eds.), *Linguistic foundations for second language teaching and learning* (pp. 1–15). Cambridge University Press.

Lin, P., & Siyanova-Chanturia, A. (2014). Internet television for L2 vocabulary learning. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 149–158). Routledge.

Marshall, J., & Werndly, A. (2002). *The language of television.* Routledge.

Meinhof, U. H. (1998). *Language learning in the age of satellite television.* Oxford University Press.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* John Benjamins.

Nesselhauf, N., & Tschichold, C. (2002). Collocations in CALL: An investigation of vocabulary-building software for EFL. *Computer Assisted Language Learning, 15*(3), 251–279.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101–119). John Benjamins.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1–28.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25–36.

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal, 29*(4), 679.

Pujola, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL, 14*(2), 235–262.

Quaglio, P. (2009). *Television dialogue: the sitcom Friends vs. natural conversation.* Amsterdam: John Benjamins.

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly, 45*(4), 689–717.

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

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 281–317.

Shei, C., & Hsieh, H.-P. (2011). Linkit: A CALL system for learning Chinese characters, words, and phrases. *Computer Assisted Language Learning, 25*(4), 319–338.

Szudarski, P., & Conklin, K. (2014). Short- and long-term effects of rote rehearsal on ESL learners' processing of L2 collocations. *TESOL Quarterly, 48*(4), 833–842.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal, 42*(4), 272–281.

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System, 18*(2), 221–234.

Vanderplank, R. (1994). Resolving inherent conflicts- autonomous language learning from popular broadcast television. In H. Jung & R. Vanderplank (Eds.), *Barriers and bridges: media technology in language learning: proceedings of the 1993 CETaLL symposium on the occasion of the 10th AILA World Congress in Amsterdam* (pp. 119–134). Frankfurt am Main: Peter Lang.

Vu, D. V., & Peters, E. (2021). incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, 1–23.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 1–27.

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics, 15*(4), 497–519.
Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366.
Webb, S., & Chang, A. C. S. (2020). How does mode of input affect the incidental learning of collocations? *Studies in second language acquisition*, First View, 1–22.
Weltens, B., & Grendel, M. (1993). Attrition of vocabulary knowledge. In R. Schreuder & B. Weltens (Eds.), *The Bilingual lexicon*. John Benjamins.
Wojcicki, S. (2015). *Keynote speech at VidCon 2015 conference*, Anaheim Convention Center, California, July 23–25, 2015.
Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning and Technology, 13*(2), 48–67.
Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language, Learning & Technology, 10*(3), 85.
Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624.
Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258.

Phoebe Lin is Assistant Professor at the Department of English and Communication, The Hong Kong Polytechnic University. Her research investigates aspects of vocabulary learning using a range of methods, including controlled experiments, corpus analysis, statistical modelling, and software development. Her latest project is IdiomsTube, an intelligent app that guides users to learn English idioms through watching any subtitled YouTube videos they like. She is also the author of paperback monograph The Prosody of formulaic sequences: A corpus and discourse approach (Continuum/Bloomsbury Academic).

Spoken Word Form Recognition with a Mobile Application: Comparing Azerbaijani and Japanese Learners

Joshua Matthews, Kriss Lange, and Gunther M. Wiest

Abstract To develop adequate levels of automaticity in second language (L2) word recognition learners need to engage with a large amount of meaningful spoken target language input. Unfortunately, there is often not enough in-class time available for this. Thus, using technology to facilitate out-of-class development of L2 word recognition holds strong potential. This study explores and compares the use of a mobile language learning application (app) by Azerbaijani and Japanese learners of English. The app was designed to improve second language (L2) learners' recognition of the spoken form of high-frequency vocabulary by giving learners repeated and self-paced opportunities to listen to and transcribe target words presented in fluently articulated, contextual sentences. First, a quasi-experimental approach was applied in both contexts to compare improvements in word recognition among those who used the app and those in a control group that did not. Next, learner interaction data collected in the app's database were used to investigate links between learner engagement and improvements in word recognition. Further, the most challenging words for learners to recognize and transcribe were identified and compared between treatment group members from each context. Finally, stimulated recall protocols in the learners' respective first languages (L1) were conducted among a subgroup from each context to investigate the origins of their difficulty with aurally recognizing their most challenging words. Suggestions are provided for the development of word recognition from speech that may be useful both within and beyond the classroom.

Keywords L2 listening · L2 word form recognition from speech (WFRS) · Mobile-assisted language learning · Out-of-class learning

J. Matthews (✉)
University of New England, Armidale, Australia
e-mail: jmatth28@une.edu.au

K. Lange
University of Shimane Matsue Campus, Matsue, Japan
e-mail: k-lange@u-shimane.ac.jp

G. M. Wiest
ADA University, Baku, Azerbaijan
e-mail: gwiest@ada.edu.az

B. L. Reynolds (ed.), *Vocabulary Learning in the Wild*,
https://doi.org/10.1007/978-981-99-1490-6_13

1 Introduction

Recognizing words in speech is a fundamental starting point in the second language (L2) listening comprehension process. Without first recognizing the phonological form of words in speech, a listener is unlikely to be able to access the associated meaning. Rapid and accurate recognition of words from speech is a key attribute of skilled L2 listeners (Field, 2008a). As well as being a fundamental element of skilled listening, the recognition of words in L2 speech can present significant challenges for L2 learners (Lange & Matthews, 2020, 2021). These challenges can be understood by considering the intrinsic nature of spoken words. Unlike static words on a page which can be revisited by the reader's eye as required, words in speech are available to the listener for only a very brief duration. In short, words in speech are temporal. Further, words in speech are often blended (i.e., coarticulated) such that boundaries between consecutive words in an utterance (intonation unit) are not explicit. This can make the phonological form of spoken words variable and dependent on the acoustic context within which they occur. The specific challenges associated with L2 word recognition from speech highlight the merit of interventions that help learners effectively deal with this fundamental aspect of L2 learning.

Efforts to enhance L2 learners' spoken word recognition from speech should focus on developing automatized phonological knowledge of words. Firstly, to account for the temporal nature of speech learners must be able to quickly recognize spoken words without drawing excessively from their finite cognitive resources, which are required for higher level L2 listening comprehension processes. Automaticity in language processing is underpinned by implicit knowledge which develops in response to the frequency effects of exposure to input (Ellis, 2002). In other words, a learner's automaticity in word recognition develops in step with the number of opportunities to successfully recognize words in the spoken target language. Furthermore, to account for the blended and variable nature of spoken words, learners must also possess adequate levels of phonological knowledge. Simply put, learners need to know what words actually *sound like* when articulated in speech. Importantly, an L2 learner's knowledge of words in written form is typically not equivalent to knowledge of words in spoken form (Cheng & Matthews, 2018). For instance, it is not unusual for a language learner to know a word in the written form but be unable to recognize that same word when it is encountered in speech (Carney, 2021; Goh, 2000; Lange & Matthews, 2021). As word knowledge is modality specific, it is important to provide learners with extensive and meaningful opportunities to engage with *spoken* target language input.

Unlike most L1 listeners who develop automatic word recognition effortlessly throughout their lifetime, L2 listeners need to devote time to actively develop word recognition from speech. However, the outlay of time needed to facilitate the automatization of L2 word recognition from speech is a significant problem for language learning. Firstly, there is inadequate time within in-class settings to adequately engage with this involved task. Additionally, in English as a foreign language (EFL) contexts, the extent of a learner's English language usage and exposure may typically not extend beyond in-class learning. For example, even at the tertiary level and after years of formal instruction, learners may still not have the phonological knowledge needed to recognize the 3,000 most frequently occurring words of the English language (Matthews & Cheng, 2015). In such circumstances, it is especially important to assist learners in the development of their L2 word recognition from speech as part of targeted out-of-class learning. Mobile-assisted language learning holds particularly strong promise in facilitating language learning that extends beyond the classroom, especially in the current era, where most language learners almost always have powerful smartphones (mini-computers) close at hand. The current study explores the potential of mobile-assisted language learning for the development of L2 word recognition in out-of-class learning conducted within two different formal EFL university contexts in Azerbaijan and Japan.

2 Literature Review

2.1 The Importance of Word Knowledge in L2 Listening Comprehension

Recent research has made clear the strong connection between L2 word knowledge and L2 listening comprehension (Cheng et al., 2022; Matthews, et al., 2023; Vafaee & Suzuki, 2019; Wallace, 2022). The special importance of word knowledge in language learning can be attributed to the strong form-meaning unity that occurs at the lexical level (Hulstijn, 2002). In terms of listening, this means that if the phonological form of a word (or string of words) can be recognized, the listener has a chance to access the appropriate corresponding meaning. However, in relation to L2 listening, the temporal, blended, and variable nature of the form of words in speech makes their recognition a considerable challenge. Meeting this challenge entails the listener skillfully utilizing both linguistic information derived from bottom-up processing (i.e., aural decoding) and contextual information from top-down processing (Flowerdew & Miller, 2005). Therefore, difficulty recognizing L2 words from speech may generally be attributable to a combination of inadequate linguistic knowledge needed for bottom-up processing and inadequate utilization of contextual information (e.g., background knowledge, comprehension of the preceding aural text, pragmatic knowledge).

Listening comprehension problems often stem from a difficulty in recognizing words in speech despite those words being known by the L2 listener in the written form (Goh, 2000; Masrai, 2020). For example, Carney (2021) conducted interviews and analysis of 15 Japanese EFL learners' difficulties in comprehending English speech consisting of high-frequency vocabulary. The most common reasons for comprehension breakdown were difficulties with L1 phonological influence, word segmentation, and word recognition. Lange and Matthews (2021) also used a mixed methods approach among Japanese EFL learners to determine that a significant cause of misunderstanding in L2 listening was an inability to recognize the phonological form of L2 words. Through the application of L1 interviews, Lange and Matthews (2021) showed that phonological representations of high-frequency words stored in learners' mental lexicons were often strongly influenced by L1 intrusion. This saw learners struggle to map words from speech onto their corresponding semantic representations despite knowing the word's meaning and written form. Learners reported that their mental representations of English words had been altered by their Japanese phonological system making even known words difficult to recognize from speech. These issues suggest the importance of helping learners more accurately recognize the spoken form of words including those that the learners are likely to have encountered many times in their previous language learning experiences (i.e., high-frequency words).

In terms of pedagogical recommendations, research makes clear that it is important for learners to apply non-linguistic knowledge such as background knowledge and strategies to assist their L2 listening comprehension (Yeldham, 2016). However, early-stage learners are unlikely to have developed sufficient levels of automaticity in L2 language processing necessary to effectively integrate both linguistic and non-linguistic knowledge sources. When trying to comprehend L2 speech, such learners are likely to experience a heavy cognitive burden simply trying to catch words here and there from largely unrecognizable sequences of spoken words. Indeed, this inevitable circumstance speaks to the importance of applying listening strategies to try to accommodate for limitations of automaticity in linguistic processing. However, as Graham et al., (2010) assert "… a minimum level of vocabulary recognition is required before nonlinguistic knowledge … can be brought into play effectively" and that "… without accurate word recognition, applications of such knowledge are little more than guesses imposed on the text" (p. 14). A central objective of the current study is to investigate approaches to help learners develop better L2 word recognition from speech.

2.2 Approaches to Develop L2 Word Form Recognition from Speech

In its fullest sense, the construct of word recognition entails the capacity to both recognize the phonological form of a word and to map this form onto an appropriate meaning. In the current research, however, we have focused on the learner's capacity to recognize the phonological form of a word and map it against its corresponding written form, so-called *word form recognition from speech* (WFRS). A clear limitation of this construct is that it does not directly measure knowledge of word meaning. However, the practical advantage is that this construct facilitates convenient provision of automated computer-mediated feedback on learner performance. Further, as learners often have a more complete knowledge of words in the written form, providing learners with systematic opportunities to map the phonological form onto the corresponding written form has pedagogical value (Field, 2008b; Hulstijn, 2003).

Although a number of researchers have suggested the potential of technology in improving WFRS (Hulstijn, 2003; Jia & Hew, 2021a), few have empirically investigated the efficacy of such approaches in language classrooms (Matthews & O'Toole, 2015; Matthews et al., 2015, 2017). Furthermore, none to our knowledge have done so specifically in out-of-class contexts by way of the affordances of mobile devices. To our knowledge, the only study that has addressed the computer-mediated development of L2 WFRS by way of a quasi-experimental design is Matthews et al. (2015), which investigated the effectiveness of a prototype online app used in an in-class context among 96 Chinese EFL tertiary level learners. Results indicated that learners in a treatment group who used the app across a five-week period had significantly greater improvements in L2 WFRS when compared to a control group that did not use the app. The app played short sections of simple speech to learners, thereby giving them repeated opportunities to transcribe the text and receive subsequent feedback on performance. These results provide preliminary empirical support for the general recommendations put forward by scholars concerning how to develop L2 WFRS (Field, 2008b; Hulstijn, 2003) and demonstrate the capacity of computers to facilitate these recommendations in authentic learning contexts. However, many questions remain, for example, little is known about the extent of mobile technology's usefulness in the development of L2 WFRS in out-of-class locations. Moreover, related research has only been undertaken in a few research contexts (Matthews & O'Toole, 2015); little is known about how generic suggestions for the development of L2 WFRS may be differentially effective in different language learning contexts.

2.3 The Current Study

The current research can be positioned within Benson's (2011) model of *language learning beyond the classroom* (i.e., location, formality, pedagogy, and locus of control) in the following way. In terms of the location, the use of the application was undertaken out-of-class. The portability and omnipresence of mobile devices is a key advantage in this regard. Learners can engage in learning at almost any time and anywhere. The learning associated with the app in both Azerbaijan and Japan was formal in the sense that it was linked with tertiary level courses, albeit undertaken in locations beyond the formal classroom itself. In terms of the dimension of pedagogy, the app was used by the learners in a self-instruction mode. As Benson (2011) describes, "in self-instruction specially designed … [affordances] … take on the role of the classroom instructor and there is a strong intention to learn on the part of the learner" (p. 11). In relation to locus of control, as the use of the app was initiated as part of formal learning in both Azerbaijan and Japan, its use can best be described as other-directed. However, as the app was used in out-of-class settings and learners needed to make decisions about their use of the app (e.g., when, where and for how long), this can also be described as self-regulated learning (Lai et al., 2022).

This study seeks to help fill the ongoing gap in the literature by exploring the efficacy of a free mobile-assisted language learning app designed to improve the L2 WFRS of early-stage L2 learners. A key feature of the current research is the out-of-class implementation of the mobile app across two EFL contexts—Azerbaijan and Japan. This will not only enable us to critically interrogate the overall potential usefulness of the app but will also cast light on how learners interact with the app in different language learning environments. A key objective of the current study is to not only investigate the use of a mobile app to enhance word recognition from speech but to also use data from the app to draw insight about how to enhance in-class learning.

The following research questions will be addressed:

1. Is out-of-class usage of the app associated with significant improvements in WFRS and if so, does the magnitude of improvement vary between L1 groups?
2. What relationship is evident between the number of times learners listen to the app and improvements in WFRS?
3. From the 1,000 target words presented in the app, which are most challenging for learners to recognize and transcribe, and what could be learned about the origin of learner difficulty with these words through stimulated recall protocols?

3 Method

3.1 Participants

3.1.1 Azerbaijani Participants

The Azerbaijani treatment group ($n = 16$) and control group ($n = 16$) consisted of first-year students (17 to 18 years old) enrolled in a year-long English foundation program in which L1 instruction is minimal. Years and consistency of English education varied from one individual to another before entering university. Both groups were involved in this study via their respective course in listening and speaking. Foundation program students receive approximately 22 hours of English instruction per week before moving on to general education courses conducted in English from their second year (out of five). According to mean scores on a locally developed, university-led English proficiency exam taken prior to the study, all participants were within the Common European Framework of Reference for Languages (CEFR) A1 level (basic user, beginner).

3.1.2 Japanese Participants.

The Japanese treatment group (n = 17) consisted of second-year students (19 to 20 years old) enrolled in an English writing course conducted mainly in English. The control group (n = 16) also consisted of second-year students in a general English course. All Japanese participants had approximately three to six hours of English courses per week during this study. Most of these English courses were conducted predominantly in the learners' L1 by Japanese instructors. Prior to entering university, learners generally had received six years of English education. Scores from the Test of English for International Communication (TOEIC) for the treatment group (M = 572.3, SD = 137.7, n = 15) and the control group (M = 537.7, SD = 101.9, n = 13) indicated that their level of English proficiency was CEFR A2 (basic user, elementary) (Educational Testing Service, 2019).

3.1.3 Stimulated Recall Protocol Treatment Subgroup

From each of the L1 treatment groups, seven participants (14 in total) were selected to participate in stimulated recall protocols. The Japanese and Azerbaijani subgroup members were matched based on their pretest scores which varied by less than 10 points between the paired learners. Pairing subgroup members from Japan and Azerbaijan was intended to enable comparison between learners of similar levels of English proficiency in the two contexts.

3.2 The Mobile App

The *C-levels Vocab* app was designed by the first and second author to develop L2 WFRS by providing learners with multiple opportunities to listen to and transcribe high-frequency words. To be clear, the app is a free resource from which the designers of the app gain no financial benefit. The app presents the first 1,000 words from the Corpus of Contemporary American English (COCA) in blocks of 100 words (i.e., *10 c-levels*). The words are presented in descending frequency of occurrence, thus generally progressing from relatively easy to relatively difficult. Figure 1 presents selected screenshots of the app's user interface. In Fig. 1A, an example of a contextual sentence is shown. At this stage the full spoken sentence is played and is heard through the learner's device, as in "I like him because he's good." All sentences were spoken by a North American English speaker.

After listening, the learner types the target word into the corresponding text box; there are four attempts to listen and do so. If not transcribed on the first attempt, the word is presented to the learner again after other words of the c-level have been engaged with. It is only when a word is transcribed correctly on the first attempt (of four) that the app's logic categorizes it as *known* and it is withdrawn from the cycling target word list of that c-level. Instant feedback makes clear to the learner that they

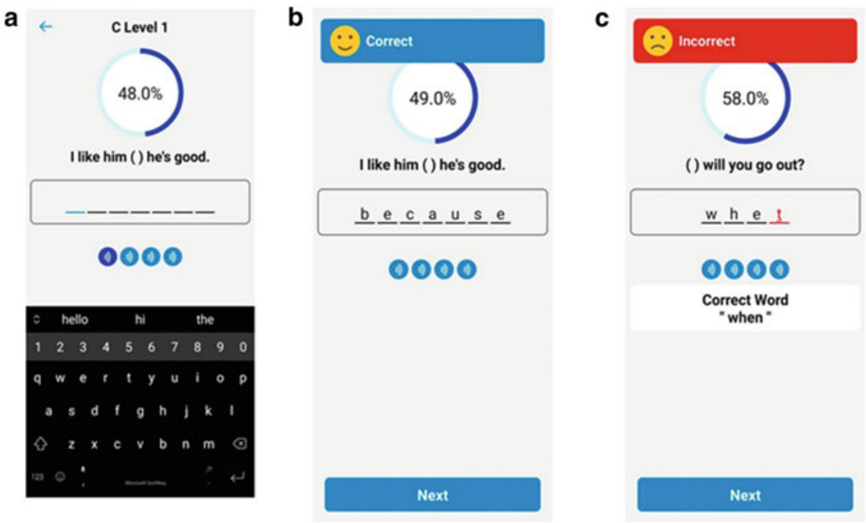

Fig. 1 Selected Screenshots of the App's User Interface

have correctly recognized and transcribed the word (Fig. 1B). It is only when all of the words of one c-level are *known* that the next c-level is made available to the learner.

Sentences and accompanying feedback were designed to provide as much contextual assistance to the learner as possible. The contextual sentences were piloted with native English language speakers to ensure the target words could be guessed without the benefit of the spoken sentence. This made sure the sentence itself afforded sufficient contextual support. Further, based on learner performance, incorrect letters were indicated in red and the correct target word was provided after four incorrect attempts (Fig. 1C). Learners were also provided with an overview of their progress through each c-level; this was shown as a percentage of words currently *known* for each c-level displayed on the screen. A range of data, such as the number of listens to each target word, was stored in the app's database for each learner.

3.3 Word Form Recognition from Speech (WFRS) Test

Before and after the intervention, the same 100-word pretest and posttest was administered to all participants from each context. Ten target words were randomly selected from each of the 10 c-levels of the app and these words were presented in both the pre- and posttest, albeit in a different order. As with the cloze format of the app, the test items consisted of a contextual sentence with the target word missing. Half of the target words were presented using contextual sentences which were identical to those used in the app and half had contextual sentences which were different from the app. This was done to confirm that learning effects from target words presented in the same contextual sentences were negligible. During administration, the test audio was played once to the respective treatment group and learners used their smartphones to enter the missing target words into an online cloze template. Words were automatically scored as either correct or incorrect.

3.4 Stimulated Recall Protocol

A structured stimulated recall protocol (Gass & Mackey, 2017) was applied in each context with the 14 subgroup members (seven from each context). Prior to the protocol, the 20 target words which had been listened to most by each subgroup member while using the app (i.e., the most challenging words) were identified from stored app data. Each subgroup member's list of words and the corresponding cloze phrases were printed onto a reference paper for use during the stimulated recall protocol. The target word's audio (the same as that of the app) was played and the subgroup member was asked to transcribe the missing target word. Afterward, the participant was asked to articulate the missing target word, provide its L1 meaning, and translate the contextual sentence with the target word into their L1. The researchers evaluated the subgroup members on each task binomially (correct as 1 or incorrect as 0). Finally, the participant was asked to explain why the target word had been especially difficult to recognize while using the app. This was done with the aim to identify and categorize the primary source of difficulty in transcribing each of the 20 target words. All of the subgroup member interviews were conducted in the participants' respective L1. The interviews lasted for approximately one hour and were audio recorded with the informed consent of each subgroup member.

Sources of WFRS error for each word were categorized based on previous research in the field (Lange & Matthews, 2020, 2021). The error categories are listed as follows:

1. Semantically unknown: The listener could not provide a L1 definition for the target word even after seeing its orthographic form.
2. Semantically known but phonologically unfamiliar: The meaning of the target word is known in the L1 but insufficient knowledge of its phonological form was primarily responsible for difficulty transcribing the target word.
3. Semantically known but phonologically unfamiliar due to the influence of connected speech: The meaning of the target word is known but attributes of connected speech, such as coarticulation, were primarily responsible for difficulty transcribing the target word.
4. Semantically known but spelled incorrectly: The meaning of the target word is known but spelling the target word incorrectly was primarily responsible for multiple failed transcription attempts while using the app.
5. Other reasons.

3.5 Procedure

Data for this study was collected independently by researchers in Azerbaijan and Japan. First, all of the participants in each context took the pretest via mobile devices. Next, the treatment group members downloaded the app and began to use it for WFRS development outside of class as homework over a period of approximately six weeks with the loose goal of completing one to two c-levels each week. After an initial practice session in class, the participants were asked to use the app outside of class at their own pace. This inevitably resulted in individual variation in how quickly participants completed the app tasks. All of the treatment group members completed the 1,000 words assessed with the app before taking the posttest. No control group participants used the app and they took the posttest during the same week as the treatment group. Finally, the subgroup members individually undertook the stimulated recall protocol with the respective researchers in each of the two contexts.

3.6 Data Analysis

Quantitative analyses included comparison of mean WFRS difference scores between groups (research question 1). Mean difference scores were determined by subtracting pretest scores from posttest scores. To test our hypothesis that those who used the mobile app (treatment groups) achieved greater mean WFRS difference scores than those that did not (control groups), independent samples *t*-tests were performed. Pearson correlation was also used to examine links between learner engagement with the app (e.g., the total number of times the participants listened to the words) and mean difference scores (research question 2).

The quantitative data collected through stimulated recall protocol were assessments of participants' ability to articulate the target word, provide an L1 definition for it, and translate the contextual sentence. Percentages of correct answers for each task were calculated to identify the most challenging target words and investigate factors which may explain suboptimal WFRS (research question 3). The qualitative data collected during stimulated recall protocols were audio recorded. Participant L1 responses to the question of why the target word had been difficult to transcribe were analyzed for explanatory themes. The themes were identified via a thorough examination of the responses from participants based on aspects of thematic analysis (Braun & Clarke, 2006). Researchers noted recurring explanations of comprehension difficulty such as the latter part of the target word being difficult to hear and influence from the L1 affecting comprehension. The majority of these responses were aligned with the error categories presented to participants during the protocol. These are explored in more detail in the discussion section.

4 Results

Research Question One —Is out-of-class usage of the app associated with significant improvements in WFRS and if so, does the magnitude of improvement vary between L1 groups?

Measures of internal consistency for both the pretest (Cronbach's $\alpha = 0.96$) and the posttest (Cronbach's $\alpha = 0.94$) were very good. The distribution of mean difference scores was sufficiently normal for these analyses (i.e., skewness & kurtosis each < 2).

For the Azerbaijani group, the assumption of homogeneity of variance was satisfied with Levene's test, $F(30) = 0.002$, $p = 0.966$. Independent samples t-test results showed a significant effect, $t(30) = 3.17$, $p = 0.004$ with Azerbaijani treatment group members achieving significantly greater WFRS difference scores (M = 24.56, SD = 9.70) than those in the Azerbaijani control group (M = 13.63, SD = 9.82) (Table 1). A Cohen's d of 1.12 suggested a large effect size (Plonsky & Oswald, 2014).

For the Japanese group, Levene's test results, $F(31) = 0.273$, $p = 0.605$, again verified the assumption of homogeneity of variance. Independent samples t-test results showed a significant effect, $t(31) = 3.17$, $p = 0.002$. As with the treatment group members from Azerbaijan, Japanese treatment group members achieved significantly greater WFRS difference scores (M = 9.47, SD = 7.25) than those in the Japanese control group (M = 1.31, SD = 6.20). As before, a large effect size was indicated (Cohen's d = 1.12).

The mean WFRS difference scores for the Azerbaijani treatment group (M = 24.56) were greater than those of the Japanese treatment group (M = 9.47). An independent samples t-test result showed that this difference reached a level of statistical significance ($t(31) = 1.731$, $p = 0.000$), with a large effect size (Cohen's d = 1.8). Of note is that the mean pretest score for the Azerbaijani treatment group (M = 56.13, SD = 17.90) was lower than that of the Japanese treatment group (M = 77.17, SD = 10.25). An independent samples t-test demonstrated that the difference in mean

Table 1 A comparison of test scores for treatment and control group members

	n	Mean pretest score	*SD*	Mean posttest score	*SD*	*MD*	*SD*
Azerbaijani treatment group	16	56.12	17.90	80.69	11.98	24.56	9.70
Azerbaijani control group	16	53.75	19.47	67.38	14.01	13.63	9.82
Japanese treatment group	17	77.18	10.25	86.65	7.11	9.47	7.25
Japanese control group	16	72.88	7.72	74.19	10.82	1.31	6.20

Note MD = mean difference between the pre- and posttest scores

Table 2 Descriptive statistics and correlations for total number of listens, difference score, and pretest score

Variable	*M*	*SD*	1	2	3
1. Total number of listens	1915.97	626.17	-		
2. Difference score	16.79	11.36	0.66**	-	
3. Pretest score	66.97	17.80	−0.87**	−0.85**	-

Note $N = 33$, ** $p < 0.001$

pretest scores between the treatment groups was statistically significant ($t(31) = –4.179$, $p = 0.000$).

Research Question Two —What relationship is evident between the number of times learners listen to the app and improvements in WFRS?

Stored data from the app provided a raw score of the number of times learners listened to each target sentence. As a learner's existing word recognition capabilities are likely to influence the patterns of interaction with computer-mediated language learning interventions (Matthews et al., 2017), another variable of interest in this analysis was pretest score. Table 2 shows the Pearson correlation matrix of all treatment group learners ($n = 33$).

A significant positive correlation was observed between total number of times listened and difference score (Table 2). The magnitude of this relationship was large (e.g., $r > 0.6$). The general trend evident is that more repeated listening to the app was strongly associated with greater WFRS improvement. There was a strong, negative correlation between pretest scores and number of times listened ($r = -0.872$, $p < 0.001$) and pretest scores and difference scores ($r = -0.851$, $p < 0.001$). The trend evident here is that learners with a lower pretest score listened more and were associated with greater difference scores than those with a higher pretest score.

Table 3 presents a breakdown of the number of times learners from each language group listened to the app across the duration of the intervention. Figure 2 visualizes the relationship between engagement with the app and WFRS difference scores.

Table 3 Descriptive statistics on mean number of times listened by L1 Group

	n	*Min*	*Max*	Mean number of times listened	*SD*
Azerbaijani treatment group	16	1,257	3,799	2,209	668.69
Japanese treatment group	17	399	2,160	1,640	445.31

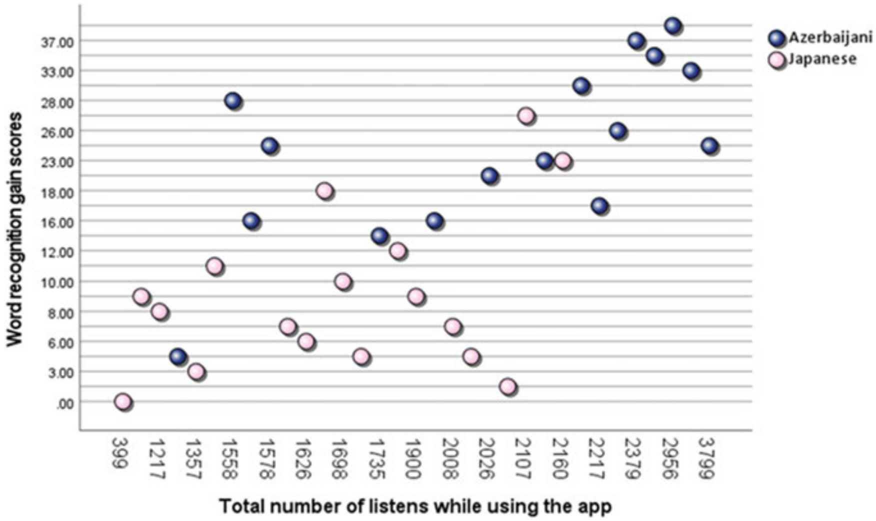

Fig. 2 Scatter Plot of WFRS Difference Score by Total Number of Listens

Research Question Three —From the 1,000 target words presented in the app, which are most challenging for learners to recognize and transcribe, and what could be learned about the origin of learner difficulty with these words through stimulated recall protocols?

Back-end data were used to rank each participant's target words according to the number of times listened when using the app. From this list, the 20 target words which were listened to most frequently were selected as a unit of analysis and are referred to here as the 20 most challenging target words. The mean number of times these words were listened to for the Azerbaijani cohort ($n = 16$) was 10.15 and 8.24 for the Japanese cohort ($n = 17$). Next, instances of the same target words in the 20 most challenging target words for each cohort were tallied. The target words which were shared five or more times within each cohort are listed in Table 4. For example, the first target word in the Azerbaijani cohort, *unidentified*, is the 274th most frequent word in the spoken section of the COCA and is shared on 10 of the 16 learners' lists of their 20 most challenging target words (indicated in the *Shared* column). Also, *unidentified, attorney,* and *correspondent* are followed by an asterisk indicating these three words were shared five or more times in both cohorts.

Table 5 presents the selection percentages for each of the five error categories for the Azerbaijani and Japanese subgroup members. These categories were selected by the participant and researcher during the stimulated recall protocol for each member's most challenging target words.

Commonalities within the subgroups as per their most challenging words were investigated by identifying the target words which repeat most often within each cohort member's list of most challenging words. The scope of investigation was

Table 4 Target words which were shared 5 or more times in lists of learners' 20 most challenging words

Japanese Cohort ($n = 17$)			Azerbaijani Cohort ($n = 16$)		
Frequency Ranking	Shared	Target word	Frequency ranking	Shared	Target word
475	9	attorney*	274	10	unidentified*
310	9	certainly	512	7	committee
264	9	senator	114	6	any
410	7	governor	256	6	court
855	7	career	425	6	sort
822	7	Iraqi	475	6	attorney*
261	6	democrat	594	6	correspondent*
274	5	unidentified*	263	5	political
594	5	correspondent*	81	5	him
73	5	our	236	5	whether
68	5	their	23	5	()n't
357	5	administration	433	5	threat
513	5	heart	705	5	able
829	5	soldier			

Note. Frequency Ranking refers to the frequency ranking for the target word in the spoken section of the COCA. *Shared* refers to the number of times the target word is shared in the lists of the 20 most difficult target words for each participant. Words followed by an asterisk (*) appear in both L1 cohorts' lists. *()n't* represents the contracted form of *not*

Table 5 Percentage of transcription error category selections

Transcription error category	Azerbaijani subgroup ($n = 7$)	Japanese subgroup ($n = 7$)
(1) Semantically Unknown	14%	26%
(2) Semantically known but phonologically unfamiliar	29%	31%
(3) Semantically known but phonologically unfamiliar due to the influence of connected speech	13%	19%
(4) Semantically known but spelled incorrectly	8%	21%
(5) Other (Could not explain why the item was difficult)	36%	4%

reduced from 20 target words to 10 target words, thus allowing a focus on only the most challenging items. The target words which repeated three or more times within the subgroups' 10 most challenging target words are listed in Table 5. To elucidate why these particular words were the most challenging, the frequency with which error categories 1 to 5 were selected are indicated in the far-right column of the table. For example, in Table 6 the first target word *attorney* is the 475th most frequently occurring word in the spoken section of the COCA. It appears on five of the seven subgroup members' lists of their 10 most challenging target words and was categorized as error category 1 by three learners and as category 2 by two learners. Table 7 presents data for the Japanese subgroup members in the same format.

Table 6 Difficult target words shared among the Azerbaijani subgroup

Freq	Shared	Target word	Target sentence	Categories				
				1	2	3	4	5
475	5	attorney*	If you have broken the law, you need to get an ()	3	2	0	0	0
274	5	unidentified*	We do not know all of their names, so some () people have been killed	1	1	1	2	0
512	5	committee	Maybe I can help. I am on the () that runs this event	2	1	0	1	1
594	3	correspondent*	The foreign () will send news back to America	0	1	0	1	1
397	3	candidate	He is the democratic () for President	2	1	0	0	0

Note. $n = 7$, *Freq.* = frequency ranking for the target word. *Shared* = times shared in lists of 10 most difficult target words. *(*)* shows words that appeared in both L1 cohorts' lists. *Categories* are defined as: (1) Semantically unknown, (2) Semantically known but phonologically unfamiliar, (3) Semantically known but phonologically unfamiliar due to the influence of connected speech, (4) Semantically known but spelled incorrectly, (5) Other

Table 7 Difficult target words shared among the Japanese subgroup

Freq	Shared	Target word	Target sentence	Categories				
				1	2	3	4	5
310	6	certainly	If there is an answer I () don't know it	1	4	1	0	0
822	4	Iraqi	The war started over American interest in () oil. They wanted more control of the oil from that foreign country	4	0	0	0	0
248	3	police	Stop now or I'll call the ()!	0	3	0	0	0
425	3	sort	You can't do that () of thing here	2	0	0	1	0
594	3	correspondent*	The foreign () will send news back to America	3	0	0	0	0
507	3	commercial	It made a lot of money so the movie did well in a () sense	1	0	0	2	0

Note $n = 7$, *Freq.* = frequency ranking for the target word. *Shared* = times shared in lists of 10 most difficult target words. *(*)* shows words that appeared in both L1 cohorts' lists. (See note in Table 5 for details of the *Categories*)

5 Discussion

The findings from research question one show that the treatment groups in both L1 cohorts made significant improvements in WFRS. The Azerbaijani treatment group gained an average of approximately 10.93 words more than the Azerbaijani control group with a large effect size (i.e., $d = 1.12$). The Japanese treatment group gained an average of 8.16 words more than the Japanese control group ($d = 1.12$). This provides evidence that WFRS was enhanced through use of the app for both L1 groups, however, the Azerbaijani learners' gains were larger. This difference may be attributable in part to the initially lower proficiency levels of the Azerbaijani learners since instruction in aural word recognition tends to produce better results for lower proficiency learners (Jia & Hew, 2021b). Another contributing factor may be the Azerbaijani group's greater use of the app with a mean difference of 569 more listening times than the Japanese cohort. Moreover, their gains may have been larger from having up to 19 more hours of English instruction per week than their Japanese counterparts.

This overall positive result for both groups provides preliminary evidence that the affordances of the app may be beneficial in enhancing WFRS among learners of diverse language backgrounds and proficiency levels. Due to the various possible contributing factors, however, more empirical research is required to determine the sole contribution of the app on WFRS development. In terms of a comparison with the one previous study that we are aware of that was specifically directed towards computer-mediated development of L2 WFRS, results from the current research appear to be positive. The Matthews et al. (2015) in-class study noted an improvement in WFRS between treatment and control group members, but the magnitude of this improvement (approximately 1.5 words difference between control and treatment group) was less than that noted in the current study and had a smaller effect size (i.e., $d = 0.47$). Although more research is needed before more assertive conclusions can be made, this comparison is at least suggestive of the feasibility of developing WFRS capacities in out-of-class contexts via the use of mobile apps.

Research question two explored the relationship between the number of times learners listened to the app and WFRS difference scores. There was a strong significant correlation between the number of times learners listened and their difference scores ($r = 0.66$). Pretest scores had a strong inverse relationship with both number of listens and WFRS score gains, which shows that learners with lower proficiency listened more and improved more than higher proficiency learners. This result is not entirely unexpected based on previous research. For example, Matthews et al. (2017) demonstrated that learners of different proficiency levels experienced significantly different WFRS gains after computer-mediated intervention, with mid-level proficiency learners achieving greater difference scores than both lower and higher proficiency learners. In sum, these findings reinforce the importance of computer-mediated approaches that are adaptive. A future area for improvement with the C-levels app (and others like it) would be the addition of an algorithm that modulated the difficulty of the target words depending on the learner's preceding performance with the task. For example, if the learner was recognizing all words correctly, the app could increase the difficulty (decrease the frequency) of the target words (e.g., skip a c-level). Also, the Azerbaijani treatment group listened to the app on average 569 times more than the Japanese treatment group. This increased level of engagement may have been influenced by several factors, including proficiency and affective factors. For example, the Azerbaijani groups' generally lower proficiency may have meant the app content was more immediately relevant to their learning needs and this may have motivated higher levels of engagement.

Suggestions for developing WFRS based on our findings are provided next. Because participants who used the app more frequently made greater improvements, it is recommend that learners focus on developing WFRS through frequent listening and cloze transcription practice. In addition to practice via the app (or similar affordances), explicit instruction focusing on patterns of phonological modification in connected speech, word stress patterns, and utilization of contextual meaning to support WFRS development is also recommended. As the analysis of participants' most challenging target words demonstrated, individual learners have unique difficulties with WFRS due to a variety of factors. Screening for difficult words via a tool such as the C-Levels Vocab app allows researchers and learners to focus on addressing the unique challenges presented by each word for each learner. Developing WFRS for words that are especially difficult may require explicit instruction tailored to the individual learner and their L1 group rather than simply more generalized practice.

The first part of research question three investigated the most challenging target words for learners to recognize and transcribe when using the app. In both groups, there were at least 13 target words which five or more learners found particularly difficult. Many of these were high-frequency words, which is consistent with previous research demonstrating difficulty with WFRS for known words (Carney, 2021; Lange & Matthews, 2021). For the Japanese participants, seven of the 14 words in Table 3 (*certainly, senator, democrat, unidentified, our, their, administration*) had frequency rankings which were higher than the 400th ranked word in the spoken section of the COCA. Among the Azerbaijani participants' lists as well, seven of the 13 words in Table 3 (*unidentified, any, court, sort, attorney, political, him, whether, ()n't*) rank higher in frequency than the 400th ranked word. To illustrate, *their* (ranked 68th in frequency) was shared on the lists of five Japanese learners' most challenging words and was a noted challenge for Azerbaijani learners also. It is likely that this difficulty is at least partially attributable to Japanese and Azerbaijani not having a voiced dental fricative /ð/. The finding that very high-frequency words can cause difficulty for language learners is of strong interest. Teachers and researchers should be aware that the phonological form of some high-frequency words may not be well known by some learners and that these problematic high-frequency words may vary depending on the learner's L1. It is important for teachers to identify these problematic high-frequency words and offer them explicit attention in an effort to raise learners' awareness of the potential challenges of recognizing their phonological forms. For example, providing learners with repeated opportunities to hear authentic samples of connected speech containing these problematic words and then again while listening and reading an accurate corresponding written transcript is an important first step. The provision of metalinguistic explanations of how the phonological form of these challenging high-frequency words may vary when articulated in fluent speech (e.g., variable acoustic contexts) is also warranted.

It is also interesting to note that scores for the Azerbaijani subgroup members on the target word articulation and transcription tasks were over 10 average percentage points greater than those of the Japanese learners. Although speculative, this slight advantage for the Azerbaijani learners may be attributable to a stronger similarity between the orthographic systems of Azerbaijani and English, when compared to that of Japanese and English. As an example, consider the words *Azerbaijan* (i.e., Azərbaycan) and *Japan* (i.e., 日本) written in Azerbaijani and Japanese respectively.

The second part of research question three addressed why learners in both contexts had experienced difficulty in WFRS for certain target words. Five error categories were used to clarify the primary reasons for transcription difficulty. The largest error category in the Japanese subgroup was (*2) Semantically known but phonologically unfamiliar* which, as confirmed via the stimulated recall protocols, represents a pervasive limitation on WFRS. Learners often recounted incongruences between their mental representation of a word's phonological form and the phonological form they perceived from the audio recording. Data suggested that these differences stem from extensive prior exposure to Japanese-accented English that seemed to have created Japanese-accented phonological forms of English words in the mental lexicon. For example, one subgroup member explained their inability to recognize *hand* (/hænd/) in the spoken utterance "It's more work than I thought. Could you give me a () with this?" was due to an inaccurate phonological representation of the target word. Japanese requires mora to have a vowel in the syllable-final position so *hand* was modified to / hɑndo/ in the learner's lexicon to conform to Japanese phonotactics. Another learner had difficulty recognizing *school* in the spoken utterance "Our child said she didn't want to go to () today." The learner explained that "The end of the word is hard to hear" due to the consonant in the syllable-final position and further stated that "[they] have a habit of hearing in *katakana*" (i.e., the Japanese phonetic syllabary). By this, we assume the learner was describing the process of mapping English words onto Japanese-accented representations in the mental lexicon. Thus, when *hand* is recognized it is associated with /hɑndo/ and if *school* is recognized, it is mapped to /sukuuru/ in the mind of the listener. In another example, the learner explained that her difficulty perceiving *police* in "Stop now or I'll call the ()!" was due to influence from Japanese. She repeatedly entered *please* for *police* despite having semantic knowledge of the target word. Every subgroup member described at least one similar difficulty related to influence from Japanese-accented English and previous studies have also documented this trend (Lange & Matthews, 2020, 2021). Influence from Japanese-accented English input can be more generally understood as L1 phonotactic constraints whereby the listener tries to apply phonological conventions of the L1 (e.g., placing vowels in the syllable-final position) erroneously to the L2 (Cutler, 2012). The largest error category for the Azerbaijani subgroup was (*5) Other* which mainly reflected the learners' difficulty in explaining why WFRS had been difficult. Speculatively, this may be attributable to learners in this context having limited experience with critical analysis of their own English language performance. This also underscores the general difficulty in self-assessing and describing limitations in one's implicit L2 linguistic knowledge, especially among learners with relatively low proficiency levels. This speaks to the particular importance of teachers working closely with lower proficiency learners to help them identify and resolve the specific challenges they have in recognizing the phonological form of words.

Although informative, this study had several limitations. First, WFRS development for the Azerbaijani learners was likely to be disproportionately affected by the 20 or more hours of English instruction they received each week during the study compared to the Japanese learners' three to six hours. Another limitation was the inclusion of culturally-bound words within the COCA (e.g., *senator, attorney, Iraqi*) which may have affected WFRS improvement differently in the two contexts. The use of knowledge-based vocabulary lists specific to L1 language groups (when they become available) to guide the order and selection of the content of similar interventions is advised for future studies (see, Schmitt et al., 2022). An additional limitation was that due to feasibility constraints only one rater in each context scored the stimulated recall protocol assessments and analyzed the qualitative data from the learners' L1 responses. Scores from multiple raters are recommended to accurately assess inter-rater reliability.

6 Conclusion

Recognizing words from speech is a fundamental skill in L2 learning and is particularly important for listening development. Unlike information processed through the orthographic modality, WFRS must be executed fluently to keep up with the transient nature of spoken input. Regular engagement with out-of-class mobile-assisted language learning like that at the center of this study is likely to help develop this capacity across different language learning contexts. Although all learners regardless of L1 shared some of the same challenging words in the current study, an important takeaway was that there was sizable variation in the specific words that learners found challenging. Another key finding was that the degree of difficulty encountered with any given word did not necessarily correspond to the frequency of occurrence of that word (i.e., as indicated by the COCA). Therefore, an individualized approach to WFRS development is suggested in which each learner's *problem* words are identified and the underlying factors responsible for suboptimal WFRS are addressed. The approach applied in the current study has enabled us to cover new ground, but there is more work to be done. Moving forward, we call for more mixed-methods research that triangulates out-of-class language learning app usage data with one-on-one interview data, both within and across learning contexts. Such research will inform us on how to strategically apply technological affordances to facilitate effective out-of-class learning.

References

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Carney, N. (2021). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly, 55*(2), 536–567. https://doi.org/10.1002/tesq.3000

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3–25. https://doi.org/10.1177/0265532216676851

Cheng, J., Matthews, J., Lange, K., & McLean, S. (2022). Aural single-word and aural phrasal verb knowledge and their relationships to L2 listening comprehension. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3137

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* MIT Press.

Educational Testing Service (2019). *Mapping the TOEIC test on the CEFR.* https://www.ets.org/s/toeic/pdf/toeic-cefr-flyer.pdf

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Field, J. (2008a). Revising segmentation hypotheses in first and second language listening. *System, 36*(1), 35–51. https://doi.org/10.1016/j.system.2007.10.003

Field, J. (2008b). *Listening in the language classroom.* Cambridge University Press. https://doi.org/10.1017/CBO9780511575945

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice.* Cambridge University Press.

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research.* Routledge. https://doi.org/10.4324/9781315813349

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning and Teaching, 4*(1), 1–20. https://doi.org/10.1080/17501220802385866

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*(3), 193–223. https://doi.org/10.1191/0267658302sr207oa

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning, 16*(5), 413–425. https://doi.org/10.1076/call.16.5.413.29488

Jia, C., & Hew, K. F. (2021a). Meeting the challenges of decoding training in English as a foreign/second language listening education: Current status and opportunities for technology-assisted decoding training. *Computer assisted language learning, 1–30*,. https://doi.org/10.1080/09588221.2021.1974051

Jia, C., & Hew, K. F. (2021b). Toward a set of design principles for decoding training: A systematic review of studies of english as a foreign/second language listening education. *Educational Research Review, 33*, 100392. https://doi.org/10.1016/j.edurev.2021.100392

Lai, Y., Saab, N., & Admiraal, W. (2022). University students' use of mobile technology in self-directed language learning: Using the integrative model of behavior prediction. *Computers & Education, 179*, 104413. https://doi.org/10.1016/j.compedu.2021.104413

Lange, K., & Matthews, J. (2020). Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension. *Studies in Second Language Learning and Teaching. 10*(4). 723–749. https://doi.org/10.14746/ssllt.2020.10.4.4

Lange, K., & Matthews, J. (2021). Analyzing trends in the aural decoding errors of Japanese EFL learners. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2020-0050

Masrai, A. (2020). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Applied Linguistics Review., 11*(3), 423–447. https://doi.org/10.1515/applirev-2018-0106

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

Matthews, J., Cheng, J., & O'Toole, J. M. (2015). Computer-mediated input, output and feedback in the development of L2 word recognition from speech. *ReCALL*, *27*(3), 321–339. http://doi:https://doi.org/10.1017/S0958344014000421

Matthews, J., Masrai, A. Lange, K., McLean, S., Alghamdi, E. A., Kim, Y. A., Shinhara, Y., Tada, S. (2023). Exploring Links Between Aural Lexical Knowledge and L2 Listening in Arabic and Japanese Speakers: A Close Replication of Cheng, Matthews, Lange and McLean (2022). TESOL Quarterly, https://doi.org/10.1002/tesq.3212

Matthews, J., & O'Toole, J. M. (2015). Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning, 28*(4), 364–382. https://doi.org/10.1080/09588221.2013.864315

Matthews, J., O'Toole, J. M., & Chen, S. (2017). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: Design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning, 30*(1–2), 22–43. https://doi.org/10.1080/09588221.2015.1129348

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, & L., Kremmel, B. (2022). *Knowledge-based vocabulary lists user manual: A supplementary resource*. British Council. https://www.britishcouncil.org/sites/default/files/kvl_user_manual_final.pdf

Vafaee, P., & Suzuki, Y. (2019). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition, 42*(2), 383–410. https://doi.org/10.1017/S0272263119000676

Wallace, M. P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning, 72*, 5–44. https://doi.org/10.1111/lang.12424

Yeldham, M. (2016). Second language listening instruction: Comparing a strategies-based approach with an interactive, strategies/bottom-up skills approach. *TESOL Quarterly, 50*(2), 394–420. https://doi.org/10.1002/tesq.233

Joshua Matthews is a Senior Lecturer and Chair of the School Education Committee at the School of Education, University of New England, Australia. His major research interests include computer-assisted language learning, L2 listening, L2 vocabulary, L2 teaching, and language testing. His previous publications have appeared in journals including Computer-Assisted Language Learning, ReCALL, Language Learning and Technology, TESOL Quarterly and Language Testing. His current research projects include those addressing L2 aural vocabulary knowledge, the role of noticing in L2 development, multi-modality in language testing, and the design of MALL.

Kriss Lange is an Associate Professor at the University of Shimane Matsue Campus, Japan. His research interests include L2 listening, vocabulary acquisition and mobile-assisted language learning. Based in Japan, his research has been mainly focused on understanding English listening difficulties for Japanese university students and on L2 listening education. His previous research has been published in journals such as TESOL Quarterly and International Review of Applied Linguistics in Language Teaching.

Gunther M. Wiest has taught English in Japan, Tunisia, Azerbaijan, and his native USA. The foundation level courses in listening, speaking, and academic vocabulary that he facilitates as Senior Instructor at English for Academic and Professional Purposes (EAPP), ADA University, inspire him to research contemporary methods of language acquisition. These include student telecollaboration or COIL with faraway peers, and as in the current volume, self-study with educational technology.