Technology-Mediated Second Language Vocabulary Development: A Review of Trends in Research Methodology
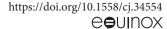
Irina Elgort

Abstract

Technology-mediated vocabulary development (TMVD) in a second language (L2) covers a wide range of instructional and learning treatments, contexts, and technologies and is situated in a broader field of second language vocabulary learning. Vocabulary knowledge is a complex, multidimensional construct that has been interpreted and categorized in second language research in many different ways. This review identifies methodological practices in research into L2 TMVD and provides a synthesis of learning and instructional approaches, aspects of vocabulary knowledge being investigated, and measures of vocabulary development used in this research. A sample of 82 primary studies from 2010 to 2017 was selected and coded for target methodological features. The results of the review show that TMVD research has achieved a certain degree of methodological maturity, while a number of issues have also been identified. These issues include: the practice of comparing technology-mediated instruction with "traditional" instruction without technology, insufficient reporting of participants' L2 proficiency, and a dearth of treatments targeting fluency development. The review also showed a preference for evaluating the knowledge of individual vocabulary items rather than the development of the L2 lexicon, and insufficient use of online and implicit measures of vocabulary knowledge. Recommendations for future TMVD research are provided.

KEYWORDS: TECHNOLOGY-MEDIATED VOCABULARY DEVELOPMENT, METHODOLOGY, ASPECTS OF VOCABULARY KNOWLEDGE, MEASURES OF VOCABULARY KNOWLEDGE

Affiliation

Victoria University of Wellington, New Zealand.
email: Irina.Elgort@vuw.ac.nz

equinoxonline

Introduction

Vocabulary development is widely recognized by learners, teachers, and researchers as a key component of learning a language. The history of technology-mediated vocabulary learning goes as far back as computer-assisted language learning itself, around the mid-twentieth century (Goodfellow, 1995). Today, vocabulary development is both mediated and shaped by digital technologies, such as the Internet, personal computers, and mobile devices (Elgort, in press). The ever-increasing capabilities to efficiently store, process, and transfer large amounts of digital text, sound, and video data afford the creation and use of large language corpora, dynamic reference tools, text and speech analysis and synthesis software, and multimedia communication tools. Research into technology-mediated vocabulary learning offers insights into how these new technologies, combined with innovative instructional and learning approaches, affect the development of different aspects of vocabulary knowledge (VK). However, apart from Chiu's (2013) meta-analysis of computer-assisted second language (L2) vocabulary instruction, there have been no recent reviews of technology-mediated vocabulary development that cut across specific sub-sections of computer-assisted language learning research, such as the use of digital reference and support tools, data-driven learning, computer-mediated communication, digital gaming, multimedia-based learning, and others.

A number of recent research syntheses in the field of technology-mediated language learning have taken a quantitative meta-analysis approach, using effects sizes across experimental studies to represent the effect of technology-mediate learning and instruction on L2 learning outcomes (Abraham, 2008; Boulton & Cobb, 2017; Lin, 2015; Montero Perez, Van Den Noortgate, & Desmet, 2013; Yun, 2011). In a second-order synthesis of 30 years of research probing the relationship between computer-assisted language learning and second language acquisition, Plonsky and Ziegler (2016) identified 14 first-order meta-analysis studies, 10 of which were conducted after 2010. The meta-analysis approach overcomes a number of limitations of primary experimental and quasi-experimental in-situ studies that are often characterized by low statistical power, as a result of small numbers of participants and items per condition, convenience sampling approaches, "dogmatic adherence to and overuse of null hypothesis significance testing" (Plonsky & Kim, 2016, p. 78), and underreporting of effect sizes (Lindstromberg, 2016; Plonsky, 2013). However, experimental designs alone may not be sufficient to evaluate novel technological and instructional methods in applied linguistic research, considering that genuine control is not always possible in real educational settings and that specific in-situ cases of technology-mediated learning may be difficult to reproduce. According to Kember (2003), using triangulations across

multi-method evaluations from several sources may be preferable when investigating pedagogical innovation.

The present article extends the scope of investigation beyond quantitative experimental and quasi-experimental research, by including observational and qualitative studies into the selected set of articles and, instead of focusing exclusively on the effect of technology-mediated learning, situates it in the context of the broader L2 vocabulary research. The present review creates a synthesis of methodological approaches currently used to study technology-mediated vocabulary development (TMVD), in an attempt to establish whether recent research in this domain is aligned with research trends in L2 vocabulary acquisition, teaching, and learning more generally.

Vocabulary Knowledge and its Development

Aspects of Word Knowledge

Vocabulary knowledge is a multidimensional construct that has been understood and categorized in L2 education research in many different ways (Elgort & Nation, 2010; Henriksen, 1999; Read, 2004; Wesche & Paribakht, 1996). One of the most comprehensive taxonomies of what it means to know a word was proposed by Nation (1990, 2001), who outlined three main aspects of knowledge and their constituent parts: (1) *knowledge of form* (spoken, written, and word parts); (2) *knowledge of meaning* (form-meaning mapping, conceptual or referential meaning, and associations), and (3) *knowledge of use* (grammatical function, collocations, and constraints on use). These aspects of knowledge can be realized *receptively* (in listening or reading) or *productively* (in speaking or writing). Notably, developmental trajectories of the different aspects of VK are not identical and may vary depending on the learning conditions and individual learner characteristics. Therefore, TMVD studies need to clarify what aspects of vocabulary knowledge are being investigated and use appropriate measures to estimate the development of these knowledge aspects (Read, 2000).

Measures of Vocabulary Knowledge and its Development

Vocabulary knowledge measures may be used to gauge an overall size of a person's lexicon (how many words she knows in the target language, without focusing on how well they are known) or to probe the quality of one or more aspects of word knowledge (Ma, 2009; Nation, 2001; Read, 2000). Vocabulary knowledge measures may focus on receptive knowledge (recognition and processing of spoken and written form and retrieval of meaning) or productive knowledge (retrieval of vocabulary in spoken or written production). Vocabulary tests may be designed to evaluate participants' ability to access VK online (i.e., under time pressure) or offline (allowing for conscious efforts to

retrieve knowledge, e.g., by using mnemonics, analyzing a word's structure, or using task strategies, such as guessing or elimination). Online lexical access can be measured by tracking eye-movements during reading, analyzing real-time interactions or using behavioral and linguistic tasks conducted under time pressure. Online access to VK is considered to better represent functional knowledge needed in real language use. However, offline measures (e.g., translation or multiple-choice vocabulary tests) are often easier to administer and may be more accessible to teacher-researchers.

Implicit measures of VK may be obtained using research methods from cognitive psychology designed to measure different aspects of language processing that are not available to learners' conscious control or report, such as their eye-movements during reading (Elgort, Brysbaert, Stevens, & Van Assche, 2017; Godfroid, Boers, & Housen, 2013) or task completion (Michel & Smith, 2017; Smith, 2012), recognition, processing, and retrieval times in priming manipulations (Elgort, 2011; Elgort & Piasecki, 2014), or neurological markers of lexical processing, such as N400 (Elgort, Perfetti, Rickles, & Stafura, 2015; McLaughlin, Osterhout, & Kim, 2004). Implicit measures may be useful in obtaining evidence of learning at early stages of vocabulary development, before learners are able to articulate their knowledge explicitly (e.g., in vocabulary interviews or translation tasks). These measures can provide insights into the processes of integration of a novel word into the lexical semantic networks of the learner. Implicit measures are also useful in probing aspects of knowledge that are not easily accessible using explicit measures, such as degree of involvement of cognitive control in language production and processing.

VK may be elicited by administering tasks that present target words or multi-word units in isolation (e.g., *LexTALE*, by Lemhöfer & Broersma, 2012), in limited contexts (e.g., VST, by Nation & Beglar, 2007), or by observing and analyzing real linguistic behavior and artifacts produced by language learners, for example, by measuring lexical richness or lexical diversity in writing (Crossley, Salsbury, & McNamara, 2012) or online chat (Sauro & Smith, 2010).

Learning and Instructional Approaches to Vocabulary Knowledge Development

The question of what learning and instructional approaches facilitate the development of VK is at the core of L2 vocabulary studies. Nation (2007) proposed that, in a well-designed language-learning curriculum, equal amounts of time should be spent on activities representing the following four learning strands: meaning-focused input (MFI, e.g., extensive listening and reading), meaning-focused output (MFO, e.g., speaking, writing, role-play), language-focused learning (LF, e.g., vocabulary exercises, intensive reading), and fluency development (FD, e.g., linked skills activities with familiar vocabulary). Employing

this "four-strands" framework, this review investigates which learning strands are represented in research into TMVD.

Vocabulary development may proceed with or without direct teacher mediation. In a meta-analysis of studies on computer-assisted L2 vocabulary instruction, Chiu (2013) reported a larger effect size for learning L2 vocabulary without the help of a teacher ($d = 1.379$, $p = 0.001$) than with teacher aid ($d = 0.214$, $p = 0.068$). Therefore, in the present review of TMVD studies, the following learning conditions are considered: teacher-mediated learning (TCH), with or without technology, versus independent technology-mediated learning (TEC), conducted in class or out of class.

Vocabulary learning is often described as deliberate (intentional) or incidental (with or without additional instructional or learning focus). Following Hulstijn (2013), vocabulary learning is considered incidental if it is not the primary focus of the activity in which a person is engaged. For example, some aspects of word knowledge may be picked up while reading, listening, or playing a game, in the absence of explicit instructions to learn it. Even when a dictionary is used to look up unfamiliar words in an effort to increase comprehension, word learning can still be described as incidental. In deliberate vocabulary learning, the learner sets out to learn new words or to improve the quality or strength of her existing VK. Examples of deliberate learning include using flashcards, word lists, word games, use of mnemonics to assist memorization; but deliberate learning may also involve studying words in context (e.g., using concordance tools, inferring word meanings from context, using reference materials while reading or listening). The incidental-deliberate learning dimension is examined in the present review because it is widely recognized and used in vocabulary learning studies. It is important to point out, however, that the incidental-deliberate distinction is not always clear-cut, as it is sometimes difficult to fully account for the learning intention of an individual. For example, a reading comprehension activity with a dictionary look-up may be considered incidental learning, but it may also be deliberate learning if the reader actively tries to commit the meaning of the looked-up word to memory.

In summary, this review article will address the following questions: (1) What learning and instructional approaches are investigated in TMVD studies? (2) What aspects of L2 vocabulary knowledge development are considered in these studies? (3) What measures are used to evaluate this development?

Methodology

Inclusion-Exclusion Criteria

The initial search was conducted in the Linguistics and Language Behavior Abstracts (LLBA) database. LLBA was selected because of its coverage of computer-assisted language learning and because, to be included in this

database, all articles have to pass through a human screening process, contributing to the reliability of the search results. The LLBA search results were checked using the Educational Resources Information Center (ERIC) database. The following search criteria were used:

- ("computer assisted" OR "technology enhanced" OR "computer mediated" OR "technology mediated") AND ("vocabulary" OR "lexical") AND ("L2" OR "second language" OR "foreign language")
- Peer reviewed
- Date: after 1 January 2010

This search yielded 565 entries. An additional search was conducted replacing ("vocabulary" OR "lexical") with ("collocation" OR "multi word" OR "idiom"), while keeping the rest of the search terms constant. This search resulted in 69 hits, a small number of which overlapped with the original search. In order to narrow down the results to the domain of TMVD, the articles were examined to confirm that they

- were published in international peer-reviewed language education journals;
- were primary (original) research;
- contained the search terms used in their intended meaning;
- measured vocabulary development in an additional language (directly or indirectly);
- involved technology-mediated language learning or language use;
- included participants who were speakers/learners of an addition language (i.e., excluding L1-only studies);
- were conducted with normal human participants (who did not have cognitive or functional impairments);
- were available in English.

This filtering process resulted in the selection of 82 articles in the domain of TMVD from the start of 2010 until April 2017.

Approach to Data Analysis

The selected articles were first examined for *general methodological attributes* (i.e., participants, items, study duration, technology used, study design, approach to data analysis) and then coded using *TMVD coding* themes (one at a time) based on the L2 vocabulary research outlined above. After close reading of the articles, initial TMVD codes were piloted using eight articles randomly chosen from the pool of the selected TMVD articles, and minor adjustments to the codes were made. The coding was conducted in two cycles,

main and verification, in order to ensure consistency in the codes' application. The following themes were used in the TMVD coding procedure (see Appendix A for the full list of codes):

- whether vocabulary development was a primary focus of the study;
- learning and instructional approaches used in the study;
- aspect of vocabulary knowledge investigated;
- measures used to evaluate vocabulary development;
- reported findings.

Analysis

Out of the 82 articles, 25 (over 30%) were published in 2016; during the period of 2010–2014, the article numbers are similar (between 8 and 12 per year). However, because of the dip in TMVD papers published in 2015 (n = 3) and only the first quarter of 2017 being examined in the present review, it would be premature to suggest that TMVD research is on the rise. The articles were published in 22 journals, with the largest proportion appearing in *Computer Assisted Language Learning* (n = 19; 23%) and *CALICO Journal* (n = 17; 21%), followed by *ReCALL* (n = 13; 16%) and *Language Learning & Technology* (n = 10; 12%) (Appendix B, Table 1).

Keywords and Abstracts

First, the keywords and abstracts of the selected articles were examined to create an overview of the studies, in the authors' own words. "Vocabulary" was the third most frequent item in the top 10% of the most frequent keywords (Appendix B Table 2). "Reading" was the most commonly mentioned keyword among the four skills (2.5%), followed by "listening" (1.7%) and "writing" (1.7%); "speaking" was listed as a keyword in only one article. "English" was the only language that was listed as a keyword.

In the abstracts' corpus, "vocabulary" was the most common content word, occurring 221 times (9.5% of all types). In the *n*-gram analysis, the most frequent 5-gram was "English as a foreign language" (n = 13); "computer assisted language learning" was the most frequent meaningful 4-gram (n = 10); the 3-gram, "data driven learning," appeared 9 times and the most frequent topic-relevant bigrams were "vocabulary learning" (n = 48), "post test"[1] (n = 30), "control group" (n = 24), "vocabulary acquisition" (n = 23), and "experimental group" (n = 18).

Technologies

The following technologies were used in the TMVD studies: specialized CALL software (n = 28), digital reference tools, such as glosses and dictionaries (n =

15); computer-mediated communication (CMC) (n = 13), digital video-based learning (n = 13), data-driven learning (DDL) and corpora-mediated learning (n = 12), digital games and gaming (n = 10), mobile learning (n = 9), learning management systems (LMS) (n = 4), e-books (n = 3), and social software (n = 2). In a number of studies, more than one technology type was used; in particular, CALL software was combined with digital reference tools, digital video-based learning, digital storytelling, mobile learning, DDL, and CMC. CMC was also combined with the use of LMS and with mobile learning in two studies; DDL was used with digital reference tools in two studies; social software—with mobile learning; and games—with reference tools. Various types of multimedia learning appeared in 15 studies, the use of captions and subtitles in 10 studies, glosses and glossing in 8 studies, and flashcard software in 6 studies.

Review of the General Methodological Attributes

On average, there were 67 participants per study (*SD* = 54). The smallest participant samples (n = 6 and n = 7) were used in two observational, qualitative studies, one of which examined vocabulary learning opportunities afforded by a virtual 3D learning environment (Milton, Jonsen, Hirst, & Lindenburn, 2012) and the other examined the nature of preschool-aged children's use of iPads for independent language learning (Terantino, 2016). The largest number of participants (n = 275, across four conditions) was used in a study that compared live and video-recorded instructions (combined with either active or passive learning) in the context of the total physical response (TPR) approach to learning a new language (Caldwell-Harris, Goodwin, Chu, & Dahlen, 2014). Two other studies with large numbers of participants (n = 230 and n = 213) were a year-long reading while listening (RWL) program study that investigated the development of VK and general proficiency in the target language (Gobel & Kano, 2014) and a study of vocabulary development through simulation games and flashcards, comparing the two TMVD conditions (Franciosi, Yagi, Tomoshige, & Suying, 2016). The majority of the studies were conducted with adult participants (n = 64, 78%), mostly university students; 7 studies used adolescents and 11 studies, children. Intermediate proficiency participants were used in 23 studies, beginner or elementary level participants in 15 studies, 14 studies used multiple proficiency levels, and 22 studies (27%) did not specify the target language proficiency of their participants.

Participants' proficiency was measured using standardized language proficiency tests: TOEFL (n = 11), CERF (n = 10), TOEIC (n = 6), Cambridge English Proficiency tests (n = 2), Oxford English Placement Test (n = 1); and VK tests: The Peabody Picture Vocabulary Test (n = 3) and Vocabulary Levels Test

(Schmitt, Schmitt, & Clapham, 2001). Eleven studies used other approaches to evaluate participants' vocabulary proficiency: self-report; in-house-developed, national or commercially available tests; and 36 studies (60% of the articles that reported participants' proficiency) did not specify which measures were used.

The greatest number of studies was conducted with Chinese (mostly, Mandarin Chinese; n = 19), Japanese (n = 11), and English (n = 11) speakers; the next most common L1s were Farsi/Persian (n = 9) and Spanish (n = 6). EFL was, by far, the most frequently used target language (n = 55, 67%), with ESL being a distant second (n = 12, 15%); there were three studies with German as a foreign language (two of which were by the same researcher), two studies with Chinese, Japanese, and Spanish (each), and one study with French, Hungarian, Patani Malay, Russian, Samoan, and Turkish (each).

The number of vocabulary items per study ranged between 5 and 156, with 38 items used on average (*SD* = 34). The studies also varied in whether the target items were the same for all participants (and selected prior to the teaching/learning intervention) or whether relevant items were mined from the study materials or interactions and selected for each study participant individually. Twenty-four articles reported no item numbers, mostly when measures of vocabulary development were not based on specific words or multi-word expressions, e.g., when lexical diversity or sophistication was measured. Treatment and study durations ranged from one-off experimental or class sessions to weeks- and months-long studies. The reported treatment duration was less than 1 hour in 18 studies, between 1 and 120 hours in 38 studies, and longer than 1 week in 26 studies.

Fifty studies used a pre- and post-test design (61%), 29 studies (35%) included a delayed post-test and 34 studies (42%) had a control group (30 of which provided some details about the control treatment). Experimental or quasi-experimental design was used the most (n = 63, 77%), either exclusively or combined with other approaches; the remaining 19 studies were observational, exploratory, or both. Over half of the studies used only quantitative data (n = 50), 28 studies used a combination of qualitative and quantitative data, and 4 studies used only qualitative data. The following instruments were used in the qualitative data analyses across the 34 studies: analysis of the participants' linguistic outputs, i.e., writing, interactions, self-report, recasts, language related episodes (LREs; n = 12); questionnaires and surveys with open-ended questions (n = 14), interviews (n = 11), observations and field notes (n = 6). Across the qualitative data analyses, no common methodological frameworks or approaches were used, even when similar data types were analyzed. Quantitative data analyses (reported in 78 articles) predominantly compared group means (using ANOVA, MANOVA, t-tests, Mann-Whitney

tests or Wilcoxon t-tests) (n = 54, 69%); other statistical analysis types were: various regression analyses (n = 11, 14%), ANCOVA (n = 7, 9%), correlational analyses (n = 6, 8%), principle component analysis (n = 1), and Pearson Chi-square analysis (n = 1). Five studies reported only descriptive statistics or percentage correct. Information about statistical modeling approaches was rarely available.

Among instructional, learner, and item variables, instructional features were the most frequent primary-interest variable type (n = 70); i.e., different TMVD treatments were compared with each other, technology-mediated learning was compared with learning without technology, comparisons were drawn between different types of in-treatment support, teacher-mediated and independent treatments, technology-mediated treatment and no treatment, and between different task types. The second most common variable of interest was learner characteristics (including proficiency and gender) and learner in-task behavior (n = 31). The third most common variable was related to the time component (n = 28), with time-on-task, amount of access, or frequency of occurrence considered 12 times, and pre- and post-test results 16 times. The final variable of interest pertained to the types of target items.

TMVD Analysis

The majority of the reviewed studies (n = 67, 82%) had vocabulary development as the primary focus of the investigation. In the remaining 15 articles, vocabulary development was a component of a broader research focus, such as written or spoken CMC (e.g., writing development in technology-mediated contexts; giving, receiving, and noticing feedback; speaker alignment in interactions; experiential learning) and general language development through CALL (e.g., in digital storytelling, multimedia reading experiences, task-based and situated language learning). Multi-word units were investigated in 16 studies (20%).

The reviewed articles can be broadly grouped into those that compared the effect of using technology-mediated learning with instructional methods that did not use technology (n = 48), those that investigated different instructional (and learning) approaches within computer-mediated learning (n = 26), and those that examined both (n = 8). Studies that used CMC, DDL, digital reference tools, digital video-based learning, or mobile learning tended to have a better balance of the two designs, while CALL software studies and digital game-based learning tended to primarily use the former design approach.

Learning and Instructional Approaches

In 60 studies (73%), participants engaged in technology-mediated learning without direct instructor mediation; 11 studies investigated instructor-mediated

TMVD and, in 11 studies, instructors were involved in some but not all treatment procedures. In studies that used independent TMVD with a control group, experimental groups[2] were compared with (a) independent learning without technology (n = 10), (b) a different technology-mediated approach (n = 8), (c) teacher-mediated learning without technology (n = 5) or (d) no learning (n = 2), in the control group. Examples of using different independent TMVD treatments include those where experimental groups received some form of additional language support, e.g., meaning-focused engagement with newscasts, with and without additional language-focused activities (Marefat & Hassanzadeh, 2016), or playing a digital game with and without access to word definitions (Alghamdi, 2016), and studies that compared two different types of technologies, e.g., vocabulary development from playing a digital simulation game and that from a flashcard game (Franciosi et al., 2016). When experimental groups used teacher-plus-technology-mediated learning, and there was a control group, the most common comparison was drawn between treatments that used technology and those that did not (aka "traditional" approach). Alvarez-Marinelli et al. (2016), for example, compared two experimental CALL curricula with an established teacher-only national English language development curriculum.

Although most of the TMVD treatments did not involve teacher mediation, 60 studies (n = 73%) were conducted in class or in a computer lab; plus, in 5 cases, a part of the treatment was in class, i.e., only 16 studies investigated TMVD outside the formal educational settings. The latter were studies that incorporated mobile-assisted language learning (e.g., Amer, 2014; Andujar, 2016; Stockwell, 2010; Terantino, 2016), intercultural exchanges (e.g., Lin, 2016) and other CMC studies (e.g., Kozar, 2016; Yilmaz & Yuksel, 2011), and longitudinal studies (e.g., Chukharev-Hudilainen & Klepikova, 2016; Gobel & Kano, 2014) or studies of extramural language use (e.g., Jensen, 2017; Sylven & Sundqvist, 2012).

In the four-strands analysis, treatments that incorporated all four learning strands were used only in 3 studies and those incorporating three strands were used in 18 studies, two strands were used in 37 studies, and one strand, in 24 studies (15 studies used LF and 9 used MFI). The studies that included all four strands were observational and longitudinal: a 25-week-long CALL curriculum study by Alvarez-Marinelli et al. (2016), a six-months Andujar (2016) study that investigated the effect of mobile instant messaging (using WhatsApp) on the accuracy and complexity of students' writing, and DeHaan, Johnson, Yoshimura, and Kondo (2012) who investigated the effect of using wikis on the development of oral communication skills in a four-week teaching module. In terms of specific strands (used individually or in combination), MFI was used in 64 studies, MFO in 27 studies, LF in 61 studies, and FD in

11 studies. The most frequent combination of strands was MFI and LF found in 44 studies, MFI and MFO were combined in 25 studies, and the combination of MFI, MFO, and LF was used in 14 studies. These results show that input-driven learning, frequently combined with language-focused activities, is the treatment of choice in TMVD studies. It also appears that opportunities for fluency development are not prioritized in the reviewed studies; there was only one study (Cobb & Horst, 2011) that explicitly emphasized fluency development in the treatment description (in the form of "lexical access speed practice"). In addition to the four-strands analysis, the articles were examined using an alternative four-skills categorization, familiar to many language teachers. The results were similar to the four-strands analysis; the number of language activities involving receptive skills was twice as high (reading: n = 16; listening: n = 11) as activities focusing on productive skills (speaking and writing: n = 7 each).

In the studies that used a control group, the four-strands analysis showed 17 instances of the same strands being used in the experimental and control groups, and 12 instances when more learning strands were used in the experimental groups than in the control group. Vyatkina (2016), for example, used DDL for teaching lexico-grammatical constructions in German; the control group followed a deductive presentation-practice-production approach using rule-based textbook exercises with a pair discussion at the end (MFO + LF), while the experimental group engaged in the illustration-interaction-induction learning, using concordance lines from a native-speaker corpus (MFI + MFO + LF).

In the incidental-deliberate learning analysis, the treatments were coded into two categories, each containing two subcategories: IL−S (e.g., reading without access to reference tools) and IL+S (reading with access to dictionaries or glosses), and DL−C (e.g., word lists or picture-word learning treatments) and DL+C (e.g., word learning with concordance lines). The two learning approaches were used in a similar number of studies: deliberate learning (n = 38), incidental learning (n =37), and seven studies used both incidental and deliberate learning. Within deliberate learning, 12 treatments were coded as DL−C and 35 treatments as DL+C; within incidental learning, 13 treatments were IS−S and 33 treatments as IL+S (some studies contained more than one category or subcategory). These results show that the majority of TMDV treatments presented vocabulary in context and that incidental learning treatments were, typically, accompanied by supplementary activities aimed at increasing learning opportunities.

Two main types of studies that used IL+S treatments were reading studies with access to reference tools, such as digital glosses and dictionaries (e.g., Hu, Vongpumivitch, Chang, & Liou, 2014; Türk & Erçetin, 2014) and video-based

activities with different types of captioning, subtitling, or audio support (e.g., Mohsen, 2016; Montero Perez, Peters, Clarebout, & Desmet, 2014; Sydorenko, 2010). Another group of studies that used IL+S treatments was CMC studies with peer- or instructor-generated feedback (e.g., Diez-Bedmar & Perez-Paredes, 2012; Dizon, 2016). Dizon (2016), for example, investigated Facebook-mediated writing development, using focused freewriting, in which participants received peer feedback. In addition, a number of CALL software and digital game-based studies incorporated both incidental learning procedures and supplementary materials that facilitated language learning; for example, Shintaku (2016) explored the interplay of game design and pedagogical mediation for learners of Japanese using a commercial off-the-shelf game with researcher-developed activity worksheets and a reference list. Another example is the French Digital Kitchen project; Seedhouse et al. (2013) created a task, in which participants cooked a dish in a real kitchen using utensils with built-in digital sensor technology, following verbal instructions in their L2 (French), with access to L1 translations and repetitions of instructions on request.

When a control group was used, the experimental and control groups tended to use the same learning approach (n = 20). However, in six studies, the category ("incidental" or "deliberate") was matched between the experimental and control groups, but the subcategories differed. For example, in a study that examined the effect of playing a simulation game on vocabulary retention (Franciosi et al., 2016), the treatment group used an online vocabulary learning application (Quizlet) in conjunction with playing the 3rd World Farmer game (DL+C), while the control group only used Quizlet (DL−C). In three cases, the experimental group experienced a treatment in a different learning category than the control group; e.g., the experimental group engaged in IL+S using a video game, while the control group used text-drill materials that contained stories adapted by the researchers from the game's plot (DL+C; Vahdat & Rasti Behbahani, 2013).

Aspects of Word Knowledge

The knowledge of spoken form was evaluated 18 times, written form 39 times, word parts 6 times, form-meaning mapping was examined 57 times, the knowledge of semantic and associative dimensions of meaning 13 times, and knowledge of use 35 times. This shows that, in the TMVD research, form–meaning mapping is the most frequently measured aspect of vocabulary development. This is aligned with an accepted understanding of form–meaning mapping as a core of VK (Nation, 2001; Schmitt, 2008). The knowledge of written form was the second most-frequently evaluated aspect and the knowledge of use was the third. The latter finding, however, needs to be considered with caution because the coding of the knowledge of use comprised

grammatical function, collocational knowledge, and constraints on use, which may have contributed to its relative prominence among other developmental measures. A closer examination suggests that over half of the KoU codes were associated with the measures of collocational aspect of word knowledge, which is a growing area of interest in second language vocabulary research (Schmitt, 2004; Wray, 2002).

The number of knowledge aspects evaluated per study varied from 1 to 5, with an average of 2 and a standard deviation of 0.89. Studies that incorporated real language use tended to cover more knowledge aspects (e.g., five aspects: Milton et al., 2012; four aspects: Andujar, 2016; Rusanganwa, 2013; Seedhouse et al., 2013). Four aspects of knowledge were also evaluated in a study that used self-report measures (Esit, 2011) and in a longitudinal study that used a battery of observational and empirical tests with school children (Cobb & Horst, 2011). Most studies evaluated two aspects of knowledge (n = 43), of which WF+FM (44%) was the most frequent combination, followed by SF+FM (14%), FM+KoU (12%), and WF+KoU and SAM+KoU (9% each). WF+FM was also the most common combination across all aspects of knowledge measured (n = 30). The following frequencies were recorded for the tri-aspect combinations: SF+WF+FM (n = 7), WF+FM+KoU (n = 5), WF+SAM+KoU (n = 4), FM+SAM+KOU (n = 4), WP+SAM+KoU (n =3), WF+WP+KoU (n = 3), WF+FM+SAM (n = 3), SF+FM+KOU (n = 3). These results point to a tendency to focus on written language and on basic form–meaning mapping as a measure of vocabulary development, rather than on the quality of L2 semantic and associative connections (lexical-semantic networks) that is considered to be indicative of more mature stages of L2 vocabulary development (Barcroft, 2016). In addition, the prevalence of measures focused on written form over spoken form may be a reflection of the nature of the target language being investigated (i.e., formal academic register) since the reviewed studies were mostly conducted with university or school students in the classroom or laboratory settings.

Measures of Vocabulary Development

First, author-supplied descriptions of measures of vocabulary development were examined, followed by the TMVD themes analysis (Appendix A). The following types of data were used (separately or in combination with each other) to evaluate vocabulary development: vocabulary tests of individual words or multi-word units (n = 69), learner outputs, including text, speech or conversation analyses (n = 15), listening and reading comprehension tests (n = 12), questionnaires (mostly, evaluating participants' attitudes and perceptions) (n = 11), observations (n = 6), fluency measures (n = 6), language proficiency tests (n = 4), interviews (n = 3), stimulated recall (n = 2), time-on-task (n = 2), learner diaries (n =

1), grammaticality judgments (n = 1), and eye-movement measures (n = 1). The most frequently used vocabulary tests were translation, multiple-choice, cloze or fill-in-the blanks, self-report (e.g., using vocabulary knowledge scales), word–picture matching, definitions matching, and sentence or text writing.

The results of the TMVD coding analysis of learning measures are provided in Appendix B, Table 3. The TMVD studies overwhelmingly measured quality (or depth) of specific aspects of VK, rather than the development of the participants' L2 lexicon (i.e., their vocabulary size). Only five studies measured participants' vocabulary size, as an indicator of vocabulary development, and three studies measured both. The number of studies that used receptive and productive measures of vocabulary development were very similar: productive (n = 29), receptive (n = 25), and a combination of receptive and productive measures (n = 28). Receptive measures were used more frequently than productive measures in CALL software studies. Productive measures were used more frequently in CMC, DDL, and studies using mobile-assisted learning and social media tools. The highest proportion of the combination of receptive and productive measures was used in studies researching digital reference tools. Digital video-based and game-based studies had an even distribution of receptive, productive and combination (receptive and productive) measures. The same number of studies used TMVD measures in-context and out-of-context, with 18 studies using both approaches. Offline (n = 61) and explicit (n = 68) measures were used much more frequently than online (n = 15) and implicit measures (n = 1). A combination of offline and online measures was used six times, and that of explicit and implicit measures[3] twelve times.

Combining Different Types of Measures

Productive measures were mostly used in context (77%), while receptive measures were used both out of context and in context. Productive knowledge was measured offline in 19 cases, online in 9 cases, and once using both online and online procedures. Receptive knowledge was measured offline 21 times and online 4 times. When receptive and productive measures were combined, they were most often used in offline evaluation procedures (n = 21). Implicit measures were used exclusively in productive procedures, with vocabulary presented in context. None of the interdisciplinary methods used in L2 vocabulary research (such as response times, priming manipulations, or eye-tracking) were used in the TMVD studies to measure vocabulary development.

Reported Study Results

The reported results were based either on the analysis of learner language data (test scores, self-ratings, analyses of output) or on the analysis of participants' attitudes to technology-mediated learning treatments, usually obtained through

surveys or interviews. In the analyses of learner language data, the reported results were overwhelmingly positive, with only ten instances where negative or no effect results were reported without any positive effects, and a further six instances of negative or no effect results reported alongside a positive effect. In the analyses of the attitudinal data (n = 16) the following was found: positive results (n = 9), both positive and negative (n = 5), negative (n = 1), and no response polarity (n = 1, i.e., an analysis of participants' preferences for different types of glossing). Studies comparing learning with and without technology reported a positive effect of TMDV much more often (96%) than studies that compared different technology-mediated learning approaches (46%). This may be due to the nature of the research questions investigated in the latter type of studies, which focused as much on *how* different technology-mediated approaches affected the development of different aspects of vocabulary knowledge, as on *whether* one of the approaches resulted in more or better learning. Studies that investigated more than one aspect of vocabulary knowledge were likely to provide a richer set of findings. Milton et al. (2012), who looked at learner behavior in the Vill@ge virtual learning environment in Second Life, found that, although opportunities for lexical development outside of controlled learning activities were poor, there was no difference between traditional learning and that in the virtual word in vocabulary focused activities. They also found a positive effect of activities in Second Life on fluency development. Studies that examined multiple variables, also produced richer findings. Rouhi and Mohebbi (2013), for example, examined the relationship between learner characteristics (measured by a multiple intelligences survey instrument) and learning achievement when using different types of vocabulary glosses (pictorial, audio and video) available in their in-house-developed software, Scaffoglossing. Contrary to their predictions, they did not find any added advantage of multimedia glosses for the high compared to the low spatial ability group. They also compared technology-mediated learning with that of a control group that used a course textbook, showing that vocabulary development was significantly greater for the former. Combining language data with participants' attitudes towards technology-mediated approaches can also add another dimension to the findings. Gordani (2013) investigated the effect integrating an online corpus-based learning into reading comprehension classes and found that it had a positive effect on vocabulary development, compared to the textbook only approach. However, the analysis of learner diaries revealed both positive and negative reactions to the use of corpora that changed over time.

Discussion and Recommendations

The main goal of this article was to review methodological aspects of recent TMVD studies in terms of (1) learning and instructional approaches, (2)

aspects of vocabulary knowledge being evaluated, and (3) measures used to evaluate vocabulary knowledge development. The findings for the general methodological attributes showed robust participant numbers (67 on average per study) and a good range of study durations, with about a third of studies taking a longitudinal approach to the data collection. The use of control groups in 42% of the studies and the inclusion of delayed post-tests in 35% of the studies also attest to a certain level of maturity of TMVD research design. The focus on multi-word units in 20% of the studies is another encouraging finding that points to an alignment with a similar trend in the L2 vocabulary research.

However, there are a number of methodological aspects that need further work. More than a quarter of the studies did not provide information about participants' target-language proficiency. Research into L2 vocabulary acquisition shows that vocabulary development progresses differently for low-and high-proficiency learners, even when level-appropriate L2 input is used. The differences in the developmental trajectories are associated with quality of lexical representations (Perfetti, 2007) and quality of L2 lexical-semantic networks, at different proficiencies (Elgort et al., 2015; Elgort & Piasecki, 2014). This means that learning new L2 words and retaining this knowledge is a more time-consuming and effortful endeavor for lower proficiency learners, and different instructional and learning approaches may be more appropriate at different proficiencies. Without information about general language proficiency or the size of the participants' L2 lexicon, effects of learning treatments are difficult to interpret, and replication studies become problematic.

TMVD research continues to compare technology-mediated learning (both teacher-mediated and independent) with learning without technology, and finding overwhelmingly in favor of the former. This line of research is not particularly interesting or informative, partly because online and digital technologies and mobile devices are now an intrinsic component of our lives and, as such, they are being used by language learners, whether or not they are also used in the language classroom. Another problem with this type of research is that the comparison made is not at all straightforward; very rarely do we find studies where the learning and instructional procedures are exactly the same, plus/minus technology. Introducing technology into a treatment changes the nature of the treatment itself. A more fruitful approach would be to investigate language learning affordances of technology-mediated learning and instructional treatments, in order to fine-tune TMVD treatments for specific learner populations, learning goals, or targeting specific aspects of vocabulary knowledge.

Building a lexicon in a second language requires time and repeated exposure to level-appropriate input, opportunities for meaning-focused output,

language focused learning, and fluency development (Nation, 2007). This review shows that the majority of TMVD studies build upon the strands of meaning-focused input and language-focused learning and, to a lesser degree, on meaning-focused output. Not enough attention, however, is devoted to fluency development in the design of instructional or learning treatments. This imbalance is especially problematic for EFL learners, who constituted a large proportion of the present sample (67%). More technology-mediated treatments should be devised to create opportunities for language learners to improve fluency of vocabulary processing and retrieval in their target languages.

Technology-mediated vocabulary development affords independent learning opportunities to language learners who can set learning goals, know how to make the most of various TMVD contexts, are able to receive, notice, and interpret feedback, and have motivation to learn. The present review shows that about two thirds of the studies provided opportunities for learners to engage in independent TMVD, and this is encouraging. However, only 20% of the studies investigated learning outside of the language classroom or laboratory environment, and this is a concern. There is nothing wrong with experimental laboratory studies, conducted under highly controlled conditions when the research goal is to zoom in on a specified variable of interest. However, this is not why the majority of the present studies were conducted in the computer laboratory. In many cases, it appears to be the matter of convenience and constraints. Furthermore, the studies are typically with university or school students, in formal educational contexts. This is problematic if the personal learning goals of the language learners are not related to academic study.

If a key advantage of TMVD treatments is, indeed, a capacity for independent learning, researchers need to start collecting data about vocabulary development that takes place outside of the language classroom, using longitudinal and observation approaches that, as this review shows, are able to cover a wider range of learning strands and vocabulary knowledge aspects. With a growing number of online multi-user language learning environments, digital gaming environments, social media tools, digital communication platforms, mobile devices, digital text, and audio and video resources with access to dynamic reference tools available to language users and learners around the world, more TMVD research is needed to trace patterns of L2 vocabulary development in these authentic, independent learning contexts.

Vocabulary development measures adopted in the reviewed studies were, typically, explicit vocabulary knowledge tests of individual items, conducted offline. Although both receptive and productive measures were used in the studies, a notable dearth of online and implicit measures makes it impossible to

predict whether vocabulary knowledge tested in these studies will be available in real language use. The only type of measure coded as implicit in this review was learner output in naturalistic production or interaction. These findings point to two important limitations of the current TMVD research. Firstly, in focusing primarily on individual target vocabulary items, patterns of development of the L2 lexicon in TMVD studies remain unclear. Secondly, failing to incorporate measures from interdisciplinary L2 vocabulary research, TMVD studies are generating findings on the development of declarative vocabulary knowledge, but not on the functional knowledge, which weakens the usefulness of these findings.

The key recommendations made for future TMVD studies, based on the review findings, are:

1. researchers should estimate and report participants' L2 proficiency and the measures used to estimate it;
2. experimental and pseudo-experimental TMVD research should move away from comparing technology-mediated treatments with those that do not use technology, and towards investigating the effects of different technology-mediated approaches on L2 vocabulary development;
3. TMVD studies should design and evaluate treatments that target fluency development;
4. TMVD studies should use vocabulary size measures as an indicator of vocabulary development overtime;
5. TMVD studies should use online and implicit measures of vocabulary development, alongside offline and explicit measures;
6. TMVD research needs to start using data from distributed technology-mediated language learning and communication environments that are used by thousands of users around the word, in order to reveal patterns of vocabulary development outside of the language classroom.

Conclusion

This review has generated rich descriptions that help us identify trends in the TMVD research methods and practices. It shows that L2 technology-mediated vocabulary learning and instruction is a vibrant research domain, which covers a wide range of technology-mediated treatments, investigates many aspects of vocabulary knowledge, and uses diverse qualitative and quantitative research approaches. The identified limitations have led to a set of recommendations for future TMVD research that could further improve the treatment and research design, make TMVD studies more robust and suitable for replications, and generate findings that would more readily generalize to real technology-mediated language use.

Notes

1.   Inconsistent spelling (pre-test, pretest, post-test, posttest, pre- and post-test, pre-post-test) in the abstracts was unified as "pre test" and "post test", for consistency.

2.   Here and elsewhere in this article, "experimental group" is used to refer to the group that experienced the technology-mediated treatment of interest (and that was compared with the control group), i.e., it refers to the "treatment group" in observational and exploratory studies.

3.   Analyses of the learner output in naturalistic free production and interaction tasks were coded as "explicit + implicit".

About the Author

Irina Elgort is Senior Lecturer in Higher Education at Victoria University of Wellington. Her research interests include computer-assisted language learning and educational technology, lexical development in a second or foreign language, the bilingual mental lexicon, and reading. She teaches a Master's level course in computer-assisted language learning.

References

Abraham, L. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning, 21*(3), 199–226. https://doi.org/10.1080/09588220802090246

Alghamdi, E. A. (2016). Untangling multimedia effects on EFL incidental vocabulary learning via playing an online hidden-object game. *IJCALLT, 6*(1), 24–39. https://doi.org/10.4018/IJCALLT.2016010102

Alvarez-Marinelli, H., Blanco, M., Lara-Alecio, R., Irby, B. J., Tong, F., Stanley, K., & Fan, Y. (2016). Computer assisted English language learning in Costa Rican elementary schools: An experimental study. *Computer Assisted Language Learning, 29*(1), 103–126. https://doi.org/10.1080/09588221.2014.903977

Amer, M. (2014). A study of learners' usage of a mobile learning application for learning idioms and collocations. *CALICO Journal, 31*(3), 285–302. https://doi.org/10.11139/cj.31.3.285-302

Andujar, A. (2016). Benefits of mobile instant messaging to develop ESL writing. *System, 62*, 63–76. https://doi.org/10.1016/j.system.2016.07.004

Barcroft, J. (2016). *Vocabulary in language teaching*. New York: Routledge.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://doi.org/10.1111/lang.12224

Caldwell-Harris, C., Goodwin, K., Chu, E., & Dahlen, K. (2014). Examining the advantage of a live instructor vs. video in a laboratory study. *Innovation in Language Learning and Teaching, 8*(3), 191–204. https://doi.org/10.1080/17501229.2013.793690

Chiu, Y. (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

Chukharev-Hudilainen, E., & Klepikova, T. (2016). The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *CALICO Journal, 33*, 334–354.

Cobb, T., & Horst, M. (2011). Does Word Coach coach words? *Calico Journal, 28*(3), 639–661. https://doi.org/10.11139/cj.28.3.639-661

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing, 29*(2), 243–263. https://doi.org/10.1177/0265532211419331

deHaan, J., Johnson, N.H., Yoshimura, N., & Kondo, T. (2012). Wiki and digital video use in strategic interaction-based experiential EFL learning. *CALICO Journal, 29*(2), 249–268. https://doi.org/10.11139/cj.29.2.249-268

Diez-Bedmar, M. B., & Perez-Paredes, P. (2012). The types and effects of peer native speakers' feedback on CMC. *Language Learning and Technology, 16*, 62–90.

Dizon, G. (2016). A comparative study of Facebook vs. paper-and-pencil writing to improve L2 writing skills. *Computer Assisted Language Learning, 29*(8), 1249–1258. https://doi.org/10.1080/09588221.2016.1266369

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Elgort, I. (in press). Teaching/developing vocabulary using ICTs and digital resources. In J. I. Liontas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118784235

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 1–26. https://doi.org/10.1017/S0272263117000109

Elgort, I., & Nation, I. S. P. (2010). Vocabulary learning in a second language: Familiar answers to new questions. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.), *Conceptualising 'learning' in applied linguistics* (pp. 89–104). Houndmills, UK: Macmillan. https://doi.org/10.1057/9780230289772_6

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience, 30*(5), 506–528. https://doi.org/10.1080/23273798.2014.942673

Elgort, I., & Piasecki, A. (2014). The effect of a bilingual learning mode on the establishment of lexical-semantic representations in the L2. *Bilingualism: Language and Cognition, 17*(3), 572–588. https://doi.org/10.1017/S1366728913000588

Esit, O. (2011). Your verbal zone: An intelligent computer-assisted language learning program in support of Turkish learners' vocabulary learning. *Computer Assisted Language Learning, 24*(3), 211–232. https://doi.org/10.1080/09588221.2010.538702

Franciosi, S., Yagi, J., Tomoshige, Y., & Suying, Y. (2016). The effect of a simple simulation game on long-term vocabulary retention. *CALICO Journal, 33*, 355–379.

Gobel, P., & Kano, M. (2014). Implementing a year-long reading while listening program for Japanese university EFL students. *Computer Assisted Language Learning, 27*(4), 279–293. https://doi.org/10.1080/09588221.2013.864314

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

Goodfellow, R. (1995). A review of the types of CALL programs for vocabulary instruction. *Computer Assisted Language Learning, 8*(2–3), 205–226. https://doi.org/10.1080/0958822940080206

Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a foreign language learners' vocabulary development. *Computer Assisted Language Learning, 26*(5), 430–445. https://doi.org/10.1080/09588221.2012.685078

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303–317. https://doi.org/10.1017/S0272263199002089

Hu, S., Vongpumivitch, V., Chang, J., & Liou, H. (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

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

Jensen, S. H. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal, 34*(1), 1–19. https://doi.org/10.1558/cj.29519

Kember, D. (2003) To control or not to control: The question of whether experimental designs are appropriate for evaluating teaching innovations in higher education. *Assessment and Evaluation in Higher Education, 28*(1), 89–101. https://doi.org/10.1080/02602930301684

Kozar, O. (2016). Text chat during video/audio conferencing lessons: Scaffolding or getting in the way? *CALICO Journal, 33*(2), 231–259. https://doi.org/10.1558/cj.v33i2.26026

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

Lin, H. (2015). Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL, 27*(3), 261–287. https://doi.org/10.1017/S095834401400041X

Lin, Y. (2016). Development of multi-word sequences by adolescent EFL Learners through online interaction: Does online contact with native English speakers lead to a more native-like use of multi-word sequences? *English Today, 32*(4), 27–32. https://doi.org/10.1017/S0266078416000201

Lindstromberg, S. (2016). Inferential statistics in *Language Teaching Research*: A review and ways forward. *Language Teaching Research, 20*(6), 741–768. https://doi.org/10.1177/1362168816649979

Ma, Q. (2009). *Second language vocabulary acquisition*. Bern, Switzerland: Peter Lang.

Marefat, F., & Hassanzadeh, M. (2016). Applying form-focused approaches to L2 vocabulary instruction through podcasts. *Language Learning and Technology, 20*, 107–127.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*, 703–704. https://doi.org/10.1038/nn1264

Michel, M. C., & Smith, B. (2017). Eye-tracking research in computer-mediated language learning. In S. Thorne and S. May (Eds.), *Language, education and technology* (pp. 453–464). (Encyclopedia of language and education). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02237-6_34

Milton, J., Jonsen, S., Hirst, S., & Lindenburn, S. (2012). Foreign language vocabulary development through activities in an online 3D environment. *The Language Learning Journal, 40*(1), 99–112. https://doi.org/10.1080/09571736.2012.658229

Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning, 29*(7), 1220–1237. https://doi.org/10.1080/09588221.2016.1210645

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology, 18*, 118–141.

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

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

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9–13.

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 2–13. https://doi.org/10.2167/illt039.0

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition, 35*(4), 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics, 36*, 73–97. https://doi.org/10.1017/S0267190516000015

Plonsky, L., & Ziegler, N. (2016). The CALL–SLA interface: Insights from a second-order synthesis. *Language Learning and Technology, 20*, 17–37.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146–161. https://doi.org/10.1017/S0267190504000078

Rouhi, A., & Mohebbi, H. (2013). Glosses, spatial intelligence, and L2 vocabulary learning in multimedia context. *3L: The Southeast Asian Journal of Language Studies, 19*, 75–87.

Rusanganwa, J. (2013). Multimedia as a means to enhance teaching technical vocabulary to physics undergraduates in Rwanda. *English for Specific Purposes, 32*(1), 36–44. https://doi.org/10.1016/j.esp.2012.07.002

Sauro, S., & Smith, B. (2010). Investigating L2 performance in text chat. *Applied Linguistics, 31*(4), 554–577. https://doi.org/10.1093/applin/amq007

Schmitt, N. (Ed.) (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.9

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

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

Seedhouse, P., Preston, A., Olivier, P., Jackson, D., Heslop, P., Plötz, T., … Ali, S. (2013). The French digital kitchen: Implementing task-based language teaching beyond the classroom. *International Journal of Computer-Assisted Language Learning and Teaching, 3*(1), 50–72. https://doi.org/10.4018/ijcallt.2013010104

Shintaku, K. (2016). The interplay of game design and pedagogical mediation in game-mediated Japanese learning. *International Journal of Computer-Assisted Language Learning and Teaching, 6*(4), 36–55. https://doi.org/10.4018/IJCALLT.2016100103

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language, Learning and Technology, 16*, 53–81.

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning and Technology, 14*, 95–110.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology, 14*, 50–73.

Sylven, L. K., & 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

Terantino, J. (2016). Examining the effects of independent MALL on vocabulary recall and listening comprehension: An exploratory case study of preschool children. *CALICO Journal, 33*, 260–277.

Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1–25. https://doi.org/10.1080/09588221.2012.692384

Vahdat, S., & Rasti Behbahani, A. (2013). The effect of video games on Iranian EFL learners' vocabulary learning. *Reading Matrix, 13*, 61–71.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning and Technology*, *20*, 159–179.

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

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Yilmaz, Y., & Yuksel, D. (2011). Effects of communication mode and salience on recasts: A first exposure study. *Language Teaching Research, 15*(4), 457–477. https://doi.org/10.1177/1362168811412873

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*(1), 39–58. https://doi.org/10.1080/09588221.2010.523285

Appendix A: TMDV Codes by Themes

Theme	Category	Code
Aspect of word knowledge being investigated (measured)		
	Spoken form	SF
	Written form	WF
	Word parts/morphology	WP
	Form–meaning mapping	FM
	Semantic and associative dimensions of meaning	SAM
	Knowledge of use, comprising grammatical function, collocational knowledge and constraints on use	KoU
Measures used to evaluate vocabulary development		
	Vocabulary size	S
	Quality of knowledge	Q
	Receptive	R
	Productive	P
	Offline	OFL
	Online	ONL
	Explicit	EXPL
	Implicit	IMPL
	In-context	IC
	Out-of-context	OC
Learning and instructional approaches		
	Teacher-mediated learning (+/− technology)	TCH
	Independent technology-mediated learning	TEC
	In-class/lab	ICL
	Out-of-class/lab	OCL
	Meaning-focused input	MFI
	Meaning-focused output	MFO
	Language-focused learning*	LF
	Fluency development	FD

Theme	Category	Code
	Incidental learning without support	IL−S
	Incidental learning with additional support	IL+S
	Deliberate learning out of context†	DL−C
	Deliberate learning in context	DL+C

* includes the use of referencing tools during reading.
† context = text or speech in which target items are embedded.

Appendix B

Table 1
Journals in which Selected Articles Were Published

Journals	No.
3L: The Southeast Asian Journal of English Language Studies	1
Applied Linguistics	1
Computer-Assisted Language Learning and Teaching	1
Computers in Human Behavior	1
Educational Media International	1
English for Specific Purposes	1
English Language Teaching	1
English Today	1
Innovation in Language Learning and Teaching	1
International Journal of Research Studies in Language Learning	1
Journal of Early Childhood Literacy	1
Language Learning Journal	1
Novitas ROYAL (Research on Youth and Language)	1
System	1
The Reading Matrix	1
Asian EFL Journal	2
Language Teaching Research	2
International Journal of Computer-Assisted Language Learning and Teaching	4
Language Learning & Technology	10
ReCALL	13
CALICO Journal	17

Table 2
Top 10% Most Frequent Words Used in Article Keywords, Provided by the Authors

Frequency	Word	% of word types	Frequency	Word	% of word types
77	learning	0.215	8	design	0.022
67	language	0.187	8	driven	0.022
47	vocabulary	0.131	8	mediated	0.022
24	second	0.067	8	teaching	0.022
21	computer	0.059	7	corpus	0.020
14	assisted	0.039	7	EFL	0.020
14	based	0.039	7	foreign	0.020
13	acquisition	0.036	7	instruction	0.020
12	comprehension	0.034	7	mobile	0.020
11	CALL	0.031	7	online	0.020
10	multimedia	0.028	6	digital	0.017
9	English	0.025	6	gloss	0.017
9	L2/L1	0.025	6	learner	0.017
9	reading	0.025	6	listening	0.017
9	video	0.025	6	subjects	0.017
8	communication	0.022	6	writing	0.017
8	data	0.022			

Table 3
Analysis of TMVD Measures Used to Evaluate Vocabulary Development

Types of measures	Coded measures	No.
Size/Quality	Vocabulary size	5
	Vocabulary knowledge (quality)	69
	Vocabulary size and quality	3
	NA*	5
Receptive/Productive	Receptive	25
	Productive	29
	Receptive and productive	28
In-context/Out-of-context	In-context	31
	Out-of-context	31
	In- and out-of-context	18
	NA	2

Types of measures	Coded measures	No.
Offline/Online	Offline	61
	Online	15
	Offline and Online	6
Explicit/Implicit	Explicit	68
	Implicit	1
	Explicit and implicit	12
	NA	1

* NA is used when insufficient information has been provided to assign a code, or where the type of code did not apply.