ACQUIRING ACADEMIC VOCABULARY FOR POST-SECONDARY ENGLISH LEARNERS: A RESEARCH SYNTHESIS

A Dissertation

by

SUNNI LYNN SONNENBURG WINKLER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Zohreh R. Eslami
Co-Chair of Committee,	Li-Jen Kuo
Committee Members,	Janet E. Hammer
	Yolanda N. Padrón
Head of Department,	Michael A. de Miranda

December 2019

Major Subject: Curriculum and Instruction

ABSTRACT

This research synthesis describes, critiques, and synthesizes intervention studies related to the academic vocabulary acquisition of postsecondary English learners. Using the article matrix and constant-comparative methods of analysis, this review includes all relevant studies so as to provide a knowledge base for both researchers and practitioners moving forward. Descriptive findings indicate the need for more experimental research in this area that reduces threats to validity, employs a combination of standardized and researcher-created instruments (with reported reliability), and reports detailed procedures of any interventions. The research focus up to this point has been on general academic vocabulary; further research is still needed in this area, though there is also a need for discipline-specific academic vocabulary at the post-secondary level. Findings regarding best practices include the following: a) there are significant benefits to combining embedded academic vocabulary learning with explicit, isolated word learning; b) technology is most effective when combined with other well-established aspects of quality vocabulary instruction; c) receptive vocabulary gains are highest when pairing the learning and assessment modes (i.e., receptive vs. productive) and pursuing consistency over structure; and d) specific tools and materials, such as using a concordance or dictionary, can significantly enhance productive academic vocabulary.

DEDICATION

I dedicate this work to every family member, friend, colleague, professor, and barista whose kind words, smiles, and cups of coffee sustained me throughout this process.

Specifically, I dedicate this to Matt, James, Mom, and Dad - I know the pursuit of this degree took me away from you for countless hours, and I am thankful for your graciousness in giving me both that time and your encouragement to keep going.

ACKNOWLEDGEMENTS

I would like to thank my committee for their feedback, support, and patience. My chair, Dr. Eslami, was my greatest advocate throughout my graduate program and this dissertation process. She trusted me to do good work, held me to high expectations, presented me with numerous opportunities to grow, and encouraged me to keep going when I felt I was at my end. My co-chair, Dr. Kuo, provided the most helpful, thorough feedback at every turn, making time to review page after page despite her busy schedule. She has set an amazing example for me of maintaining a highly productive work agenda that never compromises on quality. Dr. Hammer, whose name was the first to enter my mind when thinking of who I wanted to be on my committee, continues to be my example of what an incredible educator and kind human being looks like. She has shown me through her teaching and her friendship the kind of strong, compassionate woman and educator that I aspire to be. Dr. Padrón, who knew little about me prior to me asking her to serve on my committee, pushed me to think critically about the research process and how to thoughtfully conduct quality research. Her questions and feedback challenged me in a way that made my work so much stronger. I cannot say thank you enough to these professors who are educators and mentors in the truest sense of those words.

I also want to thank Dr. Cindy Boettcher for encouraging me to pursue this degree. She was one of my biggest supporters and proponents throughout this process - I can honestly say I would not be here were it not for her! Her support opened up incredible opportunities for me that have truly changed my life. Saying “thank you” seems wholly inadequate - the phrase does not convey my immense gratitude.

Thank you, too, to every professor, colleague, staff member, fellow graduate student, and undergraduate student who made my experience at Texas A&M a positive one. Specifically, I am grateful to Katherine Wright, my one-time office mate, co-researcher, running partner, and friend. She inspired and encouraged me in ways she probably does not even realize. I want to thank Dr. Patricia Wiese, who showed me the doors that can open through education and allowed me to walk through those with her. I am also thankful to Rachel Wiese for her unending encouragement, friendship, and laughter; she inspired and helped me to keep moving forward through the most challenging times.

To friends and family who worked alongside me, were my constant cheerleaders, and sent texts or emails of love, prayer, and support - thank you for knowing I needed every single one of those moments and messages and for being willing to give and send them.

Lastly, I would like to thank my parents, husband, and son. Thank you to my mom and dad, who have been with me since the beginning of this process, providing emotional and financial support along the way as well as doing more than their fair share of puppy-sitting while I sat writing at their kitchen counter. Thank you to my incredible husband, Matthew Winkler, for never faltering in his support of me. He never questioned me or tried to talk me out of this pursuit - even when my preliminary exams fell exactly one month before our wedding; he simply continued to say, “I love you and support whatever you want to do, babe.” That support has meant the world. Finally, thank you to my sweet baby boy, James, for making my heart so full and for reminding me of the truly important things in life.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Zohreh Eslami, advisor and chair, Professor Li-Jen Kuo, co-chair, and Dr. Janet Hammer of the Department of Teaching, Learning and Culture and Professor Yolanda Padrón of the Department of Educational Psychology.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

This work was not supported by any outside funding sources. The contents are solely the responsibility of the author.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

How do post-secondary English learners (ELs) acquire the academic language necessary to function in and contribute to academic contexts? In attempting to answer this question, I examined various L2 (second language) vocabulary and vocabulary instruction reviews; however, the majority of them were for K-12 learners. This may be a good place to start in understanding how English Learners acquire vocabulary, but the settings and motivations of K-12 learners are vastly different from adults. At the university and graduate level, students need academic vocabulary to complete their degrees and eventually use language to perform job-related duties (or get a job!). Professors require students to be skilled in both receptive (listening and reading) and productive (speaking and writing) academic vocabulary, but once students begin their post-secondary studies, what can they do or where can they turn to acquire that vocabulary? This situation is certainly becoming more and more common as numbers of post-secondary ELs increase each year. These questions are both theoretical and, by nature, highly practical. Academic vocabulary acquisition at the post-secondary level needs to be further explored. What do international students do to increase their academic vocabulary? What are the most effective strategies for these learners to acquire and improve academic vocabulary?

Even though international students have studied in English-speaking universities for decades (Martin, 1976), many still struggle with acquiring and using academic vocabulary. Reports from professors (Wang & Bakken, 2004), journal editors (Flowerdew, 2001), and international students themselves (Zhou, 2009) indicate a general awareness of the issue: needed improvement in international students' academic vocabularies, particularly as expressed in academic writing. As Ivanic (1998) notes, academic vocabulary and language are often not

taught; thus, the lack of explicit instruction in English academic language is unsurprising. Additionally, the expectations and norms across academic disciplines can vary widely, leaving many students frustrated as they attempt to navigate and communicate within different genres and academic contexts.

To compound the issue, the U.S. has a greater number of international students than 40 years ago, seeking a variety of degrees. From 2002 to 2009, enrollment of international students in U.S. universities increased by 13%, from 582,996 to 660,581 students (Choudaha & Chang, 2012). As of 2009, the U.S. enrolled about one-fifth of all globally mobile students (i.e., international students or students seeking degrees from a country that was not their country of origin). While a large portion of these students is seeking undergraduate degrees, those seeking Master's and Doctorate degrees are also increasing. From 2010 to 2011, international students seeking Master's degrees increased by 5.32% while the number of those seeking Doctorate degrees rose by 2.76%.

Even with the United States' increasingly strict immigration policies, the number of international students will likely continue to rise as many come from countries like China, India, and Brazil (Choudaha & Chang, 2012) where travel restrictions to the U.S. are not as severe. Not inconsequentially, many U.S. universities depend on the tuition and fees generated by globally mobile students and are expected to continue opening their doors to these learners. Given these ever-increasing numbers and the continued sentiment that international students are ill-equipped for the academic language demands of studying in English-medium universities, further investigation must be done to know how to help these students acquire and become skilled in using academic language.

To that end, I put forth this research synthesis, which investigates what researchers have discovered about academic vocabulary interventions for post-secondary English learners and attempts to both describe the state of this research field as well as to integrate the findings of individual studies in order to propose generalizations (Cooper & Hedges, 1994). By gathering and synthesizing this information, the practitioner and educational research communities can use it as a jumping off point for future research as well as a baseline for what works in the classroom so that, collectively, we can make progress in this area.

1.1. Why Is Vocabulary So Important?

Common sense tells us that knowing words is essential when understanding and communicating in any language. Participating in conversations, crafting emails, and reading news articles are all impossible without understanding the words in a language.

Researchers have taken that common-sense idea – the necessity of knowing words when using a language – and attempted to quantify it; they have found that the interrelationship between literacy skills is a little more complicated than common sense would dictate. Various hypotheses attempt to explain the relationship between vocabulary and other literacy skills such as word recognition (e.g., Nation & Snowling, 1998), metacognitive awareness (e.g., Kirby et al., 2012), and reading comprehension (e.g., Anderson & Freebody, 1981). Stahl and Nagy (2006) assert that vocabulary is best understood as a compilation of complementary hypotheses, each revealing a different piece of the literacy puzzle (for a review, see Stahl & Nagy, 2006).

While there are certainly aspects of vocabulary's role in various literacy skills that we still do not fully understand, researchers have continued to affirm the importance of vocabulary, and quality vocabulary instruction, particularly for success in academic settings.

1.1.1. What Do We Know About Effective Vocabulary Instruction?

In their landmark meta-analysis, Stahl and Fairbanks (1986) noted three specific aspects of effective vocabulary instruction: using contextual and definitional information to know a word, utilizing activities that encourage deeper processing, and having multiple, meaningful exposures to words. These principles of effective vocabulary instruction have endured (Graves, 2006; Marulis & Neuman, 2010).

In addition, students learn words more deeply when those words are encountered in authentic contexts, with rich instruction, where students are provided with many opportunities to use words and see them fit together meaningfully with other words (Beck et al., 2002; Graves, 2006; McKeown, Beck, Omanson, & Pople, 1985). An added layer of word knowledge relates to morphological awareness, where students who are able to deconstruct words into their subsequent parts can then decipher the meanings of individual morphemes and, eventually, much larger, more complex terms; this has been shown to contribute to multiple aspects of literacy development (Carlisle, 2010; Kuo & Anderson, 2006).

We also know that students can and need to acquire vocabulary through both intentional and incidental means. Learners increase their incidental vocabulary through their reading, which can supplement words they learn through direct instruction (Blachowicz & Fisher, 2004). Where intentional learning indicates "a deliberate attempt to commit new information to memory" (Hulstijn, 2003, p. 360), often in the form of memorizing word lists with definitions, incidental vocabulary learning occurs when learners are "focused on something other than word learning itself" (Paribakht & Wesche, 1999, p. 196). Instruction that encourages both aspects of word learning appears to be beneficial for any learner's vocabulary acquisition.

What is true regarding the importance of vocabulary and aspects of quality vocabulary instruction in general also apply to English learners with a few added considerations.

1.1.2. Why Is Vocabulary Important For English Learners?

Vocabulary is undeniably foundational in language learning (e.g., August, Carlo, Dressler, & Snow, 2005; Meara, 1980), regardless of the learners' first language (e.g., Lam, Chen, Geva, Luo, & Li, 2012), age (e.g., Proctor, Silverman, Harring, & Montecillo, 2012), or proficiency level (e.g., Golkar & Yamini, 2007).

Researchers continue to confirm the importance of vocabulary for English learners' literacy skills. Specifically, studies affirm the correlation between a strong vocabulary and reading comprehension for ELs (e.g., Laufer, 1992; Nation, 2006; Schmitt, Jiang, & Grabe, 2011). Schmitt (2008) concluded that learners need to know about 8,000 – 9,000 word families in order to read in English. Additionally, in a study with English-speaking adults, Guo et al. (2011) found that various literacy skills, such as morphological and syntactic awareness, contribute independently to vocabulary. Second-language vocabulary acquisition is complex and continues to be a high-interest research area; however, we have established certain things about effective practices for ELs' vocabulary instruction.

1.1.2.1. Receptive and productive vocabulary

Delving more specifically into one aspect of vocabulary, most researchers accept the distinction between receptive and productive language skills; these constructs appear often in the literature and are familiar to the research community (e.g., Meara & Alcoy, 2010; Sylvén & Ohlander, 2014; Townsend & Collins, 2009). Receptive skills (listening and reading) are those used primarily for taking in information; productive skills (speaking and writing) are those used for producing or generating meaningful language.

Researchers do not fully agree on the meanings of these terms, however. Some have interchanged the terms "receptive" and "passive" as well as "productive" and "active" (e.g., Laufer & Goldstein, 2004). Using the terms "passive" and "active," however, seems to bring more confusion than clarity since utilizing one's receptive vocabulary can be done actively or passively. Additionally, there is some debate as to whether the receptive-productive taxonomy is dichotomous or part of a continuum (Pinot-Shahov, 2012); for pedagogical purposes, the most practical distinction is to maintain the use of "receptive" to mean identifying terms and "productive" to mean generating language through speech or writing.

1.1.2.2. Vocabulary breadth and depth

Another often-used distinction in vocabulary knowledge is between breadth and depth of knowledge (e.g., Kieffer & Lesaux, 2012; Nagy & Herman, 1987; Read, 2004). Vocabulary breadth pertains to vocabulary size or number of known vocabulary words. Vocabulary depth, on the other hand, goes beyond a surface-level knowledge of word meaning; depth can include knowledge of a word's connotation, figurative representation, collocation, morpho-syntactic function, etc. (Read, 2004).

Much research has been conducted regarding the relationships of vocabulary breadth and depth to each other as well as to reading comprehension showing positive correlations between and among these three (e.g., Moinzadeh & Moslehpour, 2012; Qian, 2002; Rashidi & Khosravi, 2010). Researchers seem to agree that vocabulary breadth and depth correlate positively with one another (Moinzadeh & Moslehpour, 2012; Qian, 2002), though Wang (2014) notes that development of vocabulary depth tends to lag behind vocabulary breadth development. Additionally, Rashidi and Khosravi (2010) found that students who performed well on

assessments of vocabulary breadth and depth also scored higher on reading comprehension assessments.

There is some disagreement over whether vocabulary breadth or vocabulary depth plays a more important role in reading comprehension. While Moinzadeh and Moslehpour (2012) found that vocabulary breadth is a stronger predictor of reading comprehension than vocabulary depth, Rashidi and Khosravi (2010) found that vocabulary depth was a more significant predictor of reading comprehension.

Assessing vocabulary depth and breadth of 168 Chinese university students alongside their linguistic competence, which included reading and writing, Wang (2014) found that while both vocabulary breadth and depth can effectively represent participants' linguistic competence, depth is a better indicator of that competence.

Clarifying these terms and highlighting the importance of each aspect is important when discussing quality vocabulary instruction for English learners. While many vocabulary measures focus on vocabulary breadth and receptive vocabulary, effective instruction tends to incorporate multiple dimensions of word learning.

1.1.3. What Do We Know About Effective Vocabulary Instruction For ELs?

Researchers have confirmed that vocabulary practices that are effective for first language (L1) learning also work well with linguistically diverse populations (e.g., Genesee, Lindholm-Leary, Saunder, & Christian, 2005; Proctor et al., 2011). Read (2004) confirmed the benefits of direct vocabulary instruction for English learners. Gámez and Lesaux (2012) showed that increased exposure to complex vocabulary was positively correlated with students' development of vocabulary skills. In terms of metacognition, researchers have shown the perceived and actual

benefits of learners using multiple, varied vocabulary strategies to aid comprehension at the word-level and in larger contexts (e.g., Fan, 2003).

English learners, however, have some additional considerations (Short & Echevarria, 2005). First, multiple studies have shown the benefits of incorporating a learner's first language in vocabulary instruction (e.g., Carlo et al., 2008). Teachers can utilize the language students have already acquired and pair it with the appropriate English terms. While this may not be possible in all situations as equivalent translations are not always available, it can certainly be a benefit where those translations do exist.

The use of an English learners' first language can be particularly helpful when working with language cognates (Proctor & Mo, 2009). For example, the abstract English term "symbolic" is translated as "simbolico" in Spanish, which sounds nearly identical. Obviously, learners whose L1s are typologically different from English (e.g., Mandarin Chinese) would not have access to this strategy; however, recognizing and accessing cognates to learn vocabulary is a metacognitive strategy available to some English learners.

Even those for whom cognates are not available may still be able to use their existing L1 linguistic and metalinguistic skills in learning new languages. Cummins (1979) proposed the idea of linguistic interdependence, arguing that a learner's first language is interwoven with second and subsequent languages such that accessing one draws from the other, which Cummins called the "common underlying proficiency." To use a familiar analogy, language is like a massive iceberg; on the surface, it shows only two peaks that appear distinct and separate, however, beneath the surface, the two peaks are, in fact, connected to form one submerged mass that cannot be separated. Cummins postulates that the same is true with language. We must be mindful, however, in too broadly applying this analogy to language; it may work in

understanding metacognitive or metalinguistic awareness but may not translate as specifically to vocabulary. Nevertheless, drawing upon all of a learner's available language stores may prove beneficial in further language learning.

English learners also need linguistic scaffolding such as concrete representations of items about which they are learning (Harper & de Jong, 2004; Shin, 2006). These can include visual representations, realia artifacts, and video images and can provide far more meaning than a simple list of words with definitions, especially for more abstract concepts. For example, when learning about an abstract idea such as symbolism, it may benefit English learners to have a specific example of symbolism such as a nation's flag representing loyalty to that nation or a budding flower garden representing Spring and new life. Images, videos, or actual objects of these specific examples alongside first language terms and definitions may help English learners truly comprehend the idea of symbolism beyond merely drilling that term and definition.

Researchers differ as to whether it is better for learners to acquire vocabulary through explicit instruction or incidental exposure. Krashen (1989) argues that intensive, isolated vocabulary learning is limited because it relies on areas of the brain not specialized for language learning; he advocates for incidental word learning through reading. Webb and Chang (2012) studied incidental L2 vocabulary learning through assisted and unassisted reading for 82 secondary students learning English in Taiwan. They found that both assisted reading and unassisted reading led to significant gains in vocabulary knowledge with those experiencing assisted repeated reading significantly outscoring their counterparts. Recent research into game-based learning indicates the potential of gaming platforms to enhance incidental language acquisition as this may be an engaging way to help ELs develop both vocabulary and communicative competence (Peterson, 2010).

On the other hand, Calderón, Slavin, and Sánchez (2011), in their review of studies pertaining to effective instruction of ELs, assert the need for explicit vocabulary instruction before, during, and after reading across content areas, particularly for grade-school children. These conflicting viewpoints are each strongly held and widely discussed within the research community, though particular attention to the academic vocabulary acquisition of post-secondary learners is lacking.

Multiple reviews and research syntheses have discussed the effects of vocabulary interventions for English learners (Cisco & Padrón, 2012; Nagy & Townsend, 2012; Read, 2004; Taboada, 2009). While these reviews and syntheses have added to our knowledge base of vocabulary instruction for English learners, they have focused on general vocabulary acquisition (Read, 2004), the specific relationship between vocabulary and reading comprehension (Taboada, 2009), and K-12 classrooms and learners (Cisco & Padrón, 2012; Nagy & Townsend, 2012).

There remains a need to review and synthesize what we know about post-secondary English learners' academic vocabulary learning.

1.2. What Do We Know About Academic Vocabulary?

Before discussing how English Learners (ELs) grow in their academic language acquisition and vocabulary usage, we must first understand these terms. As academic vocabulary is one aspect of academic language, I begin by first discussing and defining academic language.

1.2.1. What Is Academic Language?

At the heart of academic language is academic vocabulary. Baumann and Graves (2010) note the difficulty in defining "academic vocabulary" as many scholars do not agree upon a consistent definition and use various terms. I have adapted a combination of Baumann and

Graves's (2010) basic definition of academic vocabulary and a more specific one by Scott, Nagy, and Flinspach (2008) for use in the present study: academic vocabulary is the lexis encountered in academic settings having distinctive syntactic, morphological, and stylistic features.

Multiple researchers have attempted to categorize academic vocabulary into more specific subcategories: domain-specific, general, high-frequency, symbolic representations, etc. (Baumann & Graves, 2010; Fisher & Frey, 2008; Harmon, Wood, & Hedrick, 2008). The distinction between general and domain-specific academic vocabulary is important for ELs as it distinguishes words they may encounter across their classes and academic domains (i.e., general academic vocabulary) from words that are specific to certain disciplines (e.g., science or history). Townsend et al. (2012) found that general academic word knowledge contributed unique variance (between 19% and 34%) on assessments of academic achievement across disciplines for diverse middle school students. It is easy to understand why the distinction between general and domain-specific academic vocabulary is important for K-12 classrooms, yet this distinction is also highly pertinent for post-secondary learners who must still, often simultaneously, navigate broader academic settings and specific academic disciplines.

1.2.1.1. Academic word lists

To aid teachers and learners in choosing which academic words to focus on, researchers have created various word lists. There are several general academic vocabulary lists. Among the more frequently-cited lists are Coxhead's (2000) Academic Word List (AWL), Gardner and Davies's (2013) Academic Vocabulary List (AVL), and Xue and Nation's (1984) University Word List (UWL).

Creators and proponents of these lists claim that having such a resource makes vocabulary learning manageable, arguing the value of focusing on words students will encounter

across contexts. Though these lists have been criticized (Bhatia, 2002; Hyland & Tse, 2007, 2009), others applaud their practicality (Eldridge, 2008). It is important to clarify that no single list will magically unlock academic language for ELs. Academic word lists cannot be a one-size-fits-all prescriptive remedy that, if directly instructed, will cure all English academic vocabulary woes. Rather, these lists are tools providing learners with a base of academic words from which they can infer and learn the meanings of other words and can access other aspects of academic language.

1.2.1.1.1. Coxhead's AWL

One of the most commonly-referenced lists of academic vocabulary is Coxhead's (2000) Academic Word List (AWL). This list has been used across disciplines and learner backgrounds to gauge academic vocabulary knowledge, cited in thousands of studies (e.g., Schmitt, 2008) and resources about teaching and learning vocabulary (e.g., Nation, 2013). Created largely as a resource for EAP (English for Academic Purposes) courses, Coxhead devised this list using a corpus of academic texts containing approximately 3.5 million words from across 28 subject areas. Eliminating the 2,000 most common English words, Coxhead focused on the 570 word families that account for about 10% of the total words in academic writings. These word families were then divided into ten sub-lists, ordered by frequency (i.e., Sub-list 1 contains more frequently used words than Sub-list 10); each sub-list contains 60 words with the exception of Sub-list 10, which contains 30 words.

Now that we have defined some key terms related to academic language and vocabulary, it is important to establish what the research already shows regarding instruction in this area.

1.2.2. What Do We Know About Academic Vocabulary Instruction?

More and more recent studies have focused on academic language, for English learners as

well as for English-only users. Some have examined teacher practices (Keisler & Bowers, 2012; Zwiers, 2007), others have investigated related socio-cultural and linguistic factors (Scarcella, 2003), and many have focused on how K-12 learners acquire academic language and/or academic vocabulary (e.g., McKeown, Crosson, Artz, Sandora, & Beck, 2013; Taboada & Rutherford, 2011; Townsend, Filippini, Collins, & Biancarosa, 2012).

Much of what is true for effective vocabulary instruction also appears to be true for academic vocabulary instruction. For example, a primary directive for both, it seems, is the importance of explicit instruction (Keisler & Bowers, 2012). In addition, as with general vocabulary, academic vocabulary appears to be best learned in text-based, authentic contexts (Lesaux, Kieffer, Kelley, & Harris, 2014), with multiple exposures (Ozturk, 2015), and incorporating metacognitive strategies, such as morphological awareness (Kieffer & Lesaux, 2012).

However, as discussed above, academic vocabulary has unique features and, therefore, also has unique teaching and learning considerations. As Lesaux et al. (2014) argue, an effective academic vocabulary intervention should focus on general academic words that students will encounter across disciplines but may not hear every day. In addition, learners need multiple opportunities to incorporate all four language domains – speaking, listening, reading, and writing – as they encounter and learn academic vocabulary since these words appear less frequently in everyday conversations.

Researchers have more recently explored the effects of morphological awareness (MA) and other metalinguistic skills on vocabulary learning (Bowers, Kirby, & Deacon, 2010; Kieffer & Box, 2013; Kieffer & Lesaux, 2012; Kuo & Anderson, 2006; McBride-Chang et al., 2008;

Ramirez et al., 2010). As Nagy and Townsend (2012) note, academic words are morphologically complex; thus, having an awareness of morphological structures as well as strategies to decode these morphemes in context prove beneficial for students learning academic words on their own (Jiang, Kuo, & Sonnenburg-Winkler, 2015; Kuo & Anderson, 2008). These promising findings about the role of MA in vocabulary knowledge and reading comprehension have major implications for academic settings.

The unique aspects of academic vocabulary instruction are further complicated when considering the unique aspects of English learners acquiring academic English.

1.2.3. What Do We Know About Academic Vocabulary Specific To English Learners?

Nagy and Townsend (2012), reviewing interventions of academic vocabulary for diverse K-12 classrooms, argue that principles of effective vocabulary instruction for ELs also apply to academic vocabulary instruction for these learners. These include multiple exposures to target words, practice identifying and producing these words, and using words in authentic, content-based discussions.

As Nagy and Townsend (2012) point out, research has yet to show a "direct causational relationship between vocab gain and gains in academic comprehension or academic writing" (p. 103). Nagy and Townsend (2012) advocate the use of such assessment tools as the Vocabulary Levels Test and online vocabulary profile programs that can evaluate the use of general academic words in student writing. In their review of academic vocabulary interventions for linguistically diverse K-12 learners, Nagy and Townsend (2012) conclude that when words are used as tools for thinking and communicating about content in a discipline, these help "students learn and learn to use academic vocab" (p. 105), but there is still a lot we do not know. For

example, research is limited as to the role that L1 (first language) plays in acquiring the two different types of academic vocabulary (i.e., general academic vs. discipline-specific).

Nagy and Townsend (2012) note additional issues meriting further study; specifically, they note that we currently do not have a developmental trajectory for acquiring academic vocabulary. Such a knowledge base would allow us to tap into learner characteristics at each stage so that we can identify what learners do and do not know and help them reach the next level of their development.

1.2.3.1. What are important variables for ELs' academic vocabulary development?

One of the things we still do not know about English learners' academic vocabulary is how it develops. While several studies focus on the academic vocabulary of various age groups, proficiency levels, or types of vocabulary learned, there is little research showing the progression of academic vocabulary learning over time.

In an attempt to establish a progression, the following sections are divided according to learners' grade levels in each study, which correspond to certain age ranges. This is not intended to imply that all learners in a given grade have the same level of English proficiency; in fact, the opposite is true. Learners at different grades have varying levels of language proficiency. Rather, this section is meant to show several key variables of what learners at different ages/grades are capable of and what they may need in terms of their academic vocabulary. This is also intended to provide a rough academic language learning progression from kindergarten through matriculation. These findings are summarized in Table 1.

Table 1. Academic language development by grade level

Grade Level	Key Variables
Kindergarten	• Capable of learning and using scaffolded academic language • Need explicit instruction in rich context, multiple exposures, and authentic opportunities to practice
Elementary	• Benefit from extended instruction (pre-taught vocabulary, highlighted words within text, student-created glossary entries, etc.) in both general and domain-specific academic vocabulary • Benefit from embedded definitions of domain-specific academic vocabulary
Middle School	• High capacity for learning academic vocabulary • Benefit from structured instruction including: multiple word exposures, reading words in context, connecting prior knowledge, using morphological analysis, and including target words in their own writing • Benefits of explicit academic vocabulary instruction may differ based on proficiency level with intermediate proficiency students reaping the greatest benefits
High School	• Benefit from integrating content learning with language development • Need instruction in academic vocabulary learning strategies, not just words themselves • Even students with reading difficulties can benefit from metacognitive strategies, such as morphological analysis
Postsecondary	• Excel when learner involvement load is higher • Capable of transferring learning across courses and content areas as well as from general to specific knowledge

1.2.3.1.1. Kindergarten

Young students are capable of learning and using academic language when vocabulary is thoughtfully taught in context and students are allowed multiple opportunities for authentic practice. Spycher (2009) compared intentional and implicit academic vocabulary learning for two intact classes of linguistically-diverse kindergarten students (n=39) from low socioeconomic backgrounds. The experimental group was provided with explicit instruction on 20 academic vocabulary terms in addition to the regular science curriculum; target words were taken from text in the normal curriculum (e.g., examine, larva, metamorphosis). Vocabulary instruction included

choral repetition, student-friendly definitions, multiple examples, student-generated sentences, contextualized discussions, and short-answer questions. The control group experienced the regular curriculum from the same teacher without explicit academic vocabulary instruction. Spycher found that students in the experimental group learned more academic words and were more effective in communicating their understanding of scientific concepts, expressing themselves with greater ease and confidence. Furthermore, there were no significant differences in gain score when analyzed by students' language background. Spycher concludes by noting the importance of intentional vocabulary instruction. This study indicates not only that young children are capable of learning and producing academic vocabulary but also that this learning is similar for students from varying language backgrounds.

Several other recent articles explore using academic language with kindergarteners in their science lessons, indicating an ability for this age group to acquire academic language (Gotwals & Wright, 2017; Parks & Oslik, 2018; Parsons & Bryant, 2016). Parsons and Bryant (2016) attempted to deepen students' academic vocabulary by incorporating various interrelated science terms in their lesson cycles. During an eight-week intervention, teachers utilized informational text read-alouds, explicit instruction of related science vocabulary, supported discussions, and authentic activities in centers where students could practice using scientific terms. Qualitative and quantitative results indicate an ability for these young students to comprehend and produce academic vocabulary. Though this study was not specific to English learners, the results could easily be extrapolated to young or beginning level English learners. Practices such as utilizing a variety of materials and strategies to reinforce academic terms, such as informational text read-alouds, structured discussions, and authentic opportunities to practice terms, could benefit learners of all ages. This study confirms findings by Spycher (2009) and

further illustrates that even the youngest students are capable of acquiring and producing academic language.

1.2.3.1.2. Elementary

For elementary-aged EL students, using extended isolated and embedded instruction can significantly increase these students' academic vocabulary knowledge. Additionally, older students show an increased level of academic language skills, which are vital for such academic tasks as reading comprehension.

August, Artzi, and Barr (2016) compared extended and embedded general academic vocabulary with 509 third and fourth grade students; embedded vocabulary included a direct definition of academic vocabulary within the text (e.g., " how magnets interact or work with each other") while extended vocabulary was pre-taught with pictures and sentences, posted on word walls, highlighted within the text, assessed, and used by students in creating glossary entries. This study found that while extended instruction yielded significantly greater gains (g=0.71), students scored significantly higher as a result of both conditions (extended: g=1.7; embedded: g=0.57). Additionally, when comparing whether general academic words were easier to acquire than science terms, the authors found similar gains on both types of academic vocabulary knowledge through extended instruction; however, students scored higher when learning embedded science vocabulary than when learning embedded general academic vocabulary. These findings indicate great potential gains for elementary students when using extended and embedded instruction for both general academic and discipline-specific vocabulary.

Uccelli and Galloway (2017), in a study of 218 upper elementary students, found considerable individual differences across grades as well as within grades. Overall, sixth-grade students displayed significantly greater academic language skills than the fourth- and fifth-grade

students. On a measure of core academic language skills, sixth graders showed an average correct score of .62 (SD=.26) while fourth graders obtained an average score of .52 (SD=.28). Additionally, regression analyses showed that individual differences in core academic language skills significantly contributed to differences in reading comprehension scores. These findings illustrate: a) the variability of academic language skills across students and grade levels at this critical pre-adolescent juncture, b) the potential for older students to develop greater academic language skills, and c) the importance of obtaining these skills, particularly in relation to reading comprehension.

1.2.3.1.3. Middle school

Middle-grade students show a particular capacity for learning academic vocabulary, which appears to have a significant impact on students' academic achievement. The degree of impact, however, may vary based on students' levels of English proficiency.

Lesaux, Kieffer, Kelley, and Harris (2014) conducted a 20-week intervention with over 1,469 language minority sixth-grade students from 14 large, urban middle schools. The intervention included structured exposure to target academic words using research-based methods such as multiple word exposures, reading words in context, connecting prior knowledge, using morphological analysis, and having students use target words in their writing. Results indicate a significant and substantial intervention effect for these language minority students ($d=0.49$, $p<.0001$), indicating the potential that academic vocabulary instruction can have on students' achievement.

These results confirm findings by Townsend, Filippini, Collins, and Biancarosa (2012). This study of 339 diverse seventh- and eighth-grade students showed that general academic word knowledge explained a significant amount of variance - between 19% and 34% - in students'

academic achievement across four disciplines (math, social science, science, and reading comprehension) as measured by standardized, standards-based assessments. Similar to Lesaux et al. (2014), this indicates the importance of general academic vocabulary knowledge at the middle school level as well as the potential value of targeted academic vocabulary instruction, particularly in general academic words.

Work by Hwang, Lawrence, Mo, and Snow (2015) with over 6,000 students from 13 middle schools indicates that students may benefit differently from an intervention geared toward their proficiency levels. This study compared English-only (EO) students with three groups of language minority students: IFEP students (initially fluent English proficiency; those who come to school with English proficiency such that they do not require language support), LEP students (limited English proficient; those still developing English proficiency and requiring language learning support), and RFEP students (redesignated fluent English proficient; those who have essentially "graduated" from needing language learning supports in the classroom). Results showed that RFEP students experienced significant benefits from the intervention while IFEP students experienced moderate benefits and LEP students experienced slightly less benefit than the IFEP group. In addition, the number of years since a student had been moved to the RFEP designation also impacted their academic vocabulary scores; those who had been redesignated within the past two years showed higher gains than those who had been redesignated for three or more years. Thus, there seems to be a "sweet spot" for students whose English proficiency is developing such that within a few years of being considered as English proficient, these students are able to reap the greatest benefits from academic vocabulary and reading interventions.

1.2.3.1.4. High school

English learners at the high school level begin to encounter more linguistically-demanding academic material and, therefore, require more advanced learning techniques to acquire academic language. These students may benefit from instruction that integrates content with language learning, particularly where instruction incorporates culturally-relevant material. High schoolers may also benefit from a greater emphasis on strategy instruction where they are empowered to analyze and learn academic vocabulary on their own.

One difficulty for high school English learners, whether of beginning or advanced proficiency, is that many school programs separate language development from content learning. Rather than separating content and language instruction, Short, Fidelman, and Louguit (2012) advocate the power of training content area teachers using protocols of sheltered instruction, where comprehensible input and scaffolding techniques, such as think-alouds to aid student understanding, are built in to content-area lessons. In their large-scale study comparing the scores of middle and high school students whose teachers (in various content areas) were trained in a sheltered instruction protocol (SIOP) with students whose teachers did not receive this training, results indicate that sheltered instruction has a positive impact on students' language skills. Students in the SIOP district scored significantly higher than students in the comparison group in measures of writing and oral language, though reading comprehension scores were not significantly different. These findings indicate that sheltered instruction in content areas may benefit students in their overall language development as well as in their content area knowledge.

Stewart and Walker (2017) argue for leveraging English learners' experiences, perspectives, and international travel in the secondary classroom to help these students learn language alongside content. Using a culturally-relevant historical fiction novel with supporting

texts about World War II, four refugee high school students from Myanmar developed discipline-specific higher level thinking skills while acquiring and practicing linguistic skills in all four language domains (listening, speaking, reading, and writing). These authors emphasize the need for late-arriving English learners to integrate content and language learning as opposed to separating language learning from discipline-specific content; they argue that English learners' life experiences uniquely situate them to excel in simultaneous language and content learning.

Helman (2015) saw improvement in scientific vocabulary knowledge by using a specific morphological analysis strategy with three high school English learners who had reading disabilities. Students increased and maintained the ability to analyze science terms using the Clue Word Strategy (CWS). These findings point to the potential benefits of teaching vocabulary learning strategies, particularly within specific disciplines such as science, at the high school level.

This coincides with Snow's (2010) urging to empower older students to become independent learners of science. Additionally, Slama (2012) noted the need for academic language interventions among US-born English learners as longitudinal data through standardized testing indicates that even after nine years in US classrooms, many are not sufficiently prepared to succeed in academic environments.

Interestingly, longitudinal data from the National Education Longitudinal Study of 1988 (NELS:88) shows that only one out of every eight of the ELLs in this study earned a bachelor's degree compared to one in four of their native English-speaking peers (Kanno & Cromley, 2013). While a myriad of non-linguistic factors contributed to these numbers, this is certainly a

noteworthy distinction between English-only students and English learners concerning the post-secondary academic readiness of linguistically-diverse learners.

1.2.3.1.5. Post-secondary

At the post-secondary level, English learners seem to thrive when presented with tasks that require greater learner involvement. Additionally, they are able to transfer their learning across disciplines.

In his review of 41 studies, James (2014) found that learning transfer from EAP courses can result in transfer to students' learning in other courses. James noted that this learning transfer can take place with both specific and general information as well as across varying distances (when situations are very similar or fairly different). While not specific to academic vocabulary knowledge, this review indicates that learning transfer at the collegiate level can and does take place for English learners.

Similarly, Kim (2008) found that tasks requiring a higher level of learner involvement produced more effective initial learning as well as greater retention. In addition, Kim found that different tasks having the same involvement load yield similar initial learning and retention. This study focused on the L2 vocabulary acquisition of 64 ESL learners at the post-secondary level. Again, while these findings are not specific to academic vocabulary learning, they indicate that a potential key in vocabulary learning at this level may be a learner's involvement.

The research here points to the complexity surrounding English learners' academic vocabulary development. At each age and proficiency level, various factors must be taken into account. This is certainly true for post-secondary learners, a group that will be further discussed in the following section.

1.2.3.2. What do we know about academic vocabulary for post-secondary ELs?

As we have seen, academic vocabulary can be challenging for most students at the K-12 level, but what about post-secondary English learners? Self-reports from post-secondary learners reveal that these advanced language users also struggle with understanding and using academic vocabulary (e.g., Zhou, 2009). This includes those advanced enough to pass the TOEFL and/or GRE. Researchers, professors, and even students themselves have recognized the need to further examine international students' academic vocabulary (e.g., Evans & Green, 2007; Martin, 1976; Wang & Bakken, 2004).

Vocabulary is one of the main barriers to post-secondary students' academic success (Kuehn, 1996; Santos, 2004). Flowerdew (1999), using in-depth interviews with advanced English learners, ascertained that many felt they were at a disadvantage, noting a perceived lack of rich vocabulary as one of their key difficulties. Work by Evans and Green (2007) has indicated that tertiary students in Hong Kong perceive their lack of ability in receptive and productive English vocabulary to be the most pressing issue related to their academic goals.

More specifically, in a longitudinal study of 28 university students, Evans and Morrison (2010) found that while "understanding key vocabulary" was self-reported by participants as one of the easier aspects of their English-medium studies, "understanding specialist vocabulary" was one of the most difficult. While difficulty with academic writing was the dominant concern, lack of technical vocabulary stood out as a major area of struggle as reported through surveys and interviews, partly because of the impact this perceived lack had on participants' academic reading. These participants also noted difficulty understanding and adapting to disciplinary discourse in their first year of university study; the authors, therefore, concluded that disciplinary acculturation requires more than one semester or term to occur.

Linguistic precision becomes increasingly important when discussing specific academic topics at an advanced level (Snow & Uccelli, 2009). Knowing the words with which to communicate thoughts clearly and appropriately increases the chances of understanding others and being understood – being successful in communication. It is, therefore, important to have a diverse lexical repertoire in order to communicate within and across academic disciplines.

While proficiency levels of students obviously vary, all students entering a program in which English is the language of instruction must be competent and functional in English to learn in that academic environment. In other words, if students do not know or understand what is being said or what they are assigned to read, they cannot function and learn the content of their courses. If they are overly focused on the language, they will have a much more difficult time thinking about, comprehending, navigating the actual material of a given course.

Yet, as noted earlier, academic English is no one's first language; everyone desiring to participate in academic conversations in English must acquire this language – its vocabulary and discourse features. Part of the question here is: do post-secondary English learners have the tools so that they can do that successfully?

While we can make some connections and assumptions about academic vocabulary interventions for post-secondary English learners based on what we know about vocabulary acquisition in K-12 settings or about general vocabulary acquisition, these contexts are vastly different. Professors in post-secondary academic settings often spend little time on direct vocabulary instruction, particularly for general academic words. They assume students know these or will figure them out (Santos, 2004). Professors also expect an increased level of literacy, assigning more readings than K-12 teachers and requiring more written work.

Our current knowledge level for academic vocabulary at the post-secondary level is sparse but growing. With the current research attention on academic vocabulary, we will continue to grow in our understanding of this field, but we must first pause and reflect on what we have already uncovered so that we can use that knowledge as a foundation and move forward.

1.3. Purpose Of The Current Study

As research into the post-secondary academic vocabulary of English learners continues to grow, now is an opportune time to begin synthesizing findings to establish a knowledge base. Such a base would provide a foundation upon which further research can be conducted and by which educators can discover, view, and implement best practices. To that end, this synthesis culls together, analyzes, describes, and reports on the available intervention research related to post-secondary English learners' academic vocabulary acquisition.

This study, thus, provides a review of currently available research into this area. Post-secondary English learners will likely only increase in the coming years; therefore, it is imperative to find and evaluate ways for these learners to be successful in their acquisition and usage of English academic vocabulary.

2. METHODS

Informed by Moher et al., (2010) and Cisco and Padrón (2012), I determined the initial inclusion criteria: studies must focus on academic vocabulary interventions for post-secondary English learners. I included studies with learners from various L1 backgrounds and countries of origin. Since a major aim of this research synthesis was to investigate post-secondary vocabulary learning, I excluded any studies with K-12 students. I did not set any limitations on publication year.

2.1. Search And Selection

I began by conducting a broad article search using EBSCO and ProQuest online databases to further examine 11 different education and social science databases. I included the following databases: Academic Search Complete, Education Full Text, Education Source, Educational Administration Abstracts, ERIC, MLA International Bibliography, Psychology and Behavioral Sciences Collection, PsycINFO, Social Sciences Full Text, and Teacher Reference Center. I selected the options to "find all my search terms" and "search within the full text of the articles."

I chose to only include published articles from peer-reviewed journals. While this introduces potential publication bias by excluding dissertations and theses, it also helps ensure the inclusion of articles that have undergone peer review.

My search terms included "academic vocabulary" in conjunction with "ESL", "EFL", "ELL," "English learner," "L2," "second language," "bilingual," or "linguistically diverse" anywhere in the text. Initial searches using combinations of "vocabulary" with variations of "English learner" showed upwards of 9,000 matches in the aforementioned databases. Thus, to

further narrow search results to the specific research questions under study here, I focused on "academic vocabulary" and included "advanced" or "adult" with the above terms. This yielded 894 articles.

I initially screened articles by manually reviewing titles and abstracts. Articles focused on K-12 classrooms and non-English learners were excluded from further review. Those articles not initially excluded (n=122) were then more rigorously screened through an initial review of the full publication. This was based on the following inclusion criteria: 1) English learners, 2) academic vocabulary, 3) post-secondary setting, and 4) intervention (See Table 2).

Table 2. Inclusion criteria

Criteria	Description
English learners (ELs)	Learners whose first language (L1) is anything other than English; those learning English as a second or foreign language
academic vocabulary	Studies focused on academic vocabulary, such as those on the Academic Word List (AWL; Coxhead, 2000), or studies targeting specialized, technical vocabulary. Studies examining everyday language or a specific grammar aspect (e.g., prepositions) were excluded.
post-secondary	Studies not taking place in K-12 classrooms; included university and work settings requiring technical vocabulary
intervention	Studies describing some sort of treatment, using a comparison group or reporting pre- and post-test results for a single group

To be thorough in my search, I also manually reviewed the reference lists of the initially included articles. I noted 72 articles that seemed to be vocabulary interventions based on the article titles. I then reviewed the abstracts for each of these using the same screening criteria as before.

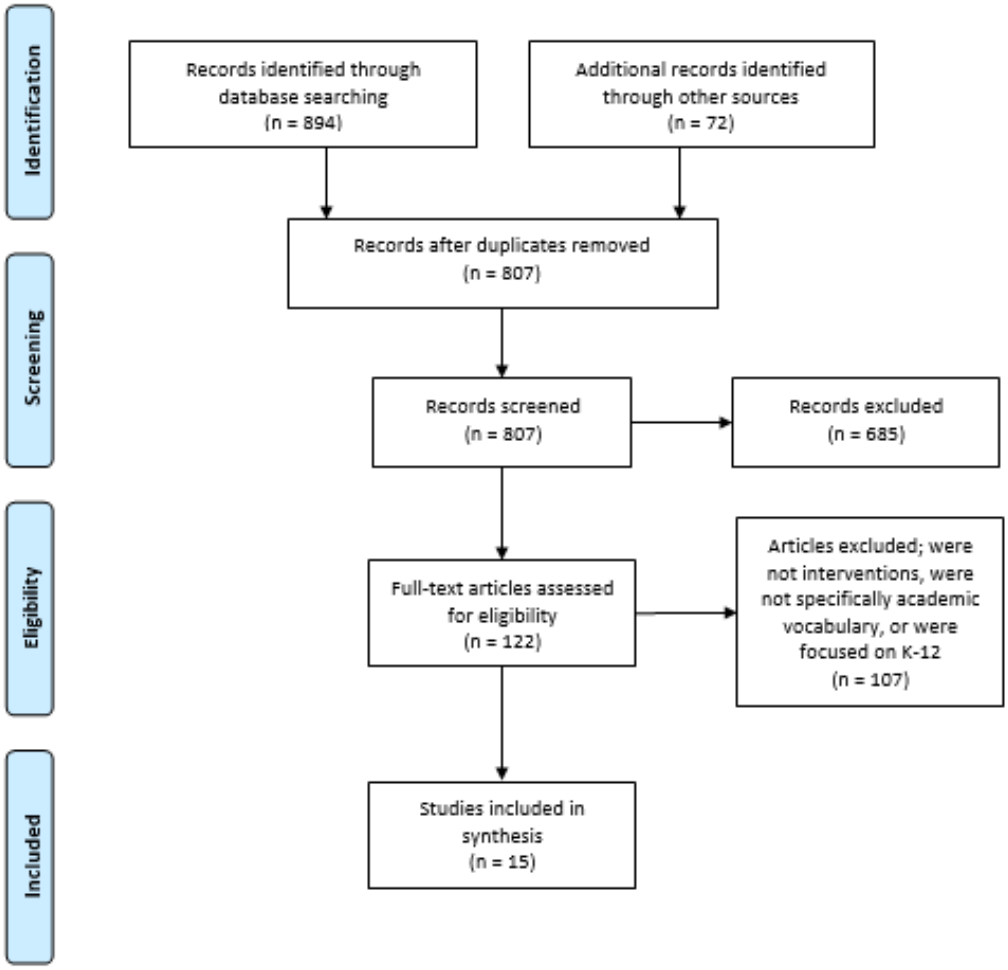


Figure 1. Search and selection

Using a detailed article matrix to code information for each study (Garrard, 2014), I examined the 122 remaining articles. This in-depth examination led to the further exclusion of 107 articles. I then went back through the remaining 15 included articles to verify that each was suitable for final inclusion (Figure 1). Thus, in total, I screened over 966 articles, which resulted in a final included list of 15 articles (Table 3).

I chose to include all 15 of the remaining studies for several reasons. First, the final number of included published articles was relatively small. Second, were I to exclude any of these articles based on study quality or quasi-experimental design, the small number of studies would further dwindle, making it nearly impossible to conduct a research synthesis. Third, following one of the points made by Cronbach et al. (1980), "Much that is written on evaluation recommends some one 'scientifically rigorous' plan. Evaluations should, however, take many forms, and less rigorous approaches have value in many circumstances" (p. 7). Finally, as even Glass, the researcher who credited with coining the term “meta-analysis” admits, “Our only regret in integrating studies on psychotherapy (Smith and Glass, 1977) is that we arbitrarily eliminated ‘pretest-posttest-only’ designs from consideration; we may have thrown away much good data” (Glass, 1978). Thus, while the 15 included studies vary in terms of rigor, quality, and design, each adds a piece to the puzzle of academic vocabulary for post-secondary English learners - and, at this point, there are so few interventions adding pieces to that puzzle that we will use what we have so that we can move forward. In keeping with a true research synthesis (Cooper & Hedges, 1994), however, each individual study will be critically analyzed and evaluated, taking study quality into consideration when integrating and synthesizing findings.

2.2. Data Analysis

Following the initial screening, I analyzed the 15 remaining articles using the article matrix (Garrard, 2014) I constructed (see Appendix). This process allowed me to see differences and similarities between studies that pointed to larger themes. I began by noting the following information for each article: research design, participant characteristics, intervention procedures and duration, outcomes, assessment instruments, intervention effects, overall study quality, and any additional notes.

Consistent with the approach described by Cisco and Padrón (2012), I analyzed the 15 included articles using the constant comparative method (Glaser & Strauss, 1965). This approach combines an analytic use of explicit coding alongside theory generation to extract, formulate, and revise themes throughout the analysis. In other words, the themes were not determined a priori; I determined the themes based on the data.

I chose this type of analysis because it allowed me to approach the data without preconceived codes or themes. Since research syntheses are sparse for this specific group of learners, I was open to whatever themes arose from the data itself and did not begin with specific themes in mind. I was thus able to recognize and articulate themes while still reading and analyzing the included articles. This method also allowed me to review and revise initial themes after reading further studies. Returning to the articles as new themes emerged, I added and adjusted themes to fit the data. For example, I initially did not include a theme regarding technology; however, after reading several articles, I noted that some interventions used no technology while others relied solely on the use of technology or compared a technologically-based intervention with a tradition pen-and-paper version. I then decided to include technology in my analysis and, thus, added a category to my article matrix where I documented each study's technology use. Though this theme was not part of my initial analysis, the constant comparative method allowed me to include the theme of technology as I encountered it in the data.

3. RESULTS

Search and selection procedures revealed 15 studies that fit inclusion criteria (see Table 3): Alijany, Mansoory, and Divsar (2015); Ángel and García (2017); Asmaa, Noorizah, and Zaini (2015); Dizon (2016); Joseph, Watanabe, Shiung, Choi, and Robbins (2009); Kaur and Hegelheimer (2005); Kiliçkaya and Krajka (2010); Lessard-Clouston (2006); Lin and Liou (2009); Moskovsky, Jiang, Libert, and Fagan (2015); Pauwels (2012); Poole (2012); Rezaei and Karbalaei, (2013); Tsai (2011); and Zhang, Song, and Burston (2011). Included study results will first be presented as individual studies then as a collective.

Table 3. Included studies

Study	Participants	Study Design	Intervention	Words Taught: Isolated, Embedded, or Both	Intervention Duration	Receptive, Productive, Both	Use technology
Alijany et al. (2015)	40 Iranian university students	Exp/Ctrl	Reading authentic academic model essays infused with AWL words	Embedded	9 weeks, 17 sessions	Receptive	No
Ángel & García (2017)	16 Columbian university students enrolled in teacher prep program	2 Exp groups	1 semester of Academic Writing Course	Embedded	1 semester, 4 hours/ week	Productive	Yes
Asmaa et al. (2015)	60 university students in Yemen	Exp/Ctrl	Data-driven learning (DDL) activities + Concordance vs. Dictionary & grammatical collocations	Both	2x each week for 2 hours X 4 weeks = 16 hours	Receptive	Yes
Dizon (2016)	9 Japanese university students	Pre-post, single group	Quizlet	Isolated	3x/week for 10 weeks	Receptive	Yes
Joseph et al. (2009).	32 Japanese university students	Exp/Ctrl	2 different softwares – iTango and IKnow	Isolated	4 weeks, Minimum of 6 hours	Both	Yes
Kaur & Hegelheimer (2005)	18 undergraduates at a university in the USA (different L1s)	Exp/Ctrl	Online concordance	Isolated	1x/week for 1 semester	Both	Yes

Table 3 Continued

Study	Participants	Study Design	Intervention	Words Taught: Isolated, Embedded, or Both	Intervention Duration	Receptive, Productive, Both	Use technology
Kiliçkaya & Krajka (2010)	38 students at a university in Turkey	Exp/Ctrl	WordChamp – web reader with dictionary capability (glossing)	Both	5 weeks, 3 hours each week (15 total hours)	Receptive	Yes
Lessard-Clouston (2006)	12 graduate students in Canada	Pre-post, single group	Graduate course – Introduction to theology class	Embedded	1 semester	Both	No
Lin & Liou (2009)	25 Chinese university students	Pre-post, single group	3 main features: 1. Explicit instruction 2. Online quizzes 3. Pair writing and individual lexical logs	Both	8 weeks 2 classes per week (Total: 800 minutes)	Both	Yes
Moskovsky et al. (2015)	120 students at a Chinese Normal University	2 Exp groups	Bottom-up vs. top-down emphasis learning AWL words	Both	48 hours (8 weeks, 6 hrs/week)	Both	No
Pauwels (2012)	59 Dutch students studying to be translators/ interpreters	Exp/Ctrl	5 sets of study materials – each set added different supports and activities	Isolated	5 weeks	Productive	Yes
Poole (2012)	26 freshmen enrolled in a large US university, varying L1s	Exp/Ctrl	3 groups: control, concordance-based, dictionary-based	Both	50 minutes	Both	Yes
Rezaei & Karbalaei (2013)	67 students at English language institutes in Iran	Exp/Ctrl	3 vocabulary learning strategies: 1. Word parts 2. Elaboration techniques 3. Context clues	Both	1 semester	Receptive	No
Tsai (2011)	129 students studying semiconductors in Taiwan	Exp/Ctrl	Multimedia learning software using narrated videos in L1 and L2	Both	7 weeks	Productive	Yes
Zhang et al. (2011)	62 Chinese students	Exp/Ctrl	Vocabulary delivered via text message (SMS) vs. paper-based	Isolated	26 days	Receptive	Yes

3.1. Included Studies: Individual Findings

This section provides a summary of each included study, offering details on participants, interventions, outcomes, and study designs. Included studies are presented in alphabetical order based on the first author's last name.

3.1.1. Alijany, Mansoory, and Divsar (2015)

Alijany et al. (2015) measured the receptive academic vocabulary of 40 Iranian EFL students at an English language institute using two different types of academic essays. Experimental group participants (n=20) were exposed to target academic vocabulary through reading 15 authentic academic model essays; a portion of each essay (8.5-11%) incorporated words from the AWL. The remaining 20 participants (control group) read 15 IELTS (International English Language Testing System) model essays that had been produced for an academic module. All essays were provided as supplementary reading to accompany the course textbook. To ensure the measurement of incidental vocabulary learning, participants did not receive instruction related to the target words. Participants were assessed on a researcher-created instrument of 40 multiple-choice items, which incorporated academic terms from the AWL model essays.

Experimental group participants demonstrated significant gains on an assessment of those vocabulary terms as compared to their pre-test scores ($t(39)=-8.39$, $p=.001$). They also scored significantly higher than a control group, who did not have access to the model essays, on both a post-test ($t(39)=-6.34$, $p=.001$) and a one month delayed post-test ($t= -6.43$, $p=.001$).

3.1.2. Ángel and García (2017)

Ángel & García (2017) did not explicitly focus on academic vocabulary learning in their semester-long academic writing course, but they did assess it as a component of academic

writing. Students in both experimental cohorts of this study, whose first language was Spanish, exhibited high scores for academic vocabulary produced in their writing (based on a rubric that included other components such as organization and syntax). Cohort 2 showed slightly higher vocabulary scores than Cohort 1 based on the raw data (4.8125 vs. 4.75), though the difference was not statistically significant. While the scoring difference is negligible, especially considering the small sample size, a possible reason for the slightly higher raw scores could be that Cohort 2 experienced extended tutoring sessions via the virtual writing lab as well as more systematized feedback throughout the course. Perhaps a replication study with a larger sample size could confirm these findings.

While the difference between the groups' productive academic vocabulary is insignificant, there are several possible reasons for such high overall scores of academic vocabulary. First, the authors report that participants' English proficiency fell between A2 and B1 levels (A1 is the highest level). Perhaps the advanced English proficiency of these learners caused a ceiling effect when using this specific measure. Similarly, the instrument itself (a writing rubric) may not have been sophisticated enough to capture participants' nuanced academic language use. Using a specific academic vocabulary measure in addition to this rubric may have yielded more robust results. Lastly, inter-rater reliability was not reported in this study; perhaps a more rigorous scoring system with multiple raters would have yielded clearer or different results.

3.1.3. Asmaa, Noorizah, and Zaini (2015)

Asmaa et al. (2015) measured academic vocabulary learning for two intact groups (comprised of 30 female students each) who were taking the same university reading course. Both groups used isolated vocabulary learning tools, such as a dictionary and concordance lines;

however, only the experimental group was taught strategies and provided with opportunities to observe embedded target words in context. In addition to two sessions where the experimental group was trained to notice contextualized words, this group was also provided with printed activities focusing on word meanings and grammatical collocations through what the authors describe as data-driven learning. Target academic terms were pulled from course readings based on their frequency according to the AVL (Academic Vocabulary List; Gardner & Davis, 2014).

Both groups improved from pre-test to post-test; however, the experimental group significantly outperformed the control group on both an initial post-test ($t(60)=3.155$, $p=.004$) and a delayed post-test ($t(60)=2.97$, $p=.006$). Interestingly, both groups scored even higher on the delayed post-test with the experimental group still significantly outscored the control group. This is likely due to the fact that the delayed post-test was administered just after the course mid-term where students had recently reviewed the target vocabulary.

While the data-driven learning activities were not clearly described as part of the intervention, it is evident that experimental group participants were offered more activities and support in their academic word learning (printed worksheets and activities). This is likely the reason for their significantly higher scores since there were no significant differences between groups prior to the course based on pre-test results.

3.1.4. Dizon (2016)

Dizon (2016) examined the impact that a vocabulary learning website (Quizlet) had on the academic vocabulary knowledge and motivation of nine Japanese university students in their second year. Students could access the online Quizlet platform, which primarily uses flashcards for study purposes. The flashcards in this study incorporated target AWL terms on one side with English definitions alongside Japanese (L1) definitions on the flip side. Participants could choose

from various study methods on the website or mobile application; these included simple identification, generating the appropriate term given the definition, or playing a word-matching game with the terms and definitions. In addition to being encouraged to review on their own, participants were introduced to target AWL words during class and provided with class time to practice these words; they could use their phones or computers to practice Quizlet during class. Dizon found that scores improved significantly from pre- to post-test on Nation's (2001) Vocabulary Levels Test ($t(9)=-2.64$, $p=0.03$), a multiple-choice measure of receptive vocabulary; these findings are inconclusive, however, as this study lacks a control group.

Dizon also used a questionnaire to examine participants' self-reported use of the Quizlet program. These results indicated that about 20% studied for less than 20 minutes outside of class each week; about 45% reported studying for 20-40 minutes, and about 33% said they studied for 40-60 minutes. No respondents claimed to study for longer than an hour outside of class each week. Additionally, 65% of participants noted their preference for using Quizlet on their smartphones compared to 35% who preferred using their computers.

3.1.5. Joseph, Watanabe, Shiung, Choi, and Robbins (2009)

Joseph et al. (2009) compared two different vocabulary-based software programs with 32 Japanese students; the iTango program presented target words in 10 lists of 25 with no structured learning process while the iKnow program used quizzes, audio support for pronunciation, sequenced re-presentation, and individualized adapted pace to support and encourage active recall. Both groups were given the same words and could study at any time. The iKnow group showed a large effect size over the iTango group of 1.67 on the recall portion (productive) and 1.29 on the multiple-choice questions (receptive). Effect sizes on the delayed post-test were also large: 1.37 for recall and 1.35 for multiple-choice questions. Interestingly, regarding receptive

vocabulary, the iKnow group improved from pre- to post-test (+10%) while the iTango group scores decreased by 4% on the post-test. In terms of productive vocabulary, Joseph et al. (2009) found that both the iKnow and iTango groups performed significantly better on the post-test where they had to recall terms when provided with synonyms (iKnow improved by 31%; iTango improved by 4%). In a delayed post-test, both groups scored lower than in the initial post-test, though the iKnow group's score remained higher than that of the iTango group. The iKnow group maintained gains in both assessment areas; the pre-test to delayed post-test gains were +7% for receptive vocabulary and +15% for productive vocabulary. The iTango group maintained a gain of +3% for productive vocabulary but showed a total loss of -9% in receptive vocabulary from pre- to delayed post-test.

While these authors admirably acknowledge various limitations present in this study, one threat to validity that merits further discussion is the high attrition rate; four students were eliminated from the iTango group because they did not meet the minimum weekly study requirement (1.5 hours each week). This potentially speaks to both the motivation of study participants or to participants' interest and engagement in the present study method. In other words, it is possible that participants in the iTango group were less motivated to study and acquire target words. Conversely, it is possible that the iTango platform did not capture participants' interest or engage them in the material such that they wanted to study words using this method. In addition to the noted effectiveness of the iKnow software, it is thus possible that iKnow outperforms iTango in terms of participant engagement and interest. Perhaps one reason for iKnow's effectiveness is its high-interest level for participants.

3.1.6. Kaur and Hegelheimer (2005)

Kaur and Hegelheimer (2005) measured both receptive and productive vocabulary for 18 ESL students at a university in the midwestern United States. Participants, who were randomly assigned, used either an online dictionary (control group) or an online dictionary and concordance software (experimental group) to learn AWL words. The experimental group outperformed the control group in a sentence-building task (productive vocabulary) and multiple-choice cloze task (receptive vocabulary), though these differences were not statistically significant. However, in a writing task where participants were encouraged to use academic words, the experimental group attempted to use more academic words than the control group and used significantly more words correctly.

Interestingly, the authors note that experimental group participants self-reportedly used the dictionary tools more than the concordance tool in their writing tasks. Use of both tools varied widely for both groups and, according to the authors, was not correlated with the overall score.

3.1.7. Kiliçkaya and Krajka (2010)

Kiliçkaya and Krajka (2010) investigated the role of technology in the receptive vocabulary learning of 38 EFL students in Turkey as they used 10 reading passages to learn target academic words. The experimental group (using a program called WordChamp) scored significantly higher than the control group (using traditional flashcards, vocabulary notebooks, and paper-based dictionaries) on both the post-test ($t(37)= -3.114$, $p=0.004$) and delayed post-test ($t(37)= -3.672$, $p=0.001$), though both groups improved significantly from the pre-test.

One concern related to this study is the role of reading comprehension in the outcome measures. The researcher-created instrument was developed based on 10 academic readings from

previous proficiency exams; participants were asked to read these then answer five multiple-choice questions. It is unclear whether the ten readings were new to participants or if these had been previously reviewed as part of the course study materials. No sample items from the assessment were provided by the authors. This assessment raises questions as to the potentially confounding role of reading comprehension, which is never explicitly acknowledged or addressed within this study.

3.1.8. Lessard-Clouston (2006)

Lessard-Clouston (2006), using a single-group pre-post design, specifically investigated the vocabulary learning of five Chinese English learners (ELs) (four Cantonese-speaking, one Mandarin-speaking). These students learned and studied target words through an introductory graduate course; they were not provided a separate list of target words but were assessed on embedded academic words they were exposed to in the course readings and lectures. Participants scored higher on both post-test measures; they answered 9.21% more accurately on a word identification task (from 77.96% to 87.17%) and 6.4% more accurately when presented with vocabulary knowledge scales (from 57.20% to 63.60%).

In this study, Lessard-Clouston also investigated the discipline-specific academic vocabulary acquisition of seven native English speakers (NES), comparing these participants to EL participants. The NES group outscored their EL counterparts on both pre-test measures (based on the raw data) as well as on post-test scores on vocabulary depth (knowledge scales). Interestingly, however, the two groups scored similarly on the word identification post-test measuring vocabulary breadth. This highlights the potentially comparable growth that ELs may experience compared to English-only peers concerning academic vocabulary taught in the context of a discipline-specific course.

3.1.9. Lin and Liou (2009)

Lin and Liou (2009) used a single-group pre-post design to evaluate the effectiveness of an eight-week academic vocabulary module for a group of 25 Mandarin-speaking EFL students in their third year at university. Participants experienced explicit instruction via wordlists, lecture notes, concordance usage, and weekly readings with highlighted words. Additionally, they accessed online quizzes, did pair writing, and kept individual lexical logs.

Using Nation's Vocabulary Levels Test (VLT), participants scored 57.44 out of 60 (SD 4.3787) on the pre-test and 58.08 out of 60 (SD 4.4527) on the post-test ($t(24) = -1.154$, $p>0.05$), an insignificant improvement potentially due to a ceiling effect. On the Vocabulary Knowledge Scales (VKS), participants scored 69.48 out of 90 (SD 10.3082) on the pre-test and 81.6 out of 90 (SD 7.3256) on the post-test, showing significant improvement ($t(24) = -9.302$, $p<0.05$). Additionally, they significantly increased in the ratio of academic words produced in academic writing (from 2.60% to 4.99%).

3.1.10. Moskovsky, Jiang, Libert, and Fagan (2015)

Moskovsky et al. (2015) examined the effect of differently-structured learning processes with 120 Chinese ELs taking an academic vocabulary course in their first year at university. In this study, both groups were exposed to target words embedded within a larger context as well as words dissected and discussed at the morphemic level; one group started at the morphemic analysis level and worked through a bottom-up process while the other group started with target words presented in the context of larger passage and worked from the top down. In other words, while both groups went through the same individual steps of learning and practicing the same target AWL words, the two groups went through those steps in opposite order from one another. These authors created an instrument they call the Academic Vocabulary Size Test, a seeming

hybrid of the Vocabulary Levels Test and the Vocabulary Knowledge Scales used to measure receptive academic vocabulary.

Moskovsky and colleagues found that the two groups (bottom-up vs. top-down) both exhibited significant gains in their post-test scores, though the bottom-up process yielded significantly higher gains on receptive academic vocabulary than the top-down method. Results show that starting at the morpheme-level and working up to a larger context may have a greater effect on receptive vocabulary, though the two methods appear equally beneficial for producing academic vocabulary.

3.1.11. Pauwels (2012)

In Pauwels (2012), six mixed-ability groups were given different types of study materials to use over the course of five weeks; these ranged from a list of alphabetized words to a list organized by topic with example sentences (concordance), definitions, and exercises to help practice those target words. In a post-test that asked participants to translate three passages, each including 10-15 target words, the group whose study materials included a list organized by topic along with L1 (Dutch) translations outperformed all other groups except the group that was given concordance-like example sentences (there was no significant difference between these two groups). Interestingly, the concordance group scored significantly higher than the group that was given concordance materials as well as definitions.

It is noteworthy that these participants were studying to be translators/interpreters. Thus, the use of L1 translations may have been a familiar study method, possibly helping to account for why the group given L1 translations scored significantly higher than almost all other groups. It is also noteworthy that the L1 here was Dutch, which is typologically similar to English and

may have aided in language transfer/acquisition. Participants with a typologically different L1, such as Chinese, may have seen different results.

While Pauwels conducted an ANOVA using pre-test information which established that there were no significant differences between groups prior to the intervention, pre- to post-test gains are not provided. Pauwels noted the average score of all six groups to be higher on the post-test than on the pre-test; however, these tests appear to be in different formats and, therefore, cannot easily be compared to show learning gains.

Additionally, initial differences between groups disappeared on a delayed post-test. This may have been due, in part, to test sensitization; the delayed post-test was a shortened version of the initial post-test asking participants to translate one passage. It is unclear, however, whether the delayed post-test passage was identical to the one used in the initial post-test.

3.1.12. Poole (2012)

Comparing dictionary glosses with concordance glosses, Poole (2012) set up a three-group study with 26 learners from various linguistic backgrounds in their first year of college; each group had the same target words contained in the same two reading passages. In the dictionary group, participants were instructed to click on the target words (which were highlighted) within the passage to see their dictionary meanings. The concordance group did the same but were exposed to five sentences containing the target word, each showing different syntactic and semantic aspects of that word. In the control group, students encountered the highlighted target words but had no links to a dictionary or a concordance. Both experimental groups scored higher than the control group on all measures, though not all differences were significant. The concordance-based group scored significantly higher than the control group on the receptive Vocabulary Levels Test (VLT), though there was not a significant difference

between the two experimental groups on this measure. There were no statistically significant differences between the three groups on a judgment task (rating correctness of a target word in a sentence) or a cloze task (fill in the sentence by producing the correct target word).

While no clear explanation is provided for why the intervention was effective on some measures but not on others, the researchers noted that the concordance-group performed better than expected on the VLT; having only example sentences without definition input, the authors assumed the concordance group would score lower than the dictionary group on this measure. They attribute the concordance group's high scores to the fact that this group responded positively to the intervention (as noted in an additional attitude survey) and were, thus, engaged with the material.

Perhaps an alternative explanation for these results is that reading words in context (with or without a dictionary and/or concordance support) transfers more readily to assessments involving syntactic awareness. In both the judgment task and cloze task, target words are presented in the context of a sentence and, thus, include built-in syntactical information. The VLT, however, presents decontextualized words and definitions.

3.1.13. Rezaei and Karbalaei (2013)

Rezaei and Karbalaei (2013) focused on teaching vocabulary strategies to 67 EFL students in Iran using three different methods: morphemic analysis, context clues, and elaboration techniques. The control group was not instructed in these strategies. Both groups were provided with the same list of 30 academic words, though the experimental group studied the words in sets of 10, using a different strategy with each set.

When asked to fill in the blanks with the correct term out of four options, the experimental group had a significantly higher gain score (M=5.89, SD=2.298) than the control group (M=0.13, SD=1.996) (F=118.989, p=.000).

3.1.14. Tsai (2011)

To ascertain the effect of courseware versus teacher-based instruction on the discipline-specific vocabulary of 129 students at a technical university in Taiwan, Tsai (2011) asked participants to explain the meaning and process or purpose of ten terms. These participants, divided based on three intact groupings according to their programs (2-year weekend program, 4-year night program, 4-year day program), were all taking an optional course entitled "English for technology." All participants used lecture notes and the course textbook; additionally, those using the online courseware were exposed to narrated videos with accompanying text presented karaoke-style.

The post-test showed no significant differences between the groups after this seven-week intervention. From this, the author concludes that courseware-based instruction is comparable to teacher-centered instruction; in other words, students perform roughly equally, so institutions should use whichever instructional method best fits their needs. The author states that all participants significantly improved but does not include pre-test scores to compare pre-post or group gain differences.

One question that goes unaddressed here is the role that student background, interest, and motivation play in these results. It is unclear if or how students were motivated to learn the target words as well as how comfortable the students were using courseware. There is also no report provided regarding whether or how often students accessed and utilized the various course features.

3.1.15. Zhang, Song, and Burston (2011)

To examine the effect of electronic delivery of academic vocabulary versus paper-based methods, Zhang et al. (2011) assessed how well 62 students could correctly identify the meaning of an underlined word within a sentence. The experimental group, who regularly received vocabulary lists via SMS (text message), scored significantly higher than the group who received a paper list ($t(61)=2.45$, $p<.05$). The two groups, however, did not score significantly differently on a delayed post-test. Interestingly, both groups significantly outscored their pre-test averages on both the post-test and delayed post-test. Thus, while the group using technology initially experienced a bump in average score, both groups, using isolated vocabulary study methods, outperformed themselves on both post-tests.

One possible explanation for these findings is that the novelty of this technology allowed for short-term interest and increased studying that then diminished over time. Additionally, if students were not motivated to continue reviewing target words between tests, these words may have been forgotten. This points to the potential need for both deeper learner involvement to establish greater recall over time as well as continued exposure to target words to solidify acquisition.

3.2. Synthesized Findings

The following section presents synthesized findings across the 15 included studies. Topics in this section include: Study features such as type of academic vocabulary and learner context, study quality, receptive and productive vocabulary, learner context, technology, and intervention effectiveness.

3.2.1. Study Features

The following section details synthesized findings regarding two key study features. These include the type of academic vocabulary focused on in the included studies as well as learner context for these studies.

3.2.1.1. General vs. discipline-specific academic vocabulary

I began the search for included studies by being open to both general academic vocabulary as well as discipline-specific vocabulary. However, almost all (13 out of 15) studies that met inclusion criteria dealt with general academic vocabulary; only two studies (Lessard-Clouston, 2006; Tsai, 2011) focused on discipline-specific vocabulary (theology; semiconductor technology, respectively). Other screened studies dealing with discipline-specific vocabulary either: were for K-12 students (e.g., Helman, 2015), were not interventions (e.g., Moon, 2017), or were for beginning proficiency English learners (e.g., Madrigal-Hopes, Villavicencio, Foote, & Green, 2014). Thus, one finding of this research synthesis is that studies for advanced adult English learners related to academic vocabulary are primarily focused on general academic vocabulary.

Going a step further, the majority of the included studies focus on the linguistic aspect of academic English with a couple focusing on the cognitive aspect; no studies investigate the sociocultural/psychological aspect of academic language based on Scarcella's (2003) description. By far, the most common focus is on the academic vocabulary knowledge; how well do participants know the meanings or uses of target words. Two studies (Asmaa et al., 2015; Rezaei & Karbalaei, 2013) accessed the cognitive dimension of academic language by teaching participants strategies for vocabulary learning (e.g., word parts, context clues). In addition to

continuing research related to the linguistic and cognitive aspects, another area of needed research is into the sociocultural/psychological aspects of academic English.

3.2.1.2. Learner context

An observation of the included studies is that the majority were conducted in EFL (English as a Foreign Language) contexts. Eleven of the 15 included studies took place in countries where English is not the dominant language of society. These EFL contexts include Iran (Alijany et al., 2015; Rezaei & Karbalaei, 2013), Yemen (Asmaa et al., 2015), Japan (Dizon, 2016), Turkey (Kiliçkaya & Krajka, 2010), China (Lin & Liou, 2009; Moskovsky et al., 2015; Zhang et al., 2011), Belgium (Pauwels, 2012), Columbia (Ángel & García, 2017), and Taiwan (Tsai, 2011). Of those studies in ESL (English as a Second Language) contexts, three take place in the United States (Joseph et al., 2009; Kaur & Hegelheimer, 2005; Poole, 2012), and one study is set in Canada (Lessard-Clouston, 2006). This prompts a caution regarding overgeneralization of the findings from the present research synthesis as well as a direction for future research, which will be discussed further in the Discussion section below.

3.2.2. Evaluating Study Quality

Evaluating research is a complex task - partly because evaluating research is, by nature, subjective; thus, I drew from established criteria (Chalmers et al., 1981; Cooper, 1984; Cooper & Hedges, 1994; Wortman, 1983, 1994) to help shed light on the quality of the studies included here.

In one approach to evaluating research, Chalmers et al. (1981) propose a methods-description approach, which seeks to objectively identify a study's descriptives, protocol, analyses, and data, providing an in-depth list of variables by which to determine a study's general quality as opposed to a reviewer attempting to establish the bias within each study. While

Chalmers and colleagues use this method specifically to evaluate randomized controlled trials in the medical field (Deeks et al., 2003), the process they outline is easily applied to evaluating research in any discipline.

Another approach, building on the work of Campbell and Stanley (1966) and Cook and Campbell (1979), is provided by Wortman (1983), who describes issues of internal, external, and construct validity as well as statistical conclusions that affect research credibility and generalizability; he advocates categorizing and listing the various threats to validity when evaluating research. Wortman also discusses the various ways in which threats to validity have historically been categorized and prioritized. Wortman's work, which is well-cited (e.g., Lipsey & Wilson, 2001; Norris & Ortega, 2000), builds on the foundation set out by Campbell and colleagues and provides a solid basis for examining study quality as it relates to validity threats.

Cooper (1984) combines both the aforementioned approaches, proposing a synthesis of the threats-to-validity approach (Wortman, 1983) and the methods-description approach (Chalmers et al., 1981). Cooper (1984) argues that while this integrative framework does not completely remove subjectivity from the review process, it is a step in the direction of "explicit objective decision making in an area previously rife with subjective and arbitrary judgments" (p. 76).

I chose to use Cooper's (1984) integrative approach to evaluating study quality. Through this integrative framework, I was able to both identify the relevant objective aspects of the included studies (as with the methods-description framework) as well as note threats to validity that are unclear from that information alone.

Using Cooper's (1984) framework of integrated evaluation and drawing from Wortman's (1983) coding form, I created an evaluation matrix (see Table 4) detailing objective aspects of

each study: design features, participant information, sufficient intervention description, instruments used, analyses conducted, and information about each study's publication source. I also included any threats to validity that were present. This approach provided me with an unbiased structure through which to examine and present the information regarding each study as provided by the original authors.

Table 4. Study quality evaluation matrix

Study	**Control group?**	**n**	**Participant selection**	**Random assign-ment to groups?**	**Is the interven-tion sufficiently described?** (i.e., able to replicate?)	**Were the instruments/ outcome measures standardized? If not, was reliability reported?**	**Threats to validity**	**Analyses conducted**	**Publication source & Journal Impact Factor**
Alijany et al. (2015)	Yes	40	Volunteers	No	No	No • Cronbach's alpha = .89	• Selection bias: Male only • Testing effect: same pre- and post-test	• T-tests, paired and independent	• *Modern Journal of Language Teaching Methods* • No Impact factor • Double-blind peer review
Ángel & García (2017)	No	16	Convenience sample	No	No	No • No reliability reported	• Selection bias: Small sample size; No control group • Instrument bias: One rater (researcher) scoring vocabulary use within writing	• Writing rubric	• *GiST: Education & Learning Research Journal* • *No impact factor* • *Double-blind peer review*
Asmaa et al. (2015)	Yes	60	Convenience sample	Yes	No	No • No reliability reported	• Selection bias: Female only • Testing effect: all tests identical, except delayed post was ordered differently • Instrumental bias: One rater (researcher) determining accuracy of responses	• T-test	• *3L: Southeast Asian Journal Of English Language Studies* • *No impact factor* • *Double-blind review* • *Paid submission*

Table 4 Continued

Study	Control group?	n	Participant selection	Random assign-ment to groups?	Is the interven-tion sufficiently described? (i.e., able to replicate?)	Were the instruments/ outcome measures standardized? If not, was reliability reported?	Threats to validity	Analyses conducted	Publication source & Journal Impact Factor
Dizon (2016)	No	9	Convenience sample	No	Yes	Yes • VLT	• Selection bias: Small sample size; No control group • Testing effect: One measure of vocabulary; study time was self-reported	• Paired t-test	• *Teaching English With Technology* • No impact factor • Double-blind peer review
Joseph et al. (2009)	No 2 Exp. groups	36	Volunteers	No	Yes	No • No reliability reported	• Selection bias • Researcher error: Technical issue during pre-test • Mortality: n=36 at beginning, n=26 at end	• Percentage • Boxplots • ANOVA • Wilcoxon (non-parametric)	• *Research & Practice In Technology Enhanced Learning* • No impact factor • Peer-reviewed • Open access
Kaur & Hegelheimer (2005)	Yes	18	Convenience sample	Yes	Yes	No • No reliability reported • Item analysis conducted for pre-test, but details not reported	• Instrument bias: Two raters (teacher and researcher) scoring vocabulary use within writing, no inter-rater reliability noted	• Item analysis on pretest scores – item difficulty & item discrimination • Used percentage of correct word use to calculate correctly used academic words	• *Computer Assisted Language Learning* • *Impact factor (JCR): 1.722* • *Blind peer review*
Kiliçkaya & Krajka (2010)	Yes	38	Convenience sample	No	No	No • No reliability reported	• Selection bias	• Means & SDs • T-tests to compare the 2 groups	• *Turkish Online Journal of Educational Technology* • *No impact factor* • *Double-blind peer review*

Table 4 Continued

Study	Control group?	n	Participant selection	Random assign-ment to groups?	Is the interven-tion sufficiently described? (i.e., able to replicate?)	Were the instruments/ outcome measures standardized? If not, was reliability reported?	Threats to validity	Analyses conducted	Publication source & Journal Impact Factor
Lessard-Clouston (2006)	No	5	Convenience sample	No	No	No • Piloted, but no reliability reported • Borrows from: Word ID & VKS	• Selection bias: Small sample size; No control group • Testing effect: same pre- and post-test	• Means & SDs • Percentages • Wilcoxon signed rank test (one-tailed)	• *Canadian Modern Language Review* • Impact factor (JCR): 0.256 • Double-blind peer review
Lin & Liou (2009)	No	25	Convenience sample	No	Yes	Yes • VLT • VKS • Timed writing task: Scored using Lexical Frequency Profile and ESL Composition Profile	• Selection bias	• T-test • One way repeated measures ANOVA • Writing quality rubric • Canonical analysis	• *English Teaching & Learning* • No impact factor • Peer-reviewed
Moskovsky et al (2015)	No 2 exp. groups	120	Volunteers	No	Yes	No • Test-retest correlations – significant at .01	• Selection bias	• ANOVAs • T-test post-hoc • ANCOVA	• *TESOL Quarterly* • *Impact factor (JCR): 2.056* • *Double-blind peer review*
Pauwels (2012)	Yes	59	Volunteers	No	No	No • No reliability reported	• Testing effect: Delayed posttest was a shorter version of posttest (same passage to translate) • Participants self-reported use of study materials • Attrition: n=79 at beginning, n=49 at end	• ANOVA • T-tests • Spearman rank correlation	• *Language Learning Journal* • *No impact factor* • *Double-blind peer review*

Table 4 Continued

Study	Control group?	n	Participant selection	Random assign-ment to groups?	Is the interven-tion sufficiently described? (i.e., able to replicate?)	Were the instruments/ outcome measures standardized? If not, was reliability reported?	Threats to validity	Analyses conducted	Publication source & Journal Impact Factor
Poole (2012)	Yes	26	Convenience sample	No	No	Yes & No • VLT Judgment task and cloze sentences; piloted with no significant differences	• Selection bias	• ANOVA • T-tests • Attitude questionn aire •	• *CALICO Journal* • *No impact factor* • *Double-blind peer review*
Rezaei & Karbalaei (2013)	Yes	67	Convenience sample	No	No	No • No reliability reported	• Selection bias: In-tact classes	• ANOVA • Kolmogoro v-Smirnov test for normality of data distribution	• *European Online Journal of Natural and Social Sciences* • No impact factor • Peer-reviewed • Open access
Tsai (2011)	Yes	417	Convenience sample	No	No	No • No reliability reported	• Selection bias • Testing effect: Same pre- and post-test	• Independe nt t-test • Mean scores on post-test	• *Journal of Educational Technology & Society* • *Impact factor (JCR): 1.58 4* • *Blind peer review*
Zhang et al. (2011)	Yes	62	Convenience sample	No	Yes	Yes • TOEFL multiple-choice vocabulary test	• Selection bias: In-tact classes • Testing effect: Same pre- and post-test • History effect: No note re: technolo gy/ mobile phone use before interventi on	• T-tests • Qualitative - low inference indicators through written reports	• *Turkish Online Journal of Educational Technology* • *No impact factor* • *Double-blind peer review*

n = number of participants; Exp = Experimental; JCR = Journal Citation Reports; SDs = Standard deviations; VLT = Vocabulary Levels Test; VKS = Vocabulary Knowledge Scales; TOEFL = Test of English as a Foreign Language

In general, the quality of these studies is mediocre. Based on work with the National Research Council, Towne and Shavelson (2002) put forth the following elements of quality research in education:

1. Pose significant questions that can be investigated empirically.
2. Link research to relevant theory.
3. Use methods that permit direct investigation of the question.
4. Provide a coherent and explicit chain of reasoning.
5. Replicate and generalize across studies.
6. Disclose research to encourage professional scrutiny and critique. (p.20)

While the majority of studies included here accomplish points one, two, and four (as listed above), several use questionable methods and instruments, many do not disclose procedures such that they could be replicated, the majority cannot be generalized, and some do not disclose their research such that adequate scrutiny and critique are possible.

Towne and Shavelson (2002) note that methods of education research have encountered much scrutiny, particularly where randomized trials are concerned. They rightly state that no research method is inherently good or better than others; rather, the method must match the question being addressed. Additionally, "scientific inferences are strengthened if they hold up under scrutiny through testing using multiple methods" (p. 20). This is an area where several included studies in the present synthesis are lacking; they conduct minimal analyses or use a single researcher-created instrument. This and other elements of study quality will be described further below.

3.2.2.1. Study design

One finding of this review is the limited number of experimental studies related to this topic; ten of the included studies used a control group, another two studies compared two experimental groups (Ángel & García, 2017; Moskovsky et al., 2015), and three studies used a pre-post, single-group design (Dizon, 2016; Lessard-Clouston, 2006; Lin & Liou, 2009). Of the 10 studies using a control group, only one (Kaur & Hegelheimer, 2005) randomly assigned participants to experimental and control groups. Several others had selection bias that limits the generalizability of findings; for example, Alijany et al.'s (2015) participants were all male, and Asmaa et al.'s (2015) participants were all female. While no original research will be without flaws, these findings indicate a clear need for further experimental research in this area that minimizes selection bias.

3.2.2.2. Lack of detailed reporting

One key finding of this synthesis is the lack of intervention details reported in individual studies. While some provide a comprehensive description of intervention (e.g., Lin & Liou, 2009), others only vaguely describe their intervention procedures and do not provide enough information to replicate the intervention or the full study (e.g., Alijany et al., 2015). Six of the 15 included studies provide intervention details such that these studies could be replicated. Lack of detail and specificity not only diminish the research community's ability to validate and replicate study findings, but this lack also prevents educators and learners from knowing which aspects of an intervention are most effective for vocabulary learning.

3.2.2.3. Journal impact factor

Eleven of the 15 included studies were published in journals that do not have an impact factor. Garfield (2006, p. 90) provides a definition of impact factor: It "is based on 2 elements:

the numerator, which is the number of citations in the current year to items published in the previous 2 years, and the denominator, which is the number of substantive articles and reviews published in the same 2 years." Impact factors for journals with included studied here range from 0.256 (*Canadian Modern Language Review*) to 2.056 (*TESOL Quarterly*).

3.2.2.4. Instruments

The majority of the included studies used researcher-created measures instead of standardized instruments. Fourteen of the 15 studies used researcher-developed measures; three used standardized measures. Two of the studies using standardized measures did so in conjunction with researcher-created measures (and are, therefore, counted in both numbers above). One study (Lessard-Clouston, 2006) adapted two common assessment formats - word identification and vocabulary knowledge scales - though these altered such that they are considered here as researcher-created instruments. While this is certainly common among researchers, the difficulty is in comparing study results, even when researchers report the reliability of their measures, which is not always the case.

Only three of the 15 included studies used a standardized measure (Dizon, 2016; Lin & Liou, 2009; Zhang et al., 2011); Dizon (2016) used only standardized measures while Lin and Liou (2009) and Moskovsky et al. (2015) employed a combination of researcher-developed and standardized measures. The other 12 included studies used solely researcher-created instruments to measure vocabulary learning and/or usage; of these 12, only four studies reported any kind of reliability related to their researcher-created measure.

Lin and Liou (2009), in addition to using two standardized measures (VLT and VKS), used a researcher-created writing task to measure vocabulary production, though they used two standardized methods to analyze those writing samples. Moskovsky et al. (2015) used modified

versions of the VLT, the VKS, and Nation's Productive Levels Test, but seemed to modify these such that they were not the same as the originals and, therefore, should be considered researcher-created. The issue of standardized and researcher-created instruments will be further addressed in the Discussion section.

3.2.3. Receptive And Productive Vocabulary

Measuring receptive vocabulary knowledge typically requires participants to identify or show that they comprehend correct words, meanings, and usages (e.g., multiple-choice questions) while a measure of productive knowledge requires participants to generate or provide accurate terms, meanings, or translations (e.g., fill in the blank or write a paragraph using target words). Common instruments used with ELs include the TOEFL exam for measuring receptive vocabulary and the Peabody Picture Vocabulary Test (PPVT) for measuring productive vocabulary.

Receptive and productive vocabulary are interrelated skills, and both are necessary in an academic context, particularly at the post-secondary level. Students are often required to utilize receptive vocabulary skills by listening to class lectures and reading academic texts. Students are then asked to take that receptive vocabulary knowledge and transfer it to productive knowledge in to do things like participate in discussions and write academic essays.

Interestingly, at the post-secondary level, many measures of English language ability assess receptive vocabulary (e.g., TOEFL, GRE), but most academic content assessments occur through productive vocabulary (e.g., essays, seminar discussions, etc.).

The majority of the included studies here focus on measuring receptive knowledge or a combination of receptive and productive vocabulary knowledge. Six studies exclusively measured receptive vocabulary (Alijany et al., 2015; Asmaa et al., 2015; Dizon, 2016; Kiliçkaya

& Krajka, 2010; Rezaei & Karbalaei, 2013; Zhang et al., 200), three studies exclusively measured productive vocabulary (Ángel & García, 2017; Pauwels, 2012; Tsai, 2011), and six studies measured both (Joseph et al., 2009; Kaur & Hegelheimer, 2005; Lessard-Clouston, 2006; Lin & Liou, 2009; Moskovsky et al., 2015; Poole, 2012).

3.2.3.1. Receptive: Reading words in context

Four studies (Alijany et al., 2015; Asmaa et al., 2015; Kiliçkaya & Krajka, 2010; Rezaei & Karbalaei, 2013) utilized reading passages for receptive vocabulary learning, and participants in all four studies experienced significant gains. Alijany et al.'s (2015) participants, who read academic texts containing target vocabulary, scored higher on the post-test ($t=-8.39$, $p=.001$) while also significantly outperforming the control group on both a post-test ($t=-6.34$, $p=.001$) and a one month delayed post-test ($t=-6.43$, $p=.001$). Asmaa et al. (2015), examining EFL students taking an academic reading course, found that both the experimental and control groups scored higher on the post-test than on the pre-test with the experimental group (who received training in noticing collocations) scoring significantly higher on the post-test than the control group ($t=3.155$, $p=.004$). Kiliçkaya and Krajka's (2010) participants, using 10 reading passages to learn target academic words, improved significantly from pre- to post-test, while the experimental group (using WordChamp) significantly outperformed the control group on both the post-test ($t(37)= -3.114$, $p=0.004$) and delayed post-test ($t(37)= -3.672$, $p=0.001$). Rezaei and Karbalaei (2013), who taught an experimental group three vocabulary learning strategies, found that this group scored significantly higher than the control group ($F=118.989$, $p=.000$) as they used word parts, elaboration techniques, and context clues to determine word meanings within their academic reading.

Participants in all four studies appear to benefit from academic reading; these include both the experimental and control groups in Asmaa et al.'s (2015) and Kiliçkaya and Krajka's (2010) studies as well as the experimental groups in Alijany et al.'s (2015) and Rezaei and Karbalaei's (2013) studies. This points to the potential of using reading to improve receptive vocabulary knowledge. While Alijany et al. (2015) focus on the role of incidental vocabulary learning, Asmaa et al. (2015) and Kiliçkaya and Krajka (2010) both had participants (experimental and control groups) study target words in isolation as well as in the context of their reading. This may indicate greater benefits for students who combine isolated wordlist studying with embedded target words. In addition, teaching learners strategies for word learning, such as noticing collocations (Asmaa et al., 2015), utilizing available tools (like WordChamp; Kiliçkaya & Krajka, 2010), and using word parts to determine meaning (Rezaei & Karbalaei, 2013) may provide added benefits when measuring receptive academic vocabulary.

3.2.3.2. Receptive: More vs. less structure

Zhang et al. (2011) and Joseph et al. (2009) both compared groups who were given a list of target words with groups whose learning of those same target words was more structured. In Zhang et al.'s (2011) study, the control group was given a paper-based list of all target words to study; the experimental group received daily text messages containing target words (structured delivery). Similarly, in Joseph et al.'s (2009) study, both groups had access to the same lists of target words alongside definitions and example sentences. One group used a software (iTango) that solely presented lists of target words (10 lists of 25 words each) with definitions and example sentences; another group used a different software (iKnow) that provided the same English terms, definitions, and examples while also providing audio support for English pronunciation, sequenced re-presentation of material, and an adapted learning pace for

individuals. For example, while the iTango group saw words and definitions concurrently, the iKnow group saw a cue word, giving participants the opportunity to retrieve information about the related target word before this cue word disappears, revealing the target word and definition.

Both studies indicated minimal or only temporary gains in receptive vocabulary. While Zhang et al.'s (2011) experimental group (using the structured SMS delivery) initially outperformed the paper-based group ($t=2.45$, $p<.05$), differences disappeared at a delayed post-test. Both groups did, however, improve from pre- to post-test. For Joseph et al. (2009), the iKnow group's receptive vocabulary score minimally increased by 10% while the iTango group's score actually decreased by 4%; a delayed post-test indicated lower scores than the initial post-test, though the iKnow group maintained a 7% gain over their pre-test scores.

These findings indicate that structuring the delivery of target words may yield initial receptive academic vocabulary gains, but simply adding this structure does not promote long-term recall. Adding examples and audio support to a more structured-delivery method, however, may provide greater benefits, as with the iKnow group from Joseph et al.'s (2009) study.

3.2.3.3. Productive: Specific tools and materials

Three studies measuring productive academic vocabulary involve specific tools and materials, or specific presentation of material, that appear to aid participants in learning target words, at least in the short-term.

Kaur and Hegelheimer (2005) found that those using an online dictionary alongside concordance software attempted to use more AWL words in their writing and to use more academic words correctly than a group using only an online dictionary. However, a sentence-building task revealed no significant differences between the two groups.

Pauwels (2012) gave six groups different sets of study materials and found that the most effective were a thematically-organized list with L1 glosses or the same list with example sentences; groups with these materials scored significantly higher on the post-test than groups with other materials (e.g., thematically-organized list alone, organized list with definitions, etc.). These differences, however, disappeared on the delayed post-test.

Joseph et al. (2009) found that both the iKnow (structured re-presentation, adapted recall) and iTango (wordlists) groups performed significantly better on the post-test where they had to recall terms when provided with synonyms (iKnow improved by 31%; iTango improved by 4%), though the iKnow group improved to a greater degree. Scores dropped on a delayed post-test but remained higher than those on the pre-test (pre to delayed post: iKnow =+15%, iTango =+3%). Tsai (2011) found that both face-to-face and courseware-based learners experienced significant gains after a seven-week unit when they were asked to explain the meaning of a discipline-specific term along with its process or purpose.

Taken together, these results indicate benefits for using specific tools in acquiring productive academic vocabulary, though these benefits may be short-lived. Findings from Kaur and Hegelheimer (2005) and Pauwels (2012) highlight the value of tools such as an online dictionary, glossing system, or concordance, though, as noted in Pauwels's study, this value may only be short-term; benefits may also be exclusive to those who languages are typologically similar to English (as participants in Pauwels's study spoke Dutch). Joseph et al.'s (2009) study seems to indicate the value of software that can adapt to individual users and provide a structured representation of the material.

3.2.3.4. Receptive vs. productive: Technology

When considering the effect of technology specifically on receptive and productive measures of academic vocabulary, it appears that technology has a greater impact on receptive vocabulary tasks than on productive ones. Studies using technology that exclusively measured productive vocabulary showed little or no significant gains or differences (Ángel & García, 2017; Tsai, 2011). However, two studies using technology that exclusively measured receptive vocabulary showed significant differences and gains (Kiliçkaya & Krajka, 2010; Zhang et al., 2011).

3.2.4. Technology

With the abundant digital resources available today, many researchers are turning to these to aid in acquiring academic vocabulary. Of the 15 included studies, 11 use some form of technology. Some only include a small aspect of technology, like an electronic log (e.g., Pauwels, 2012), while others center their research questions on the effectiveness of using technology to learn academic vocabulary (e.g., Kiliçkaya & Krajka, 2010).

Those applying technology to their research interventions related to academic vocabulary use it to varying degrees. Some use tools that are substitutes for hard copy resources such as online dictionaries and concordances (Kaur & Hegelheimer, 2005). Others utilize websites and cellular phone applications, such as Quizlet and iKnow, as study tools (Dizon, 2016; Joseph et al., 2009). A few use online databases of academic writings to help students see academic words in context (Poole, 2012). One compared academic vocabulary learning when students study using traditional pen and paper methods to when students received words via text message on their cell phones (Zhang et al., 2011).

Six of the 15 included studies (Dizon, 2016; Joseph et al., 2009; Kiliçkaya & Krajka, 2010; Lin & Liou, 2009; Tsai, 2011; Zhang et al., 2011) use technology as the crux of the experiment (e.g., paper-based vs. technology-based study methods). Three studies (Ángel & García, 2017; Kaur & Hegelheimer, 2005; Poole, 2012) used technology in a way that was integral to the experiment (e.g., online concordance). Two studies (Asmaa et al., 2015; Pauwels, 2012) used technology minimally (e.g., online log).

3.2.4.1. SAMR model

Seeing such a variety of uses for technology, I employed the SAMR model (see Table 5) to categorize how each study used technology (Puentedura, 2012). The SAMR model is a taxonomy-based way of classifying technology integration, primarily in educational settings; SAMR stands for Substitution, Augmentation, Modification, and Redefinition. I chose this model as it is already commonly referenced by educators in multiple, varied classroom settings (e.g., Peachey, 2018; Portnoy, 2018). Noting the popularity of this model among educators, Hamilton, Rosenberg, and Akcaoglu (2016), caution applying the SAMR model to judge or prescribe classroom activity; they argue for considering context and teacher agency, noting, for example, that using technology as a substitute may be the exact appropriate choice, though it is "lowest level" of the SAMR model. Hamilton et al. (2016) advocate that this model not be used hierarchically and that educators should not be made to feel that they must try to use technology only to "redefine" a lesson or activity.

Others, such as Romrell, Kidder, and Wood (2014) see the benefits of classifying technology use through the SAMR model in an attempt to push the bounds of how technology is currently employed in classrooms. These authors argue that when a technology can modify or redefine an aspect of the classroom, the learning is truly transformed. Thus, while it is used more

by practitioners than by researchers, due to its popularity and potential, I chose the SAMR model as a means of classification and discussion regarding technology in the included studies.

Table 5. SAMR model explanation, adapted from Puentedura (2012)

Category	Explanation	Example
Substitute	Technology is used as a replacement for paper-based methods (interchangeable)	Writing a daily diary entry on a Word document as a substitute for a paper journal
Augment	Technologically-based version goes beyond substitution and provides functional improvement	Digital textbook provides audio-support and linked definitions
Modify	Technology is used to transform and redesign tasks	An online discussion where participants post links to videos and articles; they can also tag and comment on others' threads
Redefine	Technology is used to create tasks that were not possible non-digitally	Using an augmented-reality software that allows students to experience a historic battle in real-time

To better understand the role of technology in academic vocabulary learning, I coded each article for technology use. I then sub-coded those articles based on how, specifically, the researchers used technology, how each usage would be classified per the SAMR model, and what the technologically-related findings were for each study (see Table 6).

Table 6. Features and findings of studies using technology

Study	**Intervention**	**Vocabulary: Receptive, Productive, Both**	**HOW they used technology** (Scale 1 to 3; 1=minimal; 3=crux of experiment)	**Technology use: SAMR** (Substitute, Augment, Modify, Redefine)	**Findings related to technology**
Ángel & García (2017)	Semester-long academic writing course, which included on-going tutoring in virtual writing lab	Productive	2 – Integral to the experiment Through the virtual writing lab, students could access materials and tutorials to aid in a range of academic writing needs, including vocabulary	Augment	Cohort 2 showed slightly higher scores on range of academic lexicon compared to Cohort 1 (4.813 vs. 4.75); while the authors focus on the holistic treatment of the course, one difference between the 2 cohorts was greater emphasis on the virtual writing lab tutoring for Cohort 2

Table 6 Continued

Study	Intervention	Vocabulary: Receptive, Productive, Both	HOW they used technology (Scale 1 to 3; 1=minimal; 3=crux of experiment)	Technology use: SAMR (Substitute, Augment, Modify, Redefine)	Findings related to technology
Asmaa et al. (2015)	Data-driven learning (DDL) activities + Concordance vs. Dictionary only	Receptive	1 – Minimal use, addition to the experiment Some control group members used a dictionary app on their phones	Substitute	Even though the control group did have access to an online resource and the experimental group did not, the experimental group still outperformed the control group on all outcome measures
Dizon (2016)	Quizlet (No control group)	Receptive	3 – Crux of the experiment Used Quizlet on their phones or computers to study words from AWL	Augment	**Pre-Post:** Gain 3.23 (SD 3.67) *t*(8)=-2.64, p=0.03 Participants had positive perceptions of Quizlet for L2 vocabulary learning and spent 20-60 minutes studying outside of class each week
Joseph et al. (2009).	2 different softwares – iTango vs. iKnow	Both	3 – Crux of the experiment iTango – material presented in 10 lists of 25 items each; no structured learning process iKnow – encouraged active recall through quizzes (recall and MCQ), audio support for pronunciation practice, adapted pace based on individual user, sequenced re-presentation of material	Augment vs. Redefine	**Receptive Vocabulary:** • *iKnow* group marginally improved from pre to post in multiple-choice question (MCQ) portion (10%) • *iTango* group scored lower on MCQ post-test than they had on the pre-test (-4%) **Productive Vocabulary:** Both groups improved significantly from pre- to post-test on recall portion iKnow: 31% gain iTango: 4% gain
Kaur & Hegelheimer (2005)	Online concordancer	Both	2 – Integral to the experiment Both groups used online dictionary (dictionary.com) Experimental group used an online concordancer (Tom Cobb's Compleat Lexical Tutor)	Augment	Experimental group outperformed the control group in each vocabulary task and in overall performance based on the raw data (though total score difference was not statistically significant) **Writing Task:** Experimental group had more attempts and significantly more correctly used academic words than control group
Kiliçkaya & Krajka (2010)	WordChamp – web reader with dictionary capability vs. traditional, paper-based vocabulary instruction	Receptive	3 – Crux of the experiment WordChamp – students could look up words as they encountered them in texts	Augment	**Receptive Vocabulary:** Experimental group scored significantly higher than the control group in both the post-test and the delayed posttest

Table 6 Continued

Study	Intervention	Vocabulary: Receptive, Productive, Both	HOW they used technology (Scale 1 to 3; 1=minimal; 3=crux of experiment)	Technology use: SAMR (Substitute, Augment, Modify, Redefine)	Findings related to technology
Lin & Liou (2009)	Web-enhanced lexical instruction 3 main features: • Explicit vocab instruction and concordance practice • Online quizzes • Pair writing and individual vocabulary logs (No control group)	Both	3 – Crux of the experiment • Moodle (course platform) • Online quizzes • TANGO – web concordance • Cambridge online dictionary for AWL AWL online highlighter	Modify	**Receptive Vocabulary:** No significant difference (since pretest scores were already so high for size, possible ceiling effect) **Productive Vocabulary:** Significant difference from pre-test to post-test (t=-9.302, p<.05) Slight score decrease from post-test to delayed post-test, but delayed post-test scores were still significantly higher than pre-test scores **Writing task:** # of words increased in essays and overall writing quality improved
Pauwels (2012)	5 sets of study materials – each set added different supports and activities; all groups used an electronic log	Productive	1 – Minimal use, addition to the experiment Only used an electronic log to specify amount of time spent studying, # of times vocabulary was repeated, and specific activities or subsets	Substitute	Use of technology had no effect on the intervention in this study
Poole (2012)	3 groups: control, concordance-based, dictionary-based	Both	2 – Integral to the experiment Web-based texts – target words read in context Both experimental groups had hyperlinked words, which led to either a dictionary or a concordance	Augment	**Receptive Vocabulary:** *Vocabulary Levels Test (VLT)* Both experimental groups scored higher on post-test than on pre-test; ANOVA showed concordance-based group scored significantly higher than control group – $F(2, 23) = 3.74$, $p=0.04$, but there was no significant difference between the two experimental groups *Judgment Task* No statistically significant differences between groups **Productive Vocabulary:** No statistically significant differences between groups
Tsai (2011)	Multimedia software using narrated videos vs. teacher-centered instruction without courseware	Productive	3 – Crux of the experiment Students with access to courseware practiced using karaoke-style, narrated videos in L1 and L2	Modify	No significant differences in post-test between groups
Zhang et al. (2011)	Structured vocabulary delivery via text message (SMS) vs. paper-based vocabulary list	Receptive	3 – Crux of the experiment 5 vocabulary words delivered each day via SMS 2x/day – lunch at noon and dinner at 5:30pm (vs. printed list of 130 words given at beginning)	Modify	**Post-test** Experimental group scored significantly higher than control group $t=2.45$, $p<.05$ **Delayed post-test:** While both groups scored higher than their original pre-tests, the difference between the two groups disappeared

Based on the SAMR taxonomy, most studies (six) use technology to augment a vocabulary study method or activity. Out of the 11 studies using technology, two studies used technology simply as a substitute for a paper-based method (Asmaa et al., 2015; Pauwels, 2012) while five used it in a way that augments traditional methods (Ángel & Garcia, 2017; Dizon, 2016; Kaur & Hegelheimer, 2005; Kiliçkaya & Krajka, 2010; Poole, 2012), according to the SAMR Model. Three studies used technology to modify a learning task (Lin & Liou, 2009; Tsai, 2011; Zhang et al., 2011), and one study (Joseph et al., 2009) compared an augmenting method with a redefining method.

Interestingly, in the included studies, how integral technology is to the intervention somewhat corresponds to the study's SAMR classification for those studies using technology to substitute or redefine - the two extremes of the SAMR model. In other words, the two studies using technology as a substitute also minimally use technology while the one study using technology to redefine (compared to an augmenting version) also incorporates technology as the crux of the intervention being tested.

Findings show that five out of ten interventions used digital means to augment traditional methods. While this may increase student engagement, Puentedura (2013) argues that using technology to modify and redefine tasks can transform learning. This may be particularly true as more and more digital natives seek to learn and use academic vocabulary.

Interestingly, however, the studies in this review seem to conflict with Puentedura's (2013) assertion that modified and redefined uses of technology yield more substantial, transformative results. Of the three studies here that compare a technologically-based method with its paper-based counterpart, only the study using an augmented method (Kiliçkaya &

Krajka, 2010) showed the experimental group outscoring the paper-based group; the two studies using a modified approach either showed insignificant differences (Tsai, 2011) or differences that disappeared in a delayed post-test (Zhang et al., 2011). This runs counter to the premise of the SAMR model that technology which modifies and redesigns tasks transforms learning where technology that substitutes and augments paper-based methods simply enhances learning.

3.2.4.2. Technology vs. paper-based

Three studies compared a technology-based method with a paper-based method: Kiliçkaya and Krajka (2010), Tsai (2011), and Zhang et al. (2011). Both Kiliçkaya and Krajka (2010) and Zhang et al. (2011) reported that the experimental groups (the ones using technology) significantly outperformed their counterparts (who used paper-based methods) on initial post-tests. However, the differences between groups in Zhang et al.'s (2011) study vanished one month later on the delayed post-test. Tsai (2011) found no significant differences between participants whose coursework was supplemented with an online component. Thus, the use of technology compared with paper-based methods shows mixed results and requires further investigation.

When considering these findings in light of the SAMR model, it is noteworthy that Kiliçkaya and Krajka (2010) used technology in a way that augmented traditional methods while Tsai (2011) and Zhang et al. (2011) both used technology to modify paper-based tasks. This may indicate that technology used to augment traditional academic vocabulary methods may yield greater gains, particularly long-term gains, than technology used to modify paper-based tasks.

In sum, while many researchers are making use of digital resources, they do so to varying degrees and with varying levels of success. These findings will be further reviewed in the Discussion section (below).

3.2.5. Intervention Effectiveness

The interventions represented in these studies vary widely in terms of procedures and outcome focus. Since it is nearly impossible to directly compare studies due to this variance, I have noted several themes across studies that highlight intervention effectiveness. These include depth of involvement with the target material, activities using a concordance, and direct vocabulary strategy instruction.

3.2.5.1. Isolated vs. embedded

One recurring theme in these studies was how academic vocabulary was taught: in isolation or context. For example, Zhang et al. (2011) taught vocabulary in isolation; they provided participants with a list of target words that included part of speech, pronunciation, translation, and an example sentence. Five studies taught vocabulary this way: purposefully, discretely, apart from a larger context (see Table 3). Contrastingly, Alijany et al. (2015) had participants learn words in context by reading model essays containing target words. Three studies examined this type of vocabulary learning: incorporating embedded academic vocabulary. Additionally, seven studies employed both embedded and isolated vocabulary learning methods. Lin and Liou (2009), for example, utilized word lists and explicit instruction over target words in conjunction with weekly readings.

It is noteworthy that some researchers use the terms “incidental” and “purposeful” (e.g., Cisco & Padrón, 2012). I chose “embedded” and “isolated” here because some studies used purposeful strategies for learning words while reading them in context; in other words, their embedded word learning was not incidental, and I did not want to confuse the two sets of terms. The terms “embedded” and “isolated” are also fairly self-explanatory as well as familiar to the research community (e.g., August, Artzi, & Barr, 2016).

In classifying studies as using embedded or isolated methods of vocabulary learning, I considered how to treat those using a concordance-based strategy. While concordances do present each word embedded within the context of a sentence, they highlight and isolate the target word with the focus being on the meaning of that word. I, therefore, categorized concordance-use as an isolated method for learning vocabulary.

3.2.5.1.1. Isolated

Four studies using exclusively isolated methods of vocabulary learning all noted gains. Zhang et al. (2011) showed that both the experimental and control groups (SMS vs. paper-based lists), using isolated study materials, experienced significant gains that were maintained in a delayed post-test of receptive vocabulary. The experimental group initially showed significantly higher scores than the control group, though these differences disappeared five weeks later in a delayed post-test.

Joseph et al. (2009) noted significant gains for the experimental group on a receptive task as well as significant improvements of both the experimental and control groups on an assessment of productive vocabulary; both groups used technology-based materials with isolated target words. Dizon's (2016) single-group pre-post study corroborates these findings, noting significant gains in receptive vocabulary for participants who studied using Quizlet virtual flashcards.

Pauwels's (2012) study further informs this idea by comparing six groups with six different sets of isolated study materials. Pauwels found that the group utilizing L1 glosses significantly outperformed all other groups with varying study materials with the exception of the group using definitions and example sentences, whose scores were not significantly different.

Additionally, Pauwels notes that all six groups experienced benefits from studying words in isolation, though specific pre- to post-test scores were omitted from the report.

These combined findings seem to indicate the clear benefits of isolated study methods as participants using isolated methods in two studies experienced significant gains compared to a control group. Additionally, those utilizing L1 glosses in Pauwels's study significantly outperformed almost all other participants using different combinations of isolated study materials. It is also noteworthy that both experimental and control group participants in Zhang et al. (2011) and Joseph et al.'s (2009) studies saw significant pre- to post-test score gains using isolated study methods.

3.2.5.1.2. Embedded

Three studies teaching academic vocabulary in context highlight the potential benefits of this method. Alijany et al. (2015) found that experimental group participants who read 15 authentic academic model essays containing target words significantly outperformed a control group, who read essays designed for an academic module, on both a post-test and a delayed post-test of receptive vocabulary. While the control group was exposed to target words embedded within essays, the type of essay differed, potentially pointing to the value of using authentic contexts through which to learn target academic vocabulary.

Results from Lessard-Clouston's (2006) pre-post single group design also align with these findings; he found that students who were exposed to key academic terms embedded in the context of an introductory graduate course scored higher on a post-test assessing both receptive and productive vocabulary knowledge (gains of 9.21% and 6.4%, respectively). This speaks to the power of learning academic vocabulary in the context of a discourse community. Ángel and Garcia (2017), likewise, noted a high level of productive academic vocabulary usage for both

cohorts in their study, though there was no pre-test measure to show gains. This points to the possible value of learning academic words by viewing, studying, and producing them within a larger context such as an academic course.

3.2.5.1.3. Combining embedded and isolated methods

Utilizing a combination of embedded and isolated methods for academic vocabulary learning appear to produce significant gains; greater gains are then experienced when words are studied from the bottom-up (progressing from more isolated to more contextualized), when various learning tools (e.g., concordance, dictionary) are employed, and when participants are taught word learning and observing strategies.

Moskovsky et al. (2015) noted significant gains when pairing embedded target words with isolated word learning activities. The bottom-up group (which began with more isolated word learning) significantly outscored the top-down group (which began with more contextualized word learning) on a measure of vocabulary breadth. These findings not only confirm the benefits of combining isolated and embedded study methods, they also highlight the role that study order and process may have on learning as the bottom-up group outperformed the top-down group.

Again, it is noteworthy that both groups in Moskovsky et al.'s (2015) study showed significant gains from pre- to post-test. Similarly, post-test scores from Lin and Liou's (2009) single-group study revealed significant gains in vocabulary depth and the ratio of academic words produced in writing. These findings, coupled with those from Moskovsky et al. (2015), highlight the potential benefits of using isolated methods of academic vocabulary learning in tandem with seeing and observing those words embedded within reading passages.

Kiliçkaya and Krajka (2010) and Poole (2012) noted that using isolated word learning tools alongside web-based tools yields greater gains. Comparing the use of a digital dictionary tool with the use of a paper-based dictionary, vocabulary cards, and vocabulary notebook, Kiliçkaya and Krajka (2010) had two groups read the same passages containing embedded target words. Both groups scored significantly higher on a post-test and delayed post-test of receptive vocabulary with the web-based group significantly outscoring the paper-based group. Similarly, Poole (2012) found that participants who used isolated web-based word learning tools (a concordance or a dictionary) with target words embedded in academic reading passages exhibited significant gains on receptive and productive vocabulary; a control group reading target words in context without these tools did not see significant improvement.

In contrast to Kiliçkaya and Krajka (2010), Tsai (2011) found that not all web-based tools provided added benefits for student learning. In this study, all participants took the same specialized course and used lecture notes and the course textbook; additionally, those using online courseware were exposed to narrated videos with accompanying text presented karaoke-style. While all three groups significantly improved, post-tests showed no significant differences between the groups. Thus, the addition of narrated, karaoke-style videos appeared to have little impact on vocabulary learning.

Both Asmaa et al. (2015) and Rezaei and Karbalaei (2013) taught participants academic word learning strategies while introducing target words, which led to significant gains in receptive vocabulary assessments of those target words. In Asmaa et al.'s (2015) study, both experimental and control groups used isolated vocabulary learning tools, such as a dictionary and concordance lines; however, only the experimental group was taught to observe target words in context. While both groups improved from pre-test to post-test, the experimental group

significantly outperformed the control group on both an initial post-test and a delayed post-test. Similarly, Rezaei and Karbalaei (2013) taught students to use word parts, elaboration techniques, and context clues. The experimental group showed significantly higher gain scores than the control group, who were not taught any word-learning strategies. In both studies, the results indicate that while utilizing tools such as concordances and dictionaries are beneficial, learning strategies to problem-solve the meanings of target words improved scores beyond those who had no strategy instruction.

3.2.5.2. Concordance-based activities

As exhibited above when discussing embedded and isolated methods of vocabulary learning, one popular activity in the included articles is that of using a concordance to learn word meanings as well as collocations. As noted above, this tool appears to benefit learners, which is why I chose to further analyze findings related to concordance-use. Three studies compared concordance use with dictionary use and/or a control group with no enhanced input. Additionally, one study used a concordance-like component. As these results are mixed, I have divided them here based on receptive and productive academic vocabulary tasks.

3.2.5.2.1. Receptive

Two studies indicate unclear results on the question of whether a concordance alongside a dictionary is more effective than just a dictionary on receptive vocabulary tasks. While Asmaa et al. (2015) found that participants who had access to both a concordance and dictionary significantly outperformed dictionary-only users, Kaur and Hegelheimer (2005) found that concordance-plus-dictionary users did not score significantly differently from a dictionary-only group (though both groups did outperform their pre-test scores). These studies seem to indicate

that using a concordance with a dictionary is potentially equivalent to or better than using a dictionary alone on tasks involving receptive vocabulary.

3.2.5.2.2. Productive

In terms of productive vocabulary, three studies provide slightly muddled results that I will attempt to untangle here. Poole (2012) found that both a concordance-based group and a dictionary-only group scored significantly higher than a group with neither resource, though the concordance and dictionary groups did not significantly differently from one another. Kaur and Hegelheimer (2005) found that a concordance-plus-dictionary group produced significantly more academic words in a writing task than dictionary-only users (there was not a concordance-only group). Pauwels (2012), comparing study materials of six groups, found that those using only a concordance scored higher on a productive task than both a group using only a dictionary and a group using a concordance with a dictionary.

Combining these results, when it comes to productive vocabulary tasks, it appears that: a) using a concordance or dictionary is more effective than using no resources (Poole, 2012), b) using a concordance and/or dictionary is more effective than other study materials (Pauwels, 2012), c) using a concordance or a dictionary may yield similar results (Poole, 2012), d) using a concordance with a dictionary helps produce more academic words than only using a dictionary (Kaur & Hegelheimer, 2005), and e) using a concordance alone is more effective than using a concordance with a dictionary (Pauwels, 2012). Thus, while a dictionary is better than nothing, a concordance is the most effective tool for producing academic vocabulary.

Interestingly, Pauwels also found that those using a concordance were only outperformed on a productive task by participants who used L1 (first language) translations to learn and study target words. With such conflicting results, further research is needed in this area to confirm the

results of these studies and to provide clarity regarding the benefits of dictionary and concordance use in learning academic terms.

3.2.5.3. Learner involvement with target material

It may seem obvious that more involvement and interaction with words will yield greater retention and/or knowledge of those words; one included study confirms that assumption while one challenges it.

Findings from Joseph et al. (2009) indicate that students who studied using a multimodal software with audio support, structured re-presentation of vocabulary, and adapted pacing significantly outperformed those who studied via online word lists, particularly on measures of productive vocabulary. The experimental group showed large effect sizes for both recall (1.67) and multiple-choice questions (1.29) on the post-test as well as large effect sizes on the delayed post-test (recall – 1.37, MCQ – 1.35). This confirms the benefits of greater learner involvement with target material via word learning strategies and multimodal studying.

Findings from Pauwels (2012), however, seem to contradict those of Joseph et al. (2009). Pauwels compared six groups, each using different sets of study materials, and found that the study materials requiring the greatest learner involvement did not yield scores as high as those of a group who had fewer study tools (e.g., organized list of target words with L1 glosses). The authors concluded that groups who had "sufficiently informative" material (p. 58) scored higher than those using more elaborate study sets. These findings will be further discussed below.

3.2.5.4. Direct instruction in vocabulary strategies

Two studies (Rezaei & Karbalaei, 2013; Asmaa et al., 2015) highlight the potential benefits of metacognition in academic vocabulary learning as participants were directly instructed in vocabulary learning strategies. Rezaei and Karbalaei (2013) found that a group who

were taught strategies for word learning (elaboration techniques, context clues, word parts) outperformed a group who studied the same target words without these strategies. In Asmaa et al. (2015), while the full intervention procedure is not entirely clear, the authors note that the experimental group received two training sessions where they learned strategies for noticing and using target words in context. Experimental group participants were trained using example words and collocations where control group participants were only provided with study materials. In this study, the experimental group significantly outscored the control group on both a post-test ($t(60)=3.155$, $p=.004$) and a delayed post-test ($t(60)=2.97$, $p=.006$).

4. DISCUSSION

This research synthesis provides a descriptive review of published research on academic vocabulary interventions for post-secondary English learners. This investigation revealed themes related to the research itself (e.g., lack of interventions) as well as to some best practices for academic vocabulary learning. Those themes will be further explored below in light of the existing research base.

4.1. Study Features

Pertaining to the research itself (i.e., studies meeting inclusion criteria), I found that research in this area is lacking, both in quantity and quality. In addition to vaguely reporting procedures and primarily using researcher-created instruments, only ten studies in the past 35 years utilized a true experimental design with a comparison group. The problems associated with these findings will be discussed further below.

4.1.1. General Vs. Discipline-Specific Academic Vocabulary

One finding of this review is that studies for adult English learners related to academic vocabulary are primarily focused on general academic vocabulary and not as much on technical or discipline-specific vocabulary. This fits with Lesaux et al.'s (2014) push to focus on words that post-secondary learners may encounter across academic disciplines. However, Green and Lambert (2018) argue for the place and importance of disciplinary literacy with discipline-specific wordlists. They note that recent research has brought forth multiple discipline-specific wordlists that may allow users a more accessible entry-point into studying academic words associated with specifics fields. The current synthesis, however, indicates an underdeveloped research field related to the actual learning of discipline-specific academic vocabulary. At the

university and graduate levels, discipline-specific vocabulary becomes increasingly important as students must read and research within their chosen fields.

While the value of general academic vocabulary knowledge is not in question, it is certainly wise to consider how discipline-specific academic vocabulary develops, particularly for post-secondary English learners. Inevitably, this line of research would come with a host of further considerations such as how to assess or measure discipline-specific academic vocabulary without confounding learners' acquisition of the content related to that vocabulary. This is admittedly no easy task, yet researchers must confront these questions to better understand academic vocabulary development and acquisition.

This begs the question: would the results of the present research synthesis differ if the data were focused on more discipline-specific vocabulary? For example, August, Artzi, and Barr (2016) note that students experienced greater success with embedded discipline-specific academic vocabulary than with general academic terms; more studies involving discipline-specific terms may reveal different findings than those presented here. Perhaps future researchers will pursue this question.

4.1.2. Learner Context

Just as this synthesis highlights the need for more research related to discipline-specific academic vocabulary, it also indicates a need for further research into the academic vocabulary learning of post-secondary ESL learners. The field has far more offerings regarding academic vocabulary learning for EFL learners at present.

There are certainly distinctions between ESL and EFL contexts that can affect English acquisition (e.g., Azkarai & Oliver, 2019). Learning and practicing English while residing in a

country where one is surrounded by the language has its obvious benefits, particularly when speaking and listening.

It is also true, however, that academic language is not as commonly heard, even in ESL settings. As noted earlier, academic English is its own unique variety of English (Scarcella, 2003) and requires a more specialized setting in which to practice – an academic setting that may be similar to those in countries where English is not the language of society. Thus, while it is important not to overgeneralize due to different contexts and considerations, it is possible to see where findings from the included studies may be viewed more broadly and applied to various academic settings.

Yet, the lack of intervention studies conducted in ESL contexts related to academic English for post-secondary students is a bit surprising. Since more and more English learners are finding their way to universities in ESL settings, the need to help these students acquire and develop their academic English vocabulary is paramount. This is another area for further study.

4.2. Study Quality

Elements of high-quality research, as noted by Towne and Shavelson (2002) and Wortman (1994) include: a strong study design, instruments that allow investigation of stated outcomes, clearly stated findings derived from the data and contextualized in prior research, attempts to limit (or, at a minimum, acknowledge) threats to validity (e.g., sampling bias, instrumentation, attrition, etc.), and a process that includes rigorous peer review.

All studies using human researchers and participants have flaws; however, some flaws are more glaring than others. Unfortunately, a high number of the included studies here contain flaws that potentially call the research results into question - or at least into further consideration.

4.2.1. Study Design: Need For Experimental Research

One finding of this review is the relative lack of intervention studies related to academic vocabulary for post-secondary English learners. As this demographic continues to grow worldwide (Choudaha & Chang, 2012; Curry, 2004), the need for studies related to how this population can most effectively acquire and produce academic vocabulary is acute. While articles describing learning theories, classroom methods, and possible vocabulary strategies certainly broaden our knowledge base about this topic, without empirical evidence from intervention studies, these descriptions simply add to a list of possible actions without providing concrete evidence of what is most effective and efficient for learning (Ioannidis, Fanelli, Dunne, & Goodman, 2015). In order to help post-secondary English learners succeed in their academic and career pursuits, researchers must continue to press the boundaries of what works best – yes, continuing to formulate theories and ideas, but also rigorously, critically testing and evaluating them and publishing those results.

4.2.1.1. Generalizability

One key issue when reviewing research findings is how generalizable the results are when trying to replicate or repeat the study. Several of the studies included here have errors in sampling and/or in study design that limit the generalizability of their findings.

4.2.1.1.1. Lack of control group

There is a striking absence of randomly-selected control groups here. Only seven of the 15 studies included a control group; of those, only two were taken from a random sample. Of those studies utilizing a control group, most were taken from a convenience sample, which limits their generalizability and validity.

As noted previously, there is no such thing as error-free research; any research that involves humans will inherently involve bias and flaws. That said, research that attempts to minimize human bias and threats to validity while also accounting for as many confounding variables as possible will have more reliable, generalizable findings. Studies using a control group help researchers do those things.

4.2.1.1.2. Convenience sample

Eleven of the 15 studies used participants taken from a sample of convenience. The obvious issue with convenience sampling is that it does not allow the researcher to control for, or in some cases to even acknowledge, extraneous or convoluting variables; in other words, the convenience samples may not be truly representative of the population under study. Samples of convenience are inherently biased (Patten, 2012). Some of these biases are minimal while others are problematic, which is why findings based on a sample of convenience must be further scrutinized and, if possible, corroborated with other similar studies and participant samples. Again, the limited number of studies in this area makes that difficult; thus, conducting research with randomized samples instead of convenience samples is encouraged.

4.2.1.2. Detailed reporting

Six of the fifteen studies provided enough details so that the intervention described could be replicated; the remaining nine studies were incomplete in their reporting such that replication would not be possible. There are a few possibilities as to why this might occur.

First, publication word count restrictions may compel some authors to trim down their full experiment reports. Unfortunately, this trimming may come at the expense of key intervention details.

Second, authors may feel that certain details are unimportant or are of little to no consequence for replication and, thus, choose to omit them. This may point to the larger failing in the research community of undervaluing replication research.

Third, authors may choose not to report certain details because the data collection or analysis was muddled or inappropriate. While it is impossible to determine whether details were omitted in an attempt to conceal poor data collection or analyses, many of the included studies omitting key intervention details also conducted minimal analyses (another flaggable issue).

When publishing findings, researchers must attend to details and provide clear explanations of their intervention procedures so that future researchers, practitioners, and learners can attempt to replicate both the intervention conditions and the findings. In order to grow as a community of learners and scholars, it is imperative that we be transparent in reporting research procedures and findings. For example, to report that an intervention is highly effective and that participants showed significant gains is only helpful to the research community insofar as those finding can: a) be replicated and validated by other researchers – in similar and different contexts, and b) be implemented by educators and individual learners to further vocabulary acquisition. One might say that a significant finding is only truly valuable when it can be utilized by others.

While research replication rates are low, they have risen in recent decades (Makel, Plucker, & Hegarty, 2012). Makel, Plucker, and Hegarty (2012) found only a 1.07% replication rate for articles published in the top 100 psychology journals (those with the highest five-year impact factors) between 1900 and 2012. Hunter (2001) notes that far more replication studies are needed to "achieve reasonable accuracy" based on the number of participants in each study. However, Larzelere, Cox, and Swindle (2015) caution that more exact replications are not

needed; rather, we need to increase the number of critical replications, where researchers are intentional about eliminating systematic bias in previous studies. Thus, while this review reveals a lack of detailed reporting and a subsequent challenge to the field to be detailed and transparent in reporting intervention procedures and findings, there is also a challenge to increase replication rates, particularly critical replications.

4.2.1.3. Instruments

Only four included studies used standardized measures (e.g., Nation's Vocabulary Levels Test). Of the 14 studies using researcher-created measures, only five mentioned conducting reliability analyses or piloting their measures to ensure a strong instrument, and only two provided the reliability scores of their measures.

While researcher-created measures have the potential to be valuable tools for data collection, those instruments must be rigorously tested and validated. Without such testing, it is difficult to determine or verify that the results from such a measure are accurate, appropriate, valid, and reliable. Those studies here that do not acknowledge any piloting or reliability analyses conducted on their measures throw their findings into question while also making it difficult for the results of their studies to be synthesized with results from other studies.

It is also noteworthy that the majority of the included studies use measures that only show how well participants learned the target words from the intervention. In other words, participants were only tested on words and strategies that were targeted in the intervention; they were not asked to translate the learning that occurred during the intervention to unknown words. It would be useful to measure how well participants score in understanding vocabulary that was not taught during an intervention. This would allow participants to use the strategies learned during the intervention to problem-solve the meanings of unfamiliar academic words, which is a

more authentic, real-world application. Such a measure could be included alongside one that measures knowledge of target words and/or strategies.

The lack of standardized instruments used in these studies begs the question: what standardized measures are available to assess the academic vocabulary knowledge (receptive or productive) of postsecondary students? As it turns out, the options are limited.

4.2.1.3.1. Standardized instrument options

There are only a few well-cited standardized measures for academic vocabulary available to researchers: Vocabulary Levels Test, Productive Vocabulary Levels Test, Vocabulary Knowledge Scales, and Vocabulary Profiler. These are presented here in the hopes that future researchers may use these in combination with newly-devised measures that may continue to be critically tested and validated in the future.

An often-used measure is Nation's (1990) Vocabulary Levels Test (VLT), which includes a University Word List (UWL) section. This instrument measures vocabulary breadth through a receptive assessment as it provides several sets of words with various possible meanings and asks assessment-takers to match the correct target words with their definitions. One consideration in using the UWL section of this measure is that it is a fairly short assessment. Additionally, the UWL section is presented as the "end" or most difficult section of the VLT assessment, so more advanced English learners may notice a ceiling effect when taking this portion (e.g., Lin & Liou, 2009).

Laufer and Nation (1999) also devised a Productive Vocabulary Levels Test (PVLT) that requires assessment-takers to fill in the remainder of a target word when presented with a sentence that includes the first letter or letters of that target word (e.g., "There are a doz_____ eggs in the basket") (Webb, 2008). Similar to the VLT, it assesses word knowledge across

multiple frequency bands, including the University Word List. The same critique is true, however, for the PVLT that is true for the VLT - advanced-proficiency participants may see a ceiling effect.

Another standardized measure is the Vocabulary Knowledge Scale or VKS (Paribakht & Wesche, 1997). This instrument has two aims: to collect self-reported knowledge of words and to elicit productive use of those words where possible. For each target word, which is decontextualized, assessment-takers choose one of five options ranging from "I do not know this word" to "I can use this word in a sentence;" if participants note that they can provide a synonym, definition, or sentence including the target word, they are asked to do so (Bruton, 2009). One key consideration when using this measure is that it requires hand-scoring in its present form and would be challenging to use with large-scale numbers of participants.

For evaluating productive academic vocabulary displayed in original writing, Cobb's (2002) Vocabulary Profiler easily extracts the academic words used within a text, even detailing the AWL sublist where produced words can be found. This method of evaluation, however, still requires hand-coding of correct usage.

One instrument that seems promising for measuring receptive academic vocabulary is Moskovsky et al.'s (2015) Academic Vocabulary Size Test (AVST). Loosely based on the Academic Word section of the VLT and employing the known/unknown format of the VKS, the authors report a test-retest correlation significant at .01. The authors provide this instrument in Appendix 3 of their article (the online version). Perhaps future researchers could continue to validate this measure through further use.

As noted here, there are limited options for well-established, standardized measures of academic vocabulary knowledge. It is, therefore, no surprise that 14 of the 15 included studies

used some version of a research-created instrument. The problem, however, is that only four of those researcher-measures were accompanied by a report of reliability. To create and proliferate more and stronger academic vocabulary knowledge assessments, researchers must disclose their instruments and any reliability testing conducted. At present, until further assessments are created, validated, and well-established, I recommend that researchers use a combination of standardized and researcher-created instruments while providing reliability reports.

4.2.1.4. Analyses

Of the 15 included studies, only six performed more sophisticated analyses than a simple t-test (e.g., ANOVA). Nine studies conducted minimal analyses (e.g., t-tests, percentage scores, writing rubric, means and standard deviations).

While it is often appropriate to conduct t-tests between two groups or between one group's pre- and post-test scores, this is often an initial test, not the sole analysis conducted. A high-quality example of t-test use is Moskovsky et al.'s (2015) study; an initial pre-post t-test was conducted to show each group's gains. A t-test was also performed to see if there were any differences between the groups at the pre-test; this highlighted a small but significant difference. The authors then conducted an ANCOVA to control for this difference. The more thoughtful levels of analyses showed that the authors were considering differences between groups as well as overall gains.

Unfortunately, most included studies did not conduct such thoughtful analyses. Perhaps a simple t-test was the only analysis warranted, but it seems as though it should be a first step instead of a conclusive final test on which all findings are based.

4.2.1.5. Journal impact factor

Only four of the 15 included studies were published in journals with an impact factor. It is important to acknowledge that there are certainly limitations in placing too high a value on impact factors. Garfield (2015) himself, who was instrumental in establishing the impact factor system, has called it a "mixed blessing" (p. 1). Yet, an impact factor is one indicator of a journal's reputability. With the abundance of published research, a journal impact factor has become a mark of prestige among researchers. Thus, while it may be a flawed system, it is an efficient way to note research that is more cited and viewed as higher quality among researchers. Based on the impact factors of the included studies, the majority of these would not be identified by the research community as highly valued research.

Thus, while a few of the included studies exhibit markers of high-quality research, many conduct minimal analyses, are not generalizable (due to sampling procedures), use only researcher-created instruments without reporting reliability, and have results that are nearly impossible to replicate due to a lack of detailed intervention reporting.

4.2.2. Receptive And Productive Vocabulary

As noted earlier, a common categorization in vocabulary, especially in L2 vocabulary, is receptive and productive language (e.g., Townsend & Collins, 2009). While I also examined vocabulary breadth and depth, these findings were not substantially different from those here regarding receptive and productive academic vocabulary, which will be examined here.

When focusing on receptive vocabulary, the included studies here indicate benefits when pairing the learning mode with the assessment mode. They also indicate that consistency is a more important factor than structure or technological innovation when studying receptive

vocabulary. For productive vocabulary, the addition of tools such as a concordance can greatly enhance the production of academic vocabulary.

4.2.2.1. Receptive: Pairing learning and assessment modes

Three studies (Alijany et al., 2015; Asmaa et al., 2015; Kiliçkaya & Krajka, 2010) utilized reading passages for receptive vocabulary learning, and participants in all three studies experienced significant gains. While the role of embedded and isolated vocabulary learning will be further discussed below, the more important finding here may be that learning through a receptive language domain (such as reading) may benefit students when taking an assessment of receptive vocabulary (where they are asked to identify).

Aligning vocabulary teaching and learning with the ways in which that vocabulary is assessed coincides with research on ecological validity (e.g., Whitehead, 2008). This raises questions regarding purpose and motivation related to learning and assessing; nevertheless, these studies indicate that matching receptive vocabulary learning with receptive vocabulary assessment is likely to yield higher assessment scores.

This fits with Henrikson's (1999) partial-precise dimension of linguistic competence. Henrikson (1999) essentially proposes a spectrum of what it means to "know" a word that allows distinction between vague recognition (i.e., word identification) and precise knowledge of meaning. For the included studies here, incidental reading may yield more of a partial vocabulary knowledge. Learners taking an assessment, such as multiple-choice identification, that then only requires that partial knowledge may see significant gains that would not be captured through an assessment requiring more precise knowledge.

4.2.2.2. Receptive: More than just structure

Joseph et al. (2009) and Zhang et al. (2011), measuring receptive vocabulary for groups using more structured study materials, both indicated minimal gains or initial gains that later disappeared. Since gains in these studies disappeared or diminished significantly in delayed post-tests, it is possible (and even likely) that participants were no longer using the aforementioned technology and structured study environment to review academic words. If this was the case, the bigger takeaway here may be the importance of consistent study for true vocabulary acquisition. In other words, were participants from either group (i.e., using either study method) to continue studying the same words consistently over time, they may have maintained their gains or even increased them.

This would align with Ozturk's (2015) argument for multiple exposures with the added consideration that those multiple exposures occur regularly over an extended time period. Thus, while simply structuring the learning process did not produce permanent significant gains, consistent study over time combined with that structure may yield significant long-term vocabulary retention, another area for further study.

Another case could be made that participants in these two studies experienced shallow or surface learning instead of deep learning. Both Joseph et al. (2009) and Zhang et al. (2011) used isolated study methods centered around technology. Perhaps the novelty of the technology allowed for initial gains that later dissipated, or student motivation for learning these words had disappeared. Regardless, the lack of retention potentially points to surface knowledge. This supports work by cognitive psychologists Craik and Tulving (1975), who noted that depth or elaboration was associated with higher retention.

4.2.2.3. Productive: Specific tools and materials

While findings from Joseph et al. (2009) and Kaur and Hegelheimer (2005) revealed little or no significant gains on receptive vocabulary, both studies showed more promising results for productive vocabulary. Joseph et al. (2009) found that both the iKnow and iTango groups performed significantly better on the productive vocabulary post-test where they had to recall terms when provided with synonyms. Similarly, Kaur and Hegelheimer (2005) note that the experimental group attempted to use more AWL words and used more academic words correctly than the control group.

These findings indicate that participants were better able to transfer their learning to an assessment of productive vocabulary than to receptive vocabulary. This goes counter to most second language research, which indicates that productive vocabulary learning is more difficult than receptive vocabulary learning (e.g., Mondria & Wiersma, 2004). However, because participants here are older and more advanced in their language learning, perhaps this points to a ceiling effect on receptive vocabulary for advanced post-secondary ELs.

Pauwels' (2012) results indicate that the most effective study materials were a thematically-organized list with L1 glosses or with examples as these groups scored significantly higher on the post-test. These differences, however, disappeared on the delayed post-test. Thus, such study materials may only have short-term effects on productive vocabulary.

Tsai's (2011) results confirm previous research regarding best practices for vocabulary learning (e.g., Keisler & Bowers, 2012; Ozturk, 2015). Tsai's findings indicate that both face-to-face and courseware-based learners saw significant gains after a seven-week unit, though differences between groups were not significant. This suggests that quality vocabulary instruction is a more important consideration than the novelty or easy access of technology.

Findings here confirm that explicit instruction is a greater factor in productive academic vocabulary learning than technology use.

4.2.3. Technology

While technology provides ever-increasing modes of learning, the research examined here shows that simply substituting technology for traditional vocabulary learning strategies is not enough to effect lasting change; the more important consideration of using technology to learn academic vocabulary is how the words are studied.

4.2.3.1. SAMR model

Perhaps one of the key factors here is that because the research and education communities are still learning how to best use the most up-to-date technology for education, we have yet to design ideal tasks that modify and redefine in a way that truly helps learners. In other words, because we know a great deal about best practices of paper-based methods, we know what will work well when we simply substitute or augment those methods using technology. However, because we are still building our knowledge base regarding technology's role in education, we are still learning how to design the best modified and redefined tasks using technology. This coincides with growing research regarding how students and teachers perceive and utilize technology in higher education classrooms (e.g., Kay & Lauricella, 2011), including the use of social media platforms (e.g., Lewis, Fretwell, Ryan, & Parham, 2013).

Providing students with engaging means of studying and practicing an academic lexis could be powerful, particularly when those modified and redefined tasks are personalized and student-paced since most post-secondary students are learning academic English outside of explicit classroom instruction. Thus, this review serves as a call to researchers to design and

implement academic vocabulary learning interventions that modify and redefine instead of merely augmenting and substituting for paper-based methods.

In terms of best practices for academic vocabulary learning, using technology seems to benefit many learners, though it is most effective when combined with other well-established aspects of quality vocabulary instruction such as increasing the amount of learner involvement with the target material, combining embedded and isolated methods of study, providing multiple exposures to target words, and integrating available tools such as concordances.

4.2.4. Embedded Vs. Isolated

Findings from multiple included studies (Asmaa et al., 2015; Kiliçkaya & Krajka, 2010; Lin & Liou, 2009; Moskovsky et al., 2015; Poole, 2012; Rezaei & Karbalaei, 2013) point to the benefits of combining embedded academic vocabulary learning with explicit, isolated word learning. Asmaa et al. (2015) and Kiliçkaya and Krajka (2010) both used a combination of intentional target word learning within the context of larger reading passages. Both studies showed significant gains. Similarly, Moskovsky et al. (2015) and Lin and Liou (2009) found that combining embedded and isolated academic vocabulary learning produced gains in productive vocabulary as well as in vocabulary breadth and depth. These findings seem to confirm both the power of intentional, direct instruction as well as the benefits of embedded, contextualized vocabulary learning.

Findings from these studies coincide with those of August, Artzi, and Barr (2016) and Keisler and Bowers (2012) regarding the primacy of explicit instruction. At the same time, these included studies also confirm findings from Lesaux et al. (2014), who note the importance of learning academic vocabulary in authentic, text-based contexts.

Worthington and Nation (1996) suggest a combined approach using some adapted texts, some unsimplified texts, and extensive reading alongside explicit attention paid to a small number of purposefully decontextualized words. They found that a gradual, structured introduction of academic vocabulary solely through the texts introduced in a course would only allow students to learn about half of all 836 target words (from the University Word List) whereas a combined approach would provide opportunities to learn more academic words in the same amount of time.

Thus, just as Pinot-Shahov (2012) suggests viewing receptive and productive language along a continuum, perhaps a similar spectrum is needed here for understanding the interplay between embedded and isolated academic language instruction. On one end of the spectrum, words can be learned solely through lists and definitions with direct, explicit instruction; on the other end of the spectrum, words can be learned incidentally, without direct instruction, solely through reading and incidental exposure. But some learning takes place in the middle of that spectrum, where words may be highlighted or discussed while being learned within a larger context. Six of the studies included here indicate significant possible gains when combining isolated and embedded methods, specifically for academic word learning. Both word learning approaches have their place in academic language learning and appear to work best in tandem.

Interestingly, Moskovsky et al. (2015) showed that the order in which students study academic words (i.e., isolated first or embedded first) made a difference in vocabulary breadth. They found that the process of first studying target words in isolation at the morphemic level (i.e., bottom-up) then progressing to reading those words in context yielded significantly greater gains in vocabulary breadth than the reverse process (beginning with embedded words). While both studying processes yielded significant improvement from pre- to post-test scores, gains for

the bottom-up group were significantly higher than those for the top-down group. This further confirms the effectiveness of the isolated/embedded combination while also indicating the role that order plays when combining these vocabulary study methods.

Additionally, the type of vocabulary being learned may be better suited for either embedded or isolated vocabulary learning strategies. As noted above, when August et al. (2016) compared embedded vocabulary learning, they found that students learned significantly more discipline-specific academic words than they did general academic words. In other words, when reading words in context, students were better able to figure out the meaning of specialized academic terms than general academic vocabulary. Thus, the benefits of learning words in context versus in isolation may partly depend on the type of words being learned. This makes sense when considering the types of academic words students are exposed to while reading. Often, technical or discipline-specific terms are unique to certain fields and typically have enough contextual information to help determine their meanings while reading (e.g., diagrams, comparisons, glossary information); general academic terms, on the other hand, appear across disciplines and have less specific contexts to help students determine their meanings. While August et al.'s (2016) findings are from work with elementary students, this same principle of seeing gains with explicit vocabulary teaching combined with the benefits of utilizing embedded discipline-specific vocabulary would likely hold true with older students as well, as is indicated by the findings from the included studies.

4.2.5. Learner Involvement

As noted earlier, Stahl and Fairbanks (1986) found that activities requiring deeper processing yielded more effective vocabulary learning; similarly, Kim (2008) found that greater learner involvement yielded stronger initial scores as well as better retention for post-secondary

ELs. One question here is whether the same holds true when learning academic vocabulary, and findings from the included studies are a bit mixed as Pauwels (2012) proposes that students with an adequate amount of involvement outperform those studying with materials requiring greater involvement.

While findings from Rezaei and Karbalaei (2013) and Joseph et al. (2009) appear to conflict with those of Pauwels (2012), there are a few things to consider when reconciling these reports. First, it is important to note that while Pauwels describes offering six different sets of study materials to six groups, Pauwels gives no indication that participants were trained in how to use these materials. Participants appear to simply have been provided with study materials then asked to log how often they studied, the length of time they studied, and with which particular materials they studied. Thus, while some groups had more options for what to study, they may not have been trained in how to study. Perhaps with a bit of training, groups with materials requiring greater involvement may have seen greater gains.

It is also possible that participants who received more elaborate, involved study materials were overwhelmed by these, especially as it appears that participants were provided with no accountability or support in using them. This could have resulted in participants being less willing to engage in regular study of the target words. If this is the case, it may be more of a question of participant motivation than of how effective it is to study using tasks and materials that require greater student involvement.

Lastly, results from Pauwels' (2012) initial post-test did not persist; differences between groups disappeared on a 4-week delayed post-test. The group using L1 glosses, who showed the greatest initial gain, also saw the greatest score loss on the delayed test. The authors note that this likely indicates shallow learning of the target words.

Thus, while these findings somewhat conflict, there is not enough information to refute the long-held stance that greater learner involvement would be anything less than beneficial in academic vocabulary learning. In fact, the large effect sizes in Joseph et al.' (2009) study combined with the disappearing differences on Pauwels' (2012) delayed post-test seem to confirm the benefits of greater learner involvement with target academic terms.

4.2.6. Future Research

This review highlights the need for future research in several areas. First, there is a stark need for high-quality intervention studies of post-secondary academic vocabulary learning that follows the elements outlined by Towne and Shavelson (2002). Not only is this vital for future researchers and practitioners, it is also necessary to establish stronger research syntheses and reviews.

In particular, the current research into academic vocabulary for post-secondary ELs focuses almost exclusively on general academic words, not on discipline-specific or technical vocabulary. While both are vital for academic learning and participation, each has its own role. Research on the specific role of technical vocabulary for post-secondary English learners is currently underdeveloped. As the field continues to conduct high-quality research related to the academic vocabulary of post-secondary English learners, it is important to investigate both general academic and discipline-specific vocabulary.

Second, most of the studies included here use inauthentic outcome measures, intended largely to test if participants remember the target words of the intervention. While valuable in a way, how much more authentic and valuable would it be to assess participants' abilities to determine academic vocabulary that was not explicitly addressed in the intervention? Thus, one direction for future research would be to design more authentic outcome measures that gauge

participants' use of intervention strategies by having them apply these to new words in context. To establish initial construct validity and researcher trust, these authentic measures could initially be combined with well-cited, standardized assessments of academic vocabulary such as the Vocabulary Levels Test (VLT). The Academic Vocabulary Size Test (AVST) instrument proposed by Moskovsky et al. (2015) is a promising step in the right direction.

Third, as we continue to discover how to best use technology for educational purposes, we must continue researching its uses as related to academic vocabulary with post-secondary learners. At present, studies simply using technology to enhance vocabulary learning, based on the SAMR model, show greater effects than those using technology to modify and redefine tasks. While certainly possible that using technology to enhance learning is the best route, it may also be true that we have yet to discover, or accurately observe and report, some of the most effective ways to use technology for modifying and redefining current learning methods. Pushing those boundaries and continuing to test results will help establish our knowledge in this area.

5. CONCLUSIONS

This research synthesis makes several things clear while also alerting the research community to the need for further research in this area. Pertaining to intervention studies on post-secondary English learners' academic vocabulary, the research base has thus far concentrated on: general academic over discipline-specific vocabulary, EFL over ESL contexts, receptive vocabulary measures, convenience samples, and researcher-created instruments. In terms of best practices, post-secondary learners appear to benefit from academic vocabulary study that combines embedded and isolated practice of target words - particularly when using a bottom-up study method, utilizes available tools such as a concordance and dictionary, and provides metacognitive awareness through direct instruction of academic vocabulary learning strategies. Additionally, simply including technology in academic vocabulary study is not as important as how academic vocabulary words are taught and studied; instead, practitioners are encouraged to combine technology use with established methods of effective academic vocabulary learning. Future research is needed and greatly encouraged based on the limited number and quality of the studies included here.

REFERENCES

Alijany, M. S., Mansoory, N., & Divsar, H. (2015). The effect of reading academic texts on learning and retention of academic vocabulary: Investigating the role of model essay. *Modern Journal of Language Teaching Methods, 5*(2), 22-32.

Anderson, R. C, & Freebody, R (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Ángel, N. l., & García, J. (2017). Improving English language learners' academic writing: A multi-strategy approach to a multi-dimensional challenge. *GIST: Education & Learning Research Journal, 14*, 49-67.

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). A review of the literature on academic English: Implications for K-12 English language learners. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.

Asmaa, A., Noorizah Mohd, N., & Zaini, A. (2015). The effect of data driven learning on receptive vocabulary knowledge of Yemeni university learners. *3L: Southeast Asian Journal of English Language Studies, 21*(3), 13-24.

August, D., Artzi, L., & Barr, C. (2016). Helping ELLs meet standards in English language arts and science: An intervention focused on academic vocabulary. *Reading & Writing Quarterly, 32*(4), 373-396.

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

Azkarai, A., & Oliver, R. (2019). Negative feedback on task repetition: ESL vs. EFL child settings. *The Language Learning Journal, 47*(3), 269-280.

Bhatia, V.K. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 21-39). London: Longman.

Bailey, A. L. (2007). The language demands of school: Putting academic English to the test. New Haven: Yale University Press.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary?. *Journal of Adolescent & Adult Literacy, 54*(1), 4-12.

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary development*. New York: Guilford.

Blachowicz, C. L., & Fisher, P. (2004). Vocabulary lessons. *Educational Leadership, 61*(6), 66-69.

Bowers, P.N., Kirby, J.R., & Deacon, S.H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research, 80*(2), 144–179. doi:10.3102/0034654309359353

Bruton, A. (2009). The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly, 6*(4), 288-297.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *Future of Children, 21*(1), 103–127.

Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly, 45*(4), 464-487.

Carlo, M. S., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D. Lively, T. J., & White, C. E. (2008). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Journal of Education, 189*(1/2), 57-76.

Chalmers, T. C., Smith Jr, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., & Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials, 2*(1), 31-49.

Choudaha, R., & Chang, L. (2012). Trends in international student mobility. *World Education News & Reviews, 25(*2), 1-21. Retrieved from: http://www.uis.unesco.org/Library/Documents/research-trends-international-student-mobility-education-2012-en.pdf

Cisco, B. K., & Padrón, Y. (2012). Investigating vocabulary and reading strategies with middle grades English language learners: A research synthesis. *RMLE Online, 36*(4), 1-23.

Cobb, T. (2002). The web vocabulary profiler. Retrievable https://www.lextutor.ca/vp/eng/

Cooper, H. M. (1984). *The integrative research review: A systematic approach*. Beverly Hills, CA: Sage.

Cooper, H., & Hedges, L. V. (1994). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 3-14), New York, NY: Russell Sage Foundation.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. http://dx.doi.org/10.1037/0096-3445.104.3.268

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R D., Hornik, R C., et al. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Bilingual Education Paper Series, 3*(2), 1-69.

Curry, M. J. (2004). UCLA community college review: Academic literacy for English language learners. *Community College Review, 32*(2), 51-68.

Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., ... & Altman, D. G. (2003). Evaluating non-randomised intervention studies: Evaluation of checklists and scales for assessing quality of non-randomised studies (ch. 4). *Health Technology Assessment 7*(27), 23-42. http://researchonline.lshtm.ac.uk/id/eprint/8742

Dizon, G. (2016). Quizlet in the EFL classroom: Enhancing academic vocabulary acquisition of Japanese university students. *Teaching English With Technology, 16*(2), 40-56.

Eldridge, J. (2008). No, there isn't an 'academic vocabulary,' but...": A reader responds to K. Hyland and P. Tse's" Is there an 'academic vocabulary'?. *TESOL Quarterly, 42*(1), 109-113.

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes, 6*(1), 3-17.

Evans, S., & Morrison, B. (2010). The first term at university: Implications for EAP. *ELT Journal, 65*(4), 387-397.

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

Fisher, D., & Frey, N. (2008). *Word wise and content rich, grades 7-12: Five essential steps to teaching academic vocabulary*. Portsmouth, NH: Heinemann.

Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing, 8*(3), 243-264.

Flowerdew, J. (2001). Attitudes of journal editors to non-native speaker contributions. *TESOL Quarterly, 35*(1), 121-149.

Gámez, P.B., & Lesaux, N.K. (2012). The relation between exposure to sophisticated and complex language and early-adolescent English-only and language minority learners' vocabulary. *Child Development, 83*(4), 1316-1331.

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA: Journal of the American Medical Association, 295*(1), 90-93.

Garfield, E. (2015). The agony and the ecstasy--the history and meaning of the journal impact factor. *TEM Journal, 4*(3), 1–22.

Genesee, F., Lindholm-Leary, K. J., Saunder, W., & Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. *Journal of Education for Students Placed at Risk, 10*(4), 363-385.

Garrard, J. (2014). *Health sciences literature review made easy: The matrix method*. Burlington, Mass.: Jones & Bartlett Learning.

Glaser, B. G., & Strauss, A. L. (1965). The constant comparative method of qualitative analysis. *Social Problems, 12*(4), 436-445.

Glass, G. V. (1978). Letters: Reply to Mansfield and Busse. *Educational Researcher, 7*(1), 3–3. https://doi.org/10.3102/0013189X007001003

Golkar, M., & Yamini, M. (2007). Vocabulary, proficiency and reading comprehension. *The Reading Matrix, 7*(3), 88-112.

Gotwals, A. W., & Wright, T. (2017). From "Plants don't eat" to "Plants are producers". *Science and Children, 55*(3), 44-50.

Graves, M. F. (2006). Building a comprehensive vocabulary program. *New England Reading Association Journal, 42*(2), 1-7.

Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes, 35*, 105-115. doi:10.1016/j.jeap.2018.07.004.

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable?. *Journal of Literacy Research, 43*(2), 159-183.

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) Model: A critical review and suggestions for its use. *TechTrends, 60*(5), 433–441. https://doi-org.srv-proxy1.library.tamu.edu/10.1007/s11528-016-0091-y

Harmon, J. M., Wood, K. D., & Hedrick, W. B. (2008). Vocabulary instruction in middle and secondary content classrooms: Understandings and direction from research. In A.E.

Farstrup & S.J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 150-181). Newark, DE: International Reading Association.

Harper, C., & de Jong, E. (2004). Misconceptions about teaching English-language learners. *Journal of Adolescent & Adult Literacy, 48*, 152-162. doi:10.1598/JAAL.48.2.6

Hulstijn, J. H. (2003). Incidental and intentional learning. *The Handbook of Second Language Acquisition*, 349-381.

Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research, 28*(1), 149-158.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly, 41*(2), 235-253.

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies, 9*(2), 111-129.

Ioannidis, J.P.A., Fanelli, D., Dunne, D.D., & Goodman, S.N. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLoS Biology, 13*(10). doi: 10.1371/journal.pbio.1002264

James, M. A. (2014). Learning transfer in English-for-academic-purposes contexts: A systematic review of research. *Journal of English for Academic Purposes, 14*, 1-13.

Jiang, Y., Kuo, L., & Sonnenburg-Winkler, S. L. (2015). Morphological awareness and reading comprehension: A qualitative study with adult EFL learners. *International Journal of Language and Linguistics, 2*(5), 18-26.

Joseph, S. R., Watanabe, Y., Shiung, Y. J., Choi, B., & Robbins, C. (2009). Key aspects of computer assisted vocabulary learning (CAVL): Combined effects of media, sequencing and task type. *Research and Practice in Technology Enhanced Learning, 4*(2), 133-168.

Kanno, Y., & Cromley, J. G. (2013). English language learners' access to and attainment in postsecondary education. *TESOL Quarterly, 47*(1), 89-121.

Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning, 18*(4), 287-310.

Kay, R., & Lauricella, S. (2011). Exploring the benefits and challenges of using laptop computers in higher education classrooms: A formative analysis. *Canadian Journal of Learning and Technology / La revue Canadienne de l'apprentissage et de la technologie, 37*(1). Canadian Network for Innovation in Education. Retrieved January 3, 2019 from https://www.learntechlib.org/p/42756/

Keisler, L., & Bowers, E. (2012). Beyond vocabulary: Focusing the lens on academic language. *California Reader, 45*(3).

Kieffer, M. J., & Box, C. D. (2013). Derivational morphological awareness, academic vocabulary, and reading comprehension in linguistically diverse sixth graders. *Learning and Individual Differences, 24*, 168-175.

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading & Writing, 25*(2), 347-373.

Kiliçkaya, F., & Krajka, J. (2010). Comparative usefulness of online and traditional vocabulary learning. *Turkish Online Journal of Educational Technology-TOJET, 9*(2), 55-63.

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

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading & Writing, 25*, 389-410.

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

Kuehn, P. (1996). Assessment of academic literacy skills: Preparing minority and LEP (Limited English Proficient) students for postsecondary education. Fresno, CA: California State University (ERIC Document Reproduction Service No. ED415498).

Kuo, L. J., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist, 41*(3), 161-180.

Kuo, L. J., & Anderson, R. C. (2008). Conceptual and methodological issues in comparing metalinguistic awareness across languages. In K. Koda & A.M. Zehler (Eds.), *Learning to Read Across Languages* (pp. 51-79). New York: Routledge.

Lam, K., Chen, X., Geva, E., Luo, Y. C., & Li, H. (2012). The role of morphological awareness in reading achievement among young Chinese-speaking English language learners: A longitudinal study. *Reading & Writing, 25*(8), 1847-1872.

Larzelere, R. E., Cox Jr, R. B., & Swindle, T. M. (2015). Many replications do not causal inferences make: The need for critical replications to test competing explanations of nonrandomized studies. *Perspectives on Psychological Science, 10*(3), 380-389.

Laufer, B. (1992). How much lexis is necessary for reading comprehension?. In *Vocabulary and Applied Linguistics* (pp. 126-132). Palgrave Macmillan, UK.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399–436.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents evidence from a randomized field trial. *American Educational Research Journal*, 0002831214532165.

Lessard-Clouston, M. (2006). Breadth and depth specialized vocabulary learning in theology among native and non-native English speakers. *Canadian Modern Language Review, 63*(2), 175-198.

Lewis, C.C., Fretwell, C.E., Ryan, J., & Parham, J.B. (2013). Faculty use of established and emerging technologies in higher education: A unified theory of acceptance and use of technology perspective. *International Journal of Higher Education, 2*(2), 22-34. doi:10.5430/ijhe.v2n2p22

Lin, M.-C, & Liou, H.-C. (2009). Expansion of EFL academic vocabulary for writing via web-enhanced lexical instruction. *English Teaching & Learning, 33*(2), 95-146.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Madrigal-Hopes, D. L., Villavicencio, E., Foote, M. M., & Green, C. (2014). Transforming English language learners' work readiness: Case studies in explicit, work-specific vocabulary instruction. *Adult Learning, 25*(2), 47-56.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur?. *Perspectives on Psychological Science, 7*(6), 537-542.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335.

McBride-Chang, C., Tardif, T., Cho, J. R., Shu, H. U. A., Fletcher, P., Stokes, S. F., ... & Leung, K. (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics, 29*(03), 437-462.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, *20*(5), 522-535.

McKeown, M. G., Crosson, A. C., Artz, N. J., Sandora, C., & Beck, I. L. (2013). In the Media: Expanding Students' Experience With Academic Vocabulary. *Reading Teacher*, *67*(1), 45-53. doi:10.1002/TRTR.1179

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching*, *13*(3-4), 221-246.

Meara, P., & Alcoy, J.C.O. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language, 22*(1), 222–236.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group (2010). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *International Journal of Surgery 8*(5), 336-341. doi:10.1016/j.ijsu.2010.02.007

Moinzadeh, A., & Moslehpour, R. (2012). Depth and breadth of vocabulary knowledge: Which really matters in reading comprehension of Iranian EFL learners?. *Journal of Language Teaching & Research, 3*(5), 1015-1026. https://doi-org.srv-proxy1.library.tamu.edu/10.4304/jltr.3.5.1015-1026

Mondria, J. A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make. *Vocabulary in a Second Language: Selection, Acquisition, and Testing, 15*(1), 79-100.

Moon, Y. (2017). Are different types of vocabulary needed for comprehending general and academic texts?. *English Teaching, 72*(3), 47-68. doi:10.15858/engtea.72.3.201709.47

Moskovsky, C., Jiang, G., Libert, A., & Fagan, S. (2015). Bottom-up or top-down: English as a foreign language vocabulary instruction for Chinese university students. *TESOL Quarterly, 49*(2), 256-277. doi:10.1002/tesq.170

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

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*(1), 91-108.

Nation, I.S.P. (1990). *Teaching and Learning Vocabulary*. Boston, MA: 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.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*, 2nd ed. Cambridge: CUP.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*(1), 85-101.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528.

Ozturk, M. (2015). Vocabulary growth of the advanced EFL learner. *The Language Learning Journal, 43(*1), 94-109.

Paribakht, T.S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 174–200). Cambridge, UK: Cambridge University Press

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

Parks, M., & Oslick, M. E. (2018). Goldilocks and the engineers. *Science & Children*, *55*(6), 66-72.

Parsons, A.W., & Bryant, C.L. (2016). Deepening kindergarteners' science vocabulary: A design study. *The Journal of Educational Research*, *109*(4), 375-390. doi: 10.1080/00220671.2014.968913

Patten, M.L. (2012). *Understanding Research Methods: An Overview of the Essentials* (8th ed.). Glendale, CA: Pyrczak Publishing.

Pauwels, P. (2012). Vocabulary materials and study strategies at advanced level. *Language Learning Journal*, *40*(1), 47-63. doi:10.1080/09571736.2011.639899

Peachey, N. (2018). *Digital Tools for Teachers – Trainers' Edition V2*. Available from https://peacheypublications.com/digital-tools-for-teachers-trainers-edition-v-2

Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers, 4*(1), 37-45.

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal*, *29*(4), 679-693.

Portnoy, L. (2018, Feb 1). How SAMR and tech can help teachers truly transform assessment. EdSurge News. Retrieved from https://www.edsurge.com/news/2018-02-01-how-samr-and-tech-can-help-teachers-truly-transform-assessment

Proctor, C.P., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading & Writing, 24*, 517-544. doi:10.1007/s11145-009-9218-2

Proctor, C. P., & Mo, E. (2009). The relationship between cognate awareness and English comprehension among Spanish-English bilingual fourth grade students. *TESOL Quarterly, 43*(1), 126-136.

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading & Writing, 25*(7), 1635-1664. doi:10.1007/s11145-011-9336-5

Puentedura, R. R. (2013, May 29). SAMR: Moving from enhancement to transformation [Web log post]. Retrieved from http://www.hippasus.com/rrpweblog/archives/2012/08/14/SAMR_SixExemplars.pdf

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Ramirez, G., Chen, X., Geva, E., & Kiefer, H. (2010). Morphological awareness in Spanish-speaking English language learners: Within and cross-language effects on word reading. *Reading & Writing, 23*(3-4), 337-358.

Rashidi, N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL Learners. *Journal of Pan-Pacific Association of Applied Linguistics, 14*(1), 81-108.

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

Rezaei, F. S., & Karbalaei, A., (2013). The effect of vocabulary strategy training among autonomous and non-autonomous learners in Iranian EFL context. *European Online Journal of Natural and Social Sciences, 2*(2), 35-49.

Romrell, D., Kidder, L.C., & Wood, E. (2014). The SAMR model as a framework for evaluating mlearning. *Journal of Asynchronous Learning Networks, 18*(2), 1-15. Retrieved from http://files.eric.ed.gov/fulltext/EJ1036281.pdf

Santos, M. G. (2004). Some findings on the academic vocabulary skills of language-minority community college students. *Focus on Basics, 6*(D), 7-9.

Scarcella, R. (2003). Academic English: A conceptual framework. The University of California Linguistic Minority Research Institute Technical Report, Santa Barbara, 2003-1.

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

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.

Scott, J. A., Nagy, W. E., & Flinspach, S. L. (2008). More than merely words: Redefining vocabulary learning in a culturally and linguistically diverse society. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 182-210). Newark, DE: International Reading Association.

Shin, J.K. (2006). Ten helpful ideas for teaching English to young learners. *English Teaching Forum, 44*(2), 2-13.

Short, D., & Echevarria, J. (2005). Teacher skills to support English language learners. *Educational Leadership, 62*(4), 8-13.

Short, D. J., Fidelman, C. G., & Louguit, M. (2012). Developing academic language in English language learners through sheltered instruction. *TESOL Quarterly, 46*(2), 334-361.

Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology, 104*(2), 265.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328(5977), 450-452.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D.R. Olson & N. Torrance (Eds.), *The Cambridge Handbook of Literacy*, (pp. 112-133). Cambridge: Cambridge University Press.

Spycher, P. (2009). Learning academic language through science in two linguistically diverse Kindergarten classes. *The Elementary School Journal, 109*(4), 359-379. https://doi.org/10.1086/593938

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110.

Stahl, S.A., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.

Stewart, M. A., & Walker, K. (2017). English as a Second Language and World War II: Possibilities for language and historical learning. *TESOL Journal, 8*(1), 44-69.

Sylvén, L. K., & Ohlander, S. (2014). The CLISS project: Receptive vocabulary proficiency in CLIL and non-CLIL groups. *Moderna Språk, 108*(2), 81-119.

Taboada, A. (2009). English language learners, vocabulary, and reading comprehension: What we know and what we need to know. *College Reading Association Yearbook, 30*, 307-322.

Taboada, A., & Rutherford, V. (2011). Developing reading comprehension and academic vocabulary for English language learners through science content: A formative experiment. *Reading Psychology, 32*, 113-157. doi:10.1080/02702711003604468

Towne, L., & Shavelson, R. J. (2002). Scientific research in education. Washington, DC: National Research Council.

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading & Writing, 22*(9), 993–1019. doi:10.1007/s11145-008-9141-y

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal, 112*(3), 497-518.

Tsai, S.-C. (2011). Integration of multimedia courseware into ESL instruction for technological purposes in higher technical education. *Educational Technology & Society, 15*(2), 50-61.

Uccelli P., & Galloway E. P. (2017). Academic language across content areas: Lessons from an innovative assessment and from students' reflections about language. *Journal of Adolescent & Adult Literacy, 60*(4), 395–404. doi: 10.1002/jaal.553

Wang, Z. (2014). A correlation analysis on the depth and breadth of ESL learners' vocabulary knowledge and their overall linguistic competence. *Theory & Practice in Language Studies, 4*(12).

Wang, M. & Bakken, L. L. (2004). An academic writing needs assessment of English-as-a-Second-Language clinical investigators. *Journal of Continuing Education in the Health Professions, 24,* 181-189.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79-95.

Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review, 68*(3), 267-290.

Worthington, D., & Nation, P. (1996). Using texts to sequence the introduction of new vocabulary in an EAP Course. *RELC Journal, 27*(2), 1–11. https://doi.org/10.1177/003368829602700201

Wortman, P. M. (1983). Evaluation research: A methodological perspective. *Annual Review of Psychology, 34*(1), 223-260.

Wortman, P. M. (1994). Judging research quality. In H. Cooper & L.V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 97-109). New York, NY: Russell Sage Foundation.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215-229.

Zhou, A. A. (2009). What adult ESL learners say about improving grammar and vocabulary in their writing for academic purposes. *Language Awareness, 18*(1), 31-46.

Zhang, H., Song, W., & Burston, J. (2011). Reexamining the effectiveness of vocabulary learning via mobile phones. *Turkish Online Journal of Educational Technology,10*(3), 203-214.

Zwiers, J. (2007). Teacher practices and perspectives for developing academic language. *International Journal of Applied Linguistics, 17*(1), 93–116. doi:10.1111/j.1473-4192.2007.00135.x

APPENDIX

ARTICLE MATRIX

Citation	Year	Method & Design	Who are the participants?	What is the context?	What is the intervention?	Vocab in context or discrete?	How long is the treat-ment?	What are the outcomes? How were they measured?	Did they use techno-logy?	What were the effects of Intervention?	Effect size(s)	What indicates the quality of this study?	Additional Notes
Alijany, M. S., Mansoory, N., & Divsar, H. (2015). The effect of reading academic texts on learning and retention of academic vocabulary: Investigating the role of model essay. Modern *Journal of Language Teaching Methods, 5*(2), 22-32.	2015	Quan (but just t-tests?) exp/ctrl	▪ 40 Iranian EFL learners; ctrl = 20, exp = 20 ▪ all male, ages 17-21 ▪ pre-university and university students ▪ intermediate level proficiency (based on Oxford Solution Proficiency Test ▪ have at least 5 total years of English learning ▪ in last term of English	▪ EFL ▪ Nikan language institute in Gilan, Iran	Reading academic texts with embedded AWL words Exp group - read academic model essays comprised of 8.5-11% AWL words Control group read model essays from the IELTS (presumably without the high percentage of AWL words) Neither group was provided with vocabulary instruction --> INCIDENTAL vocabulary learning was the focus Novelty test prepared from 97 lexical items chosen from academic model essays used in Exp group --> students asked to write synonyms for words they knew; known words eliminated from study Pre-test40 MCQ contrived using 81 academic word list words from study intervention (model essays) - choose the word that matches the provided meaning? Read 15 model academic essays over 9 weeks - ?	Context	9 weeks, 17 sessions	**Receptive** academic vocab learning (MCQ) Academic vocab retention after 1 month Knowledge of the same academic words asked about on the pre-test (same test) Delayed test 1 month later Researcher-created test (40 MCQ using 81 academic words from the model essays)	No	Exp group had significantly higher scores on their post-test compared to the pre-test (t=-8.39, p=.001) Exp group scored higher on post-test than on delayed post-test, but delayed scores were still higher than pre-test scores Exp group scored significantly higher than control group on post-test (t=-6.34, p=.001) and delayed post-test (t= -6.43, p=.001) My calculation (ctrl vs. exp post-test based on t-statistic): d=-2.00488	Receptive -2.00488	only use t-test analysis do not describe the actual intervention procedure written English in this article has several errors – the abstract has no capital letters?	Intermediate level proficiency (based on Oxford Solution Proficiency Test – but do not state what the "required score" was to be considered intermediate);
Ángel, N. l., & García, J. (2017). Improving English language learners' academic writing: A multi-strategy approach to a multi-dimensional challenge. *GIST: Education & Learning Research Journal, 14*, 49-67.	2017	Quantitative measures of academic writing	▪ 16 future English teachers ▪ Ages 18-25 ▪ L1 = Spanish ▪ English proficiency = between A2 and B1 [Advanced; A1 is highest level]	▪ EFL ▪ Colombia ▪ medium-sized, public, co-ed university	Academic Writing course ▪ 4 hours/week ▪ 1 semester 4 features: ▪ Process writing approach ▪ systematized feedback from prof & peers ▪ on-going tutoring in writing lab (virtual platform with materials) ▪ repeated test-taking practice of TOEFL writing Vocabulary - topics were academic in nature, lab tutorials built around academic writing textbooks; TOEFL practice of academic vocab based on 2-million, academic word data bank reported by Fox et al. (2007)	Context Vocab within their writing/essays	1 semester 4 hours/week	**Productive** (written essays) All Writing Outcomes: Discourse, Organization, Syntax, Vocabulary, Conventions "vocabulary focuses on the range of academic lexicon" (p. 58) Essays were scored on 1-5 scale (1=low, 5=high) based on rubrics; assumingly scored by the professors/researcher	Yes; virtual platform for writing lab where they reviewed materials	No true pre-post Students seemed to excel 2 Cohorts (one in spring, one in fall) The Cohort who took the course in the fall scored higher than Cohort 1 (spring cohort); assumingly they benefited from the professor's reflection and course modifications *Specifically in vocabulary, Cohort 1 received an average of 4.75 almost every student got a 4.75) Cohort 2 received an average of 4.8125	--	Small sample size (2 sets of 8 students each = 16 total) They averaged the writings in order to compare across groups instead of doing a pre-post comparison More descriptive than actually showing effectiveness No explanation of how essays were scored or by whom; no inter-rater reliability reported -- only 1 rater, presumably?	

Asmaa, A., Noorizah Mohd, N., & Zaini, A. (2015). The effect of data driven learning on receptive vocabulary knowledge of Yemeni university learners. *3L: Southeast Asian Journal of English Language Studies, 21*(3), 13-24.	2015	Quan (t-tests only) exp/ctrl	▪60 university students in Yemen ▪L1 Arabic speakers ▪All female; ages 21-25 ▪passed English entry exam, which parallels intermediate proficiency	▪EFL ▪University course – 2nd year reading course ▪Course in Department of English, by educational faculty (?) ▪2 intact classes – 30 in control (dictionary group) and 30 in experimental (DDL)	Both groups were taking the same reading course All were taught the same unit 1 of the reading course. Then all were introduced to the meanings and types of collocations. Experimental group went through 2 training sessions where they were taught to observe the target words in context and use concordance lines. Used 5 words/collocations that were not on the post-tests to train. Control group had same list but only used dictionary and collocation materials (not DDL activities) 1. Experimental group used DDL (data-driven learning) with ~~an electronic corpus~~ → of printed worksheets - 4 printed activities, each involving 2 parts (word meaning & word collocation) - 30 target words were presented in the context of 5 complete sentences (concordance) 2. Dictionary group – provided a dictionary for students – most used Oxford Word Power or they downloaded the Longman dictionary to their phones -- it sounds like some target words did have examples for both groups?? But then students were asked to look up the words in the context of the passage for the target words that weren't accompanied by an example - Control group used the Oxford collocation dictionary to learn collocations → 4 printouts that included an example for each collocation as well as an activity for each set of target collocation words	Both	2x each week for 2 hours X 4 weeks = 16 hours (Full semester was 15 weeks, but this was only 1 unit?)	**Receptive** academic vocabulary knowledge of collocations – knowledge rating + choose correct collocation as used in sentences Pretest – Posttest – Delayed Posttest (all 3 identical, except delayed post was ordered differently) Researcher-created Tests: 50 words taken from course reading book after running them through online academic word tool to pull out academic words; these 50 high-frequency academic words were then matched with their 2 most frequent collocations 2 parts: 1. knowledge rating – choose either 1: I don't know what this word means, or 2: I know this word. It means (can explain in English or Arabic) 2. choose the correct collocation (of the words they DO know – that they selected choice 2 in the previous section); 2 sentences that use the collocations, 3 choices for each sentence)	Yes? electronic corpus of texts representing authentic language; HOWEVER, the DDL group used printed worksheets in class (not enough computers in the lab and no internet-??); dictionary group could use printed dictionary or Longman dictionary downloaded to their phones Data-driven learning using COCA (Corpus of Contemporary American English) – free online usage, public access to corpus inside and outside of classroom, help guides available, easy search for vocab/collocation	No sign. Difference in pre-test between the 2 groups (independent t-test) Both groups improved from pre to post Experimental group had significantly higher **post-test mean** scores than control group (t=3.155, p=.004) My calculation: d=0.814617 Delayed post – both groups further improved (delayed post administered just after class mid-term, so students had just reviewed the vocab) Experimental group had significantly higher scores on the delayed post-test compared to control group mean (t=2.97; p=.006) Looking just at **collocation scores**, the experimental group had sign. Higher post-test scores compared to control group (t=3.71, p=.021) Following post-test trend, the experimental group had sign higher scores on the delayed collocation section compared to the control group (t=2.443, p=.021) Both study methods yield gains, but DDL is more effective than dictionary use alone Both groups do better in the delayed posttest for definitions because it occurs after they've just taken a mid-term and reviewed the material – I think the more important find here is that both interventions improved scores (dictionary and DDL), and that continued review of the material increased scores even more (p 20) Scores are similar for the collocation tests – both groups improve on the posttest and them improved again on the delayed post (p. 21)	Receptive 0.81462	No external measure of language proficiency – researcher stated all participants were similar b/c they all passed the same entry exam Procedures are confusing and would be difficult to replicate Again, only using t-tests… and only comparing the two groups; I don't think they ever actually did a paired t-test for each group to see the individual differences…	I'm still not entirely sure what DDL is… I know it stands for data-driven learning, but the authors don't really explain how their intervention is DDL I'm also confused on why they went into such detail in the lit. review about computer-assisted learning and technology when they didn't actually use any?! Authors note that their results differ from Poole (2012), which said that DDL instruction had an insignificant effect

Dizon, G. (2016). Quizlet in the EFL classroom: Enhancing academic vocabulary acquisition of Japanese university students. *Teaching English With Technology, 16*(2), 40-56.	2016	Quan t-test + questionnaire Pre/post	▪9 students ▪2nd year at university ▪Belonged to Faculty of Foreign Studies ▪"among the highest level of English learners within the faculty based on their TOEIC exam scores" (p. 46) ▪Convenience sample	▪EFL ▪Japanese university ▪Enrolled in a course	Quizlet AWL as target vocab During class, sub-lists of AWL words were introduced starting with the most frequent words; 20-30 words per session were given; students were allowed 10 mins at the start of class to review and study these words. They could use their phones or classroom computers with Quizlet. Were encouraged to review the words outside of class Questionnaire showed a little over 20% studying for less than 20 mins outside of class each week, about 45% studying 20-40 mins, and about 33% studying 40-60 mins; no one studied more than 60 minutes outside of class each week	Discrete Definition in English and in L1, pronunciation	3x/week for 10 weeks	**Receptive** (VLT) **Breadth** VLT, academic vocab section; Version 1 as pretest, Version 2 as posttest	Yes Questionnaire showed that 65% preferred using Quizlet on their smartphones to 35% who preferred computer	Scores improved significantly from pre to post Pre M = 20.33 (SD 5.55) Post M = 23.56 (SD 5.34) T=-2.64, p=0.03	Receptive 1.24451	Small sample size No control group Only 1 measure of vocab Questionnaire asked about study time – all self-reported	Did the instructor/researcher use the target words during the class or have students do anything with these words other than view them on Quizlet (writing assignment, readings, etc.)? Study is more about the questionnaire than the vocab itself; they care about validating their constructs and present more about the questionnaire tables thn the vocab pre-post
Joseph, S. R., Watanabe, Y., Shiung, Y. J., Choi, B., & Robbins, C. (2009). Key aspects of computer assisted vocabulary learning (CAVL): Combined effects of media, sequencing and task type. *Research and Practice in Technology Enhanced Learning, 4*(2), 133-168.	2009	Quan percentages, boxplots, ANOVA, Wilcox's (non-parametric) 2 groups – kind of like exp/ctrl	▪L1 = Japanese ▪ages 18-25 ▪10 male, 26 female initially ▪*Major attrition ▪from various classes at the college ▪group assignment based on training schedule availability ▪n=19 – iKnow (7 male, 12 female ▪n=13 – iTango	▪ESL ▪studying English at TransPacific Hawaii College (2 year college) ▪Not related to a class (individual)	2 different softwares – iTango and IKnow iTango – material presented in 10 lists of 25 items each; no structured learning process (had same text content as iKnow with definition and example sentences) iKnow – encouraged active recall through quizzes (recall and MCQ), audio support for pronunciation practice, adapted pace based on individual user, sequenced re-presentation of material iTango group essentially had lists of words without structured presentation Both groups had the same words and could study whenever and for however long they chose as long as they spent at least 1.5 hours each week for 4 weeks	Discrete	4 weeks Minimum of 1.5 hours per week → minimum study time = 6 hours	Knowledge of words from AWL (list of 250 words) – **receptive** (MCQ) and **productive** (recall given synonyms) Researcher-created test with 2 parts: 1. 75 recall items – given 2 synonyms, participants had to type in the most appropriate English word 2. 75 MCQ – 6 possible responses: correct def., 3 distractors, "I don't know," and "none of the above" Delayed post-test done 5 weeks after initial post-test	Yes the 2 different software applications were the bases of the study and the marked difference between the two groups ▪Self-paced ▪Target word presentation differed They claim the technologies are different based on task type, media, and scheduling (?)	Both groups improved significantly from pre to post recall test (iKnow improved by 31%; iTango improved by 4%) iKnow group marginally improved from pre to post in MCQ test (10%) iTango group scored lower than the pre-test on MCQ (-4%) MCQ Post: iKnow – 83% (SD 15%) iTango – 66% (SD 11%) Recall Post: iKnow – 39% (SD 26%) iTango – 8% (SD 4%) Delayed post-test scores were lower for both groups, though the iKnow groups still scored higher than the iTango group Significant effect of group (Q=19.11, p<.0001), test (Q=335.46, p<.0001), and time (Q=6.99, p<.001) Significant effects of Group*Test (Q=7.25, p<.001) and Group*Time (Q=3.75,p=.024) but no sign effect of Test*Time or Group*Test*Time Effect sizes: recall = 1.67; MCQ (rec) = 1.29 Delayed test effect size: recall = 1.37; MCQ = 1.35	Receptive 1.29 Productive 1.67	**Weaknesses** – attrition; technical issue during pre-test meant that iTango group saw fewer words than the iKnow group on the pre-test… so scores were converted to percentages instead of raw numbers **Strengths** – full report of those dropped from the study, including how those factors might change the group detailed procedure including directions in appendix – could actually replicate this study They note several of their own issues/limitations (p. 157)	Attrition – 4 students eliminated from the iTango group when they didn't meet the weekly 1.5 hours of study ● Might speak to how boring/interesting the two apps were; students weren't motivated to use the one that just presented lists iKnow used recall and MCQ quizzes, which were the types of assessment used here – could part of the gain be due to familiarity with the test type? Noted history of participants including language learning experiences, avg. length of stay in English speaking countries, avg. time in Hawaii, time taking college level non-ESL courses – see p. 148 & table 3

													
Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning, 18*(4), 287-310.	2005	Quan Item analysis on pretest scores – item difficulty & item discrimination exp/ctrl	▪18 undergraduates ▪n=9 exp; n=9 control ▪11 male, 7 female ▪From Korea (5), Indonesia (6), Malaysia (2), Hong Kong (3), Thailand (1), Mexico (1) ▪TOEFL scores between 180 and 250 ▪inadequate score in composition writing for English Placement Test (EPT) and were now enrolled in writing course for ESL undergrads (101C) ▪randomly assigned to ctrl or exp	▪ESL ▪USA - Midwestern research university ▪Writing course for ESL undergrads (101C) ▪Training session and questionnaire conducted by course instructor ▪Vocabulary tasks conducted by primary researcher	all students had the same list of target words: ▪30 AWL words - previous instructors for the course compiled a list of 30 AWL words that would be most appropriate to complete the writing task both groups used online dictionary - dictionary.com Exp group also used an online concordancer (Tom Cobb's Compleat Lexical Tutor) – uses the BNC written corpus After pretest analysis, 23 of the words were used in the <u>vocab tasks</u>, which were practiced during class time: 1. Cloze task sentence completion (given word bank to fill in for part A, given MCQ to fill in for part B) 2. Sentence-building task (asked to create a sentence using the target words; needed to specific the word form they used (noun, verb, etc.)	Discrete	1x/week for 1 semester – 14 weeks-ish?	**Receptive** (cloze task with MCQ) and **productive** (sentence-building, writing task) knowledge of academic vocabulary – measured in discrete tests and in context of written essay *Don't have a true pre-post for any 1 measure? But they have control vs. experimental comparisons for every measure* <u>Questionnaire</u> - years of exposure to English, vocab learning strategies, experience with concordancing <u>Pretest</u> - included all 30 target words; receptive knowledge - each word used in 4 sentences, but only one where it was correctly used After pretest analysis, 23 of the words were used in the <u>vocab tasks</u>, which were practiced during class time: 1. Cloze task sentence completion 2. Sentence-building task Compared scores on vocab tasks to compare groups <u>Writing task</u> – completed outside of class, though they discussed them in class and did self and peer evaluations ▪Prompt: "analyzing an issue;" asked students to write about 550 words (2 pages, typed & double-spaced) and use at least 4 published sources, explaining an issue from both sides; students were handed a list of 23 target academic	Yes computer software – concordance program, online dictionary, Camtasia (screen capture)	**Vocab Tasks & Tool Use**: Exp group outperformed control group in each vocab task and in overall performance (though total score difference was not statistically significant) Cloze (Rec): Exp: 18.67 (SD 4.87) Ctrl: 15.89 (SD 3.06 t=1.45, p=0.1664 **[NOT significant]** Sentence build (Pro): Exp: 30.33 (SD 7.81) Ctrl: 28.33 (SD 10.01) t=0.4726, p=0.6429 **[NOT significant]** Writing task (Pro): Exp: 3.11 (SD=3.59) Ctrl: 2.44 (SD=2.50) t=0.4595, p=0.6521 **[NOT significant]** Exp: 168 point – Cloze 273 points in sentence-building Control: 143 points – Cloze 255 points in sentence-building Interaction with online tools: Exp – used dictionary more than concordance (7.44 vs. 10.56); **Writing Task**: Exp group had more attempts and significantly more correctly used academic words than control group (used percentage of correct word use to calculate) Exp = 78%, Mean: 3.11 (SD=3.59) Control = 67%, Mean: 2.44 (SD=2.50)	Receptive n.s. Productive n.s. Productive (writing) n.s.	**Weaknesses –** Don't clearly report # of control vs. # of exp (inferred from the results) or how they kept the groups "separate" (so there wasn't any crossover of using the concordancing software -they were instructed in separate rooms); writing task was only evaluated by the researcher and the instructor – better to get an outside source to help evaluate for reliability; Not much of an analysis… **Strengths** – included all measures in appendices; transparent about some aspects of the process, could potentially replicate	It seems odd that they ended up using the vocab tasks to evaluate the students. If these were in-class, practice exercises, why did they score them? Unless they had planned to use them in the analysis and then something went wrong… So, the exp group used the dictionary more than the concordance and the control group had access to the same dictionary…is the difference in the fact that one group had more tools available (even if they didn't use them)??

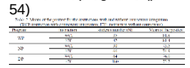
								words and encouraged to use them in this essay Post-questionnaire – 10 questions about computer experience, benefit of using concordancer and/or dictionary in vocab tasks, and if they used the concordancer and/or dictionary in their writing task					
Kiliçkaya, F., & Krajka, J. (2010). Comparative usefulness of online and traditional vocabulary learning. *Turkish Online Journal of Educational Technology-TOJET, 9*(2), 55-63.	2010	Quan t-test comparing post-test scores exp/ctrl	▪ 38 students (20 control, 18 exp) ▪ Different departments, studying English to pass university's proficiency exam ▪ Ages 17-19	▪ EFL ▪ Ankara, Turkey ▪ Private university ▪ Upper-intermediate Academic English class	WordChamp – web reader with dictionary capability (glossing) – "online glossing tool" (p. 62) WordChamp (experimental) – this group practiced their contextualized vocab using WordChamp Traditional instruction group (control) – used vocabulary notebooks, cards, and paper dictionary as they practiced learning vocab with 10 reading passages Both groups – same passages and vocab; both groups regularly reviewed vocab items; both groups had the same instructor and met for the same amount of time	Both	5 weeks, 3 hours each week (15 total hours)	**Receptive** (MCQ) Performance on a post-test and delayed post-test evaluating vocabulary knowledge Pretest, post-test, delayed posttest Test developed from previous proficiency exams 10 academic readings with 5 MCQs about the vocab in the readings ▪ 5 total questions? Or 5 questions for each passage? Unclear from article Items were scored as either correct or incorrect No sign. Difference between groups on pre-test Significant differences on both post-test and delayed post-test with experimental group performing better than control	Yes Compared Online vocab learning (WordChamp) to "traditional" vocab learning (in class) Only used it passively (in my opinion) – the students looked up words as they read them then had to identify the definitions of words on an assessment	Experimental group scored significantly higher than the control group in both the post-test and the delayed posttest. Post-test: Control mean: 32.7500 (SD 4.92977) Experimental mean: 38.3889 (SD6.21326) t = -3.114 (p=0.004) d=-1.01171 g=-0.99049 Delayed post-test: Control mean: 29.3500 (SD 5.21410) Experimental mean: 36.1111 (SD 6.13465) t = -3.672 (p=0.001) d=-1.19301 g=-1.16798	Receptive -0.99049	**Weaknesses** – Writing is not great – major run-on sentence as the first line of the abstract; other noticeable, distracting errors Mislabeled graphs (post-test results listed as "pretest" in one column) Never really explore the other potential factors involved – novelty of the technology, ease of use, amount of coaching; conclusion seems pretty simplistic (the technology worked better!) Only used the technology for passive recall – never asked students to produce the vocabulary Never gave a sample item from the assessment to show what kinds of definitions were given; also never discussed if they used the same passages during the practice that they used during the actual tests – did they have the context?	
Lessard-Clouston, M. (2006). Breadth and depth specialized vocabulary learning in theology among native and non-native English speakers. *Canadian Modern Language Review, 63*(2), 175-198.	2006	Mixed? Uses quan & qual; Author calls it case study… only gave means and SDs + percentages	▪ 12 graduate students ▪ 7 native English speakers ▪ 5 non-native (4 Cantonese, 1 Mandarin); all had been in Canada between 7 months and 10 years except 1 student who immigrated from US 1	▪ ESL ▪ Theology graduate program ▪ Canada ▪ Intro to theology class	Graduate course – all were taking an intro to theology class; had same instructor	Context	Semester	**Receptive** (Word ID) and **Productive** (Knowledge scale, use in sentence) Test of Theological Language (TTL) assessing theological vocabulary knowledge	No Not mentioned	Pre-test: overall mean on WI = NNES = 77.96%, NES = 85.07% 82.11% for whole group Pre-test Overall mean on VKS = 63%; NNES = 57.20%, NES =67.14% Post-test: Overall mean on WI = NNES = 87.17%, NES = 87.79%	Receptive 0.71785 Productive n.s.	**Strengths**: Compiled vocab from the sources participants would already be used for the class Piloted TTL measure with 3 NES and 4 NNES; thoughtful about the measure – didn't	Notes an increased gap between NES and NNES on the VKS section, though both groups improved in that area; NES improved far greater than NNES – primary difference was that NES offered more example sentences by far than NNES participants

		Wilcoxon signed rank test Pre-Post	week before class started ▪ How long had he been in the US? ▪ Studying theology ▪ Ages 25-over 55					▪ Developed using class notes from 2 previous students + handouts from the course + the 4 course textbooks ▪ Compiled a list of over 100 specialized theological terms TTL assessed: 1. **Breadth** (WI – word identification) - identify all words believed to be theological out of 100 (60 real, 40 from other disciplines); 2. **Depth** (VKS – Vocab Knowledge Scales) - Use knowledge scale to indicate knowledge of 10 real theological terms by using in a sentence Pre-test conducted within first 2 weeks of course; researcher met individually with each participant; took ~20 minutes to complete Same test used as pre and post-test		Overall mean on VKS = 77.50%; NNES = 63.60%, NES = 87.43% Both groups (NES and NNESs) had similar results in the breadth section. Both increased scores in depth section, but NES increased by more than 20% while NNESs increased to only 63.60% Wilcoxon signed rank test (one-tailed) WI Z value = -2.04, p<.02 VKS Z value = -2.85, p<.002 Combined group increased overall My calculations: NES, Pre-Post on WI: t=3.0246, p=0.0390 d=0.717853 NES, Pre-post on VKS t=-6.40000, p=0.1254 [NOT Significant]		just use pseudowords as distractors because other studies had done so; he compiled lists of specialized words from 4 other fields to use as distractors; provided reasoning and references to justify Provided measures in appendix Up front about an error in the test and how he accounted for it (hamartiology was spelled "harmatology" and considered a distractor now) Inter-rater agreement of 92% **Weaknesses**: Use of breadth/depth when it seems more like receptive/productive vocab Test retest effects (same test) No control group Small sample size	
Lin, M.-C, & Liou, H.-C. (2009). Expansion of EFL academic vocabulary for writing via web-enhanced lexical instruction. *English Teaching & Learning, 33*(2), 95-146.	2009	Mixed Quan ▪ t-test ▪ one way repeated measures ANOVA Qual ▪ Writing quality using rubric Pre-Post	▪ 25 EFL learners ▪ 3rd year of college ▪ L1 Mandarin ▪ At least 8 years of English instruction + required writing courses in first 2 years of college	EFL In-class – all taking "Reading and Writing II" course	3 main features: 1. Explicit vocab instruction re: academic words (wordlists, weekly lecture notes, weekly reading with highlighted words, concordances) 2. Online quizzes 3. Pair writing and individual lexical logs Target words – from each of the 8 sublists of the AWL Class procedure: ▪ Go over weekly target words ▪ Show the words in readings/context ▪ Learn collocations using the concordance ▪ Use target words in paired discussions (with reading comprehension questions) and in a paired writing assignment (2 people wrote 1 paragraph) ▪ Individual vocabulary log with reflection, 5 words per session (definition, form, usage)	Both	8 weeks 2 classes per week (100 minutes) Total: 800 minutes	**Receptive** (VLT) and **Productive** (VKS + use in sentence; writing task) Vocabulary size test (**breadth**) – Vocabulary Levels Test (VLT) – academic words section: 30 items Vocabulary **depth** test – Vocabulary Knowledges Scales (VKS) – 15 words from the AWL (2 from sublists 1-7 and 1 from sublist 8); wrote definition for each word + sentence using target words (60 minutes) Timed writing task – identical prompt in pretest, posttest, and delayed posttest (describing a buying decision); 60 minutes, paper and pencil	Yes "web-enhanced lexical instruction" (p. 96) Moodle Online quizzes (fill in the blank, crosswords) TANGO – web concordance Cambridge online dictionary for AWL AWL Highlighter - ?	Difference in lexical depth but not in lexical size (since pretest scores were already so high for size, ceiling effect) Vocab Tests: *Pretests* ▪ VLT: M = 57.44/60 (SD 4.3787) ▪ VKS: M = 69.48/90, 77.2% (SD 10.3082) *Posttests* ▪ VLT: M = 58.08 (SD 4.4527) ▪ VKS: M = 81.6 (SD 7.3256) t-*tests* ▪ VLT: -1.154 (p>0.05) ▪ VKS: -9.302 (p<0.05) Essays: ▪ Wrote most # of words on posttest, delayed posttest # was still higher than pretest ▪ Included more word families in posttest than in pretest and had the most	Receptive n.s. Productive 1.35538 Productive (writing) 2.46699	**Strengths**: Lit review is solid Checked academic words in essays with a prof. of applied linguistics Thoughtful analysis – computes results for the overall group and then also looks at individual essay results and their use of AWL words **Weaknesses**: The researcher and 1 other TEFL grad student were the only ones who rated/scored the VKS exams – what were the criteria? What made a 6 point answer from a 1 point answer? Paper and pencil essay when you're	Look up the AWL Highlighter (http://www.nottingham.ac.uk/~alzsh3/acvocab/awlhighlighter.htm) With all of the different strategies they were using (log, paired writing, concordance, etc.), how can they definitively attribute learning to the online components? Did a one-paired t-test for inter-rater reliability? Raters treated as the 2 variables? (p. 111) ANOVA of 5 subcomponents of rubric Used Canonical analysis to show relationship between lexical and writing skills

								▪ Scored using the Lexical Frequency Profile and the ESL Composition Profile ▪ ANOVA of 5 subcomponents of rubric ▪ Canonical Analysis (Multivariate Exploratory techniques) Delayed questionnaire – 24 questions re: perceptions of instructional effectiveness (design of online materials, vocab learning & writing, future vocab learning); 5 point Likert scale (strongly agree – strongly disagree)		word families in delayed posttest ▪ AWL ratio increased from pretest to posttest while the other 3 lists declined; all but 1000 list declined between posttest and delayed post, but AWL declined the least, 100-word-level rose slightly ▪ ANOVA – F(2, 72) = 8.5188, p< 0.001 – post-hoc showed AWL use in pretest sign. Lower than both posttests *Writing Quality (ESL comp rubric)*: Pretest: M=68.88 (SD 3.571) Posttest: M=80.64 (SD 5.718) Delayed: M=80.14 (SD 5.338) ▪ ANOVA of 5 rubric sub-components – all significant except "mechanics" ▪ Vocabulary section: Pre-test: 12.96 (SD 0.889) Post-test: 16.22 (SD 1.444) Delayed: 15.84 (SD 1.048) Questionnaire: ▪ Moderately positive attitudes to the online vocab instruction		doing an online project? No control group	
Moskovsky, C., Jiang, G., Libert, A., & Fagan, S. (2015). Bottom-up or top-down: English as a foreign language vocabulary instruction for Chinese university students. *TESOL Quarterly, 49*(2), 256-277. doi:10.1002/tesq.170	2015	Pre-post between groups design Quasi-experimental No control group ANOVAs, t-test post-hoc, ANCOVA	▪ 120 students ▪ 1st year at university ▪ Pretest showed lexical abilities to be low - ?? ▪ Authors state participants had similar English proficiency, ed. Background, and age based on university's entrance exam and selection criteria (p. 262) ▪ Matched groups	▪ EFL ▪ China ▪ Hebei Normal University	Bottom-up vs. top-down emphasis learning AWL words From p. 265: Each lesson was structured in six distinct steps (or events): • Event 1: Introducing the spoken and written form, as well as the morphemic structure, of the target word • Event 2: Introducing the definition and L1 translation of the target word • Event 3: Introducing related forms from the same lexical family • Event 4: Introducing different phrases and collocations of the target word • Event 5: Presenting the target word at the sentence level • Event 6: Presenting the target word at the level of the whole context Bottom-up group (A) went from 1 to 6; Top-down (B) went from 6 to 1	Both	48 hours 8 weeks 6 hrs/week	**Receptive** & **Productive** 1. Academic Vocabulary Size Test [AVST] (loosely based on VLT and VKS) (**breadth**) ▪ 50 words from AWL ▪ √ if they recognized a word followed by at least 1 translation equivalent, ? if they thought it was familiar, x if they didn't recognize it at all ▪ Scored like VKS 1 point if correct, 0.5 points if checked but incorrect or if ?, x =0 points 2. Controlled Productive Knowledge Test [CPKT] (Nation's Productive Level Tet) ▪ 18 sentences with 1 target word each	No? Not explicitly mentioned	Both groups showed gains from pre to posttest – both sign. Higher in posttest (both showed large effect sizes for pre-post differences) AVST: Bottom-up Pre: 29.65 (SD 4.88) Post: 44.82 (SD 3.40) t=-20.13, p=.000 Top-down Pre: 27.67 (SD 6.38) Post: 42.70 (SD 4.20) t=-20.13, p=.000 Bottom-up group scored significantly higher than top-down group; ANCOVA used to control for slight difference btwn groups at pretest – still showed sign difference $F(2,117) = 6.01$, $p=.02$, $n^2=.05$ CPKT: Bottom-up group scored slightly higher but the difference was not sign	Productive n.s.	Strengths: Well-written High quality journal – TESOL Quarterly Very thorough in reporting – provide list of resources used, how groups were evenly matched, etc. Provide example of procedure with a target word (transport) in the appendix Provide measures	

					English academic vocabulary course constructed specifically for this experiment			▪ Fill in the blank with the rest of the word given the word part ▪ Ex: The Far East is one of the populated reg_____ of the world. Test-retest correlations – significant at .01		Bottom-up: Pre: 8.52 (SD 2.30) Post: 14.67 (SD 2.62) t=34.94, p=.000 Top-down Pre: 9.58 (SD 3.63) Post: 13.85 (SD 2.28) t=-10.77, p=.000			
Pauwels, P. (2012). Vocabulary materials and study strategies at advanced level. *Language Learning Journal, 40*(1), 47-63. doi:10.1080/09571736.2011.639899	2012	Mixed Quasi-experimental Quan ▪ ANOVA ▪ t-tests ▪ Spearman rank correlation Qual Used study logs for more robust data multi-group + ctrl	▪ 59 students ▪ 2nd year in college ▪ Studying to be translators/interpreters ▪ Upper-intermediate learners of English (studied English at least 4-5 years) ▪ L1 Dutch	EFL Flemish university	Participants divided into 6 mixed-ability groups based on vocab test scores Compiled list of 163 target words from AWL Developed 5 sets of study materials with range of dictionaries to provide different learner supports – each set added different supports and activities (materials description – see pp. 50-51 + table 2) ▪ Group 1 = Control (alphabetical list of words) ▪ Groups 2-6 = different set of materials for each with more or less involvement Table 7. Smallscreen load of study task. Materials / Need / Search / Evaluation Type 1 Organized list of vocabulary / 0 / 4 / — TDK 2 Vocabulary with example sentences / 1 / 3 / — Type 3 Vocabulary with Dutch glosses / 0 / 0 (+ —) / — Type 4 Type 3 + definitions / 0 / — / — Type 5 Type 4 + exercises / 1 / — / 0 Self-regulated, self-reported use of study materials (log)	Discrete	5 weeks	**Productive** (give explanation or translation + use in sentence); translate words in context of passage Vocabulary knowledge of target AWL words Pretest a. Knowledge scale – selected one of the following for all 163 words: 1. I recognize this word, 2. I know the meaning of this word, 3. I can use this word in a sentence b. Sample of 15 words – provided an explanation or translation and used correctly in a sentence Posttest 3 translation passages based on authentic materials, included 10-15 target words (that majority of participants had marked 'unknown') ▪ *Unclear procedure – did they just translate and use those target words?* Delayed posttest Shorter version of posttest -translated 1 passage ▪ Passage *was the SAME as one of on the previous post-test??* Added simple word-translation task (L1-L2) of 8 words so that students still had 23 target words to translate Brief questionnaire about study time between the 2 post-tests	Yes (slight) Electronic logs (specify amount of time spent studying, # of times vocab was repeated, further details re: specific activities or subsets)	ANOVA showed no sign difference between groups on pretest Group 4 (+ glosses) scored higher than all other groups on post-test, except Group 3 (+examples); Group 3 scored higher than Group 1 (control) ▪ *Results reporting is not precise…* Differences between groups disappeared between posttest and delayed posttest Amount of study time was not sign correlated with pretest or initial posttest scores ▪ ANOVA – no sign between-group diff on time studied ▪ Top scorers started studying early, lower scores earned by those who started studying w/in last week ▪ Group 6, which had the most involved study materials, had the fewest early studiers 18 highest scorers on post-test used the most strategies (average more than five) 4 students put in effort to study between posttest and delayed posttest – those had higher scores in the 2nd test Claim that materials which are "sufficiently informative" are the most effective (p. 58)	Productive 0.3013	**Strengths**: **Weaknesses:** Seems to heavily defend classroom-based approach (as opposed to a less authentic experiment) Doesn't describe post-tests fully – what was the actual test? Just translating the passages and seeing if they used the target words? Attrition – started with 77, only 59 did pretest, log, and posttest; only 49 did pretest, log, posttest, delayed posttest Did t-tests following an ANOVA – why didn't they just do post-hoc tests? Fuzzy/loose reporting of stats results – "t-tests comparing groups still show some level of significance for the difference between Group 1 and all other groups except for Group 3, while Group 3 significantly outperforms Group 1" (p. 56) – okay, but was stat sign? Did group 4 score stat sign. Higher than all other groups (except 3)? All based on self-report; all about self-directed study	Gives a nice argument in the Lit review about using a classroom-based approach that isn't as "artificial" as some experiments – ecological validity

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal, 29*(4), 679-693.	2012	Mixed Quan ANOVA T-tests Qual Attitude questionnaire exp x 2/ctrl	▪ 26 NNES ▪ Enrolled in freshman composition course ▪ L1 - 14 Chinese, 5 Korean, 3 Arabic, 2 Swedish, 1 Japanese, 1 Spanish ▪ Intermediate to advanced proficiency (college admission + TOEFL score for admission) ▪ Group A (glosses), n=9 ▪ Group B (dictionary), n=9 ▪ Group C (control), n=8)	▪ ESL ▪ Freshman composition course ▪ US public university, large	1 control group, 2 experimental groups Experimental group A – Concordance gloss – exposed to glosses that had 5 modified sentences (from Corpus of contemp. American English); had to read the text, click on highlighted words to see the sentences (showing different syntax and semantic variety), and study the words; website required students to study words before moving on to the next passage or test Experimental group B – Dictionary gloss – read same texts as group A in the online setting; exposed to dictionary definitions when they clicked on target, highlighted words Control group read same texts on the same website; same target words were highlighted (bold and underline) but had no links to a concordance or dictionary	Context	50 mins Given 50 minutes to read the 2 passages	Outcomes: **Receptive** vocab knowledge of AWL words in context (VLT + rating correctness of words in sentence), **productive** knowledge (cloze task), attitude toward glossing modalities Pre-test & post-test c. Vocab Levels Test (**breadth**) d. Judgment task - 30 sentences containing the target words that participants rated on 3 point scale (correct, I don't know, incorrect) e. 10 cloze sentences where students filled in correct word *Confusing procedure –From what I can tell, because the author is not explicit – the students took a test of vocab. Knowledge then read 2 passages that contained those words and took the test again.* Attitude questionnaire 5 point Likert (Strongly agree – strongly disagree)	Yes Web-based texts Online concordance with 5 sentences linked to target words Online dictionary definitions linked to target words	Both experimental groups scored higher on post-test than on pre-test; ANOVA showed concordance-based group scored sign. Higher than control group – F(2, 23) =3.74, p=0.04, but there was no sign difference between the 2 experimental groups d=0.821753 g=0.795803 VLT *Pretest* ▪ Concordance – M= 9.70 (SD 3.53) ▪ Dictionary – M = 8.88 (SD 3.18) ▪ Control – M = 11.38 (SD 2.67) ▪ Posttest ▪ Concordance – M = 13.5 (SD 1.58) ▪ Dictionary – M = 12.12 (SD 1.96) ▪ Control – 11.87 (SD 3.56) Judgment tasks (Receptive) – all groups improved on post-test with both experimental groups appearing to improve more than control group (based on raw data), but there was **not stat sign difference** between groups Cloze task (Pro) – concordance group improved more than dictionary group based on raw data (control group scored lower on posttest), but there were **no stat sign differences** between groups Attitude questionnaire further showed similarities between the 2 experimental groups	Receptive 0.7958 Productive n.s.	**Strengths**: Test validity – used 8 MA-TESOL students to take the test and comment on or circle anything confusing, misleading, etc.; test approved by 3 professors; 4th section of students enrolled in this same comp. class took the tests, t-tests to see if there were sign differences between the 3 test sections **Weaknesses**: Confusing procedure – didn't really do an intervention? Just used 2 different technologically-enhanced passages to work on target words…	Justification of proficiency level based on college admission (p. 684)
Rezaei, F. S., & Karbalaei, A., (2013). The effect of vocabulary strategy training among autonomous and non-autonomous learners in Iranian EFL context. *European Online Journal of Natural and Social Sciences, 2*(2), 35-49.	2013	Quan ANOVA Kolmogorov-Smirnov test for normality of data distribution (?) exp/ctrl	▪ 67 students at English language institutes ▪ 35 experimental, 32 control ▪ 20 years - average age ▪ Male and female ▪ Intermediate level proficiency: TOEFL test to determine proficiency – excluded participants with 1 SD above or below the mean	▪ EFL ▪ Talk Institute, Giti Institute, Arian Institute in Gorgan, Iran (3 separate institutes) ▪ 80 students from 4 in-tact classes originally selected then each group was randomly	Impact of learning vocab strategies on vocab knowledge Used 3 different vocabulary learning strategies – a different strategy for each set of 10 words: word parts, elaboration technique, and context clues All words were presented in the context of a larger passage, were then taught how to use the given strategy to figure out word meanings of	Both? Mostly in context but then discussed discretely…	Course of 1 semester?	**Receptive** (MCQ) with cloze tasks (fill in the blank with the best of the 4 choices offered) Pre- and post-tests were from Amy Olsen's "Active Vocabulary: General and Academic Vocab," 3rd ed; general vocab words were selected from SAT, GRE word lists & magazines like	No	Experimental group had a significantly higher mean gain score than the control group F=118.989, p=.000 Effect size: d=2.667975 Hedges g=2.637072	Receptive 2.63707	Writing is choppy, the organization is unclear, and results are confusingly presented	

				assigned as ctrl or exp	the 10 target words in that passage, and were quizzed on the words for that strategy/passage The articles does NOT state what the control group did instead of learning these strategies. Did they just take the pre and post-test?			Newsweek; Academic vocab came from textbooks and AWL Conducted pilot study on the 30 target words to make sure each group was equally difficult					
Tsai, S.-C. (2011). Integration of multimedia courseware into ESL instruction for technological purposes in higher technical education. *Educational Technology & Society, 15*(2), 50-61.	2011	Quan exp/ctrl	▪ 129 students in 3 different programs ▪ Minimum 8 years of English ▪ 3 groups of students taking "English for Technology" optional course	▪ EFL ▪ Taiwan ▪ Technical University ▪ Studying semiconductors in this course	Multimedia learning software – presenting information via narrated videos – text provided in English and in Chinese, colored as in karaoke style Compare to teacher-centered instruction without courseware addition For both: the teacher used textbook and PowerPoint files to note difficult content; students with access to courseware could then practice and study those topics	Both	7-week module embedded in semester long course	**Productive** Pre- and post-test + 2 questionnaires (learning effectiveness and attitude) Pre- and post-test = identical; 10 questions, students explained the meaning of a term and its process or purpose	Yes Learner-centered courseware integration Individual computers provided	All 3 groups saw significant gains in the post-test (but pre-test scores and t-test are not reported) No significant differences in post-test (independ. T-tests), either between groups with the courseware and those without (f2f) or across the 3 different programs (p. 54)  Their conclusion: student-centered courseware is just as effective as teacher-centered instruction. *But both groups still had teacher-centered time; does these results justify the time and energy spent making that online course?*	Receptive n.s. Productive n.s.	Writing – study is terribly organized, there's no "participants" section! It goes from method (which is really the procedure) to the results, divided by theme… Pre-test sensitization – they try to spin this to make it sounds like a benefit (p. 53, "Assessment") No SDs, nothing more than independent t-tests	A little difficult to note the role of technical vocabulary here – first, it says that the teacher presented the terminology then students studied the text and used the courseware? It seems that student background and interest had more to do with their post-test results than whether or not they used the online course. Focused on the questionnaire – satisfaction & attitude
Zhang, H., Song, W., & Burston, J. (2011). Reexamining the effectiveness of vocabulary learning via mobile phones. *Turkish Online Journal of Educational Technology,10*(3), 203-214.	2011	Quan Quasi-experimental t-tests Qual – low inference indicators through written reports exp/ctrl	▪ 62 college students ▪ L1 Chinese – Mandarin? ▪ Exp – 5 M, 26 F; Ctrl – 4 M, 26 F ▪ Ages 18-22	▪ EFL ▪ University in North China ▪ 2 In-tact classes	Studying same list of vocabulary delivered differently – SMS vs. paper material List of 130 words taken from the TOEFL vocab test – list covered phonetic pronunciation, part of speech, Chinese translation, sentence examples (p. 206) Experimental group – Class A – studied vocab using text messages; 5 vocab words delivered each day via SMS 2x/day – lunch at noon and dinner at 5:30pm Control group – Class B – studied vocab using paper materials; received printed list of all 130 vocab words face-to-face at the beginning of the study; had to self-regulate study	Discrete	26 days	**Receptive** (MCQ) Vocabulary knowledge Pretest – TOEFL vocab test 30 MCQ – vocab unlined in stated, chose 1 of 4 answers as the meaning of the target word Post-test – same as pretest Delayed posttest – same as other tests, delivered 5 weeks after initial posttest Experimental group asked to write report of their learning experience – 8 open-ended questions Comments on learning experience and how learning could be enhanced – discuss effectiveness (consider both side), advantages, and disadvantages of	Yes Text messages as method of delivery for target vocab	Pretest – Control group did better than experimental group, but it was not a stat sign difference Posttest – Experimental group scored significantly higher than control group t=2.45, p<.05 Delayed posttest – experimental group had higher retention rate, but score was not significantly diff from control group *Initially, this method showed sign gains over paper materials, and both groups improved from pretest to posttest; however, in delayed posttest, while both groups scored higher than their original pretests, the difference between the two groups disappeared	Receptive 0.61481	**Strengths**: Detailed procedure **Weaknesses**: Only did 1 test of vocabulary – receptive knowledge – so could only perform a t-test Only asked for written report from experimental group; didn't ask for study logs or hours of study time from either group No questionnaire about technology/mobile phone use before or after the intervention	They use the TOEFL vocab tests – on the TOEFL (ETS) website, they state that the exams "use 100 percent academic content to evaluate the English-language proficiency of nonnative English speakers, giving you confidence about your students' ability in a real-world academic setting" (https://www.ets.org/toefl_itp/content/) retrieved on June 28, 2016 ▪ So, the vocab here is academic vocab

								using mobile phones for vocab learning					