# A Meta-analysis on the Effects of Vocabulary Instruction for English Learners

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Ellina Z. Xiong

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Dr. Jennifer McComas, Advisor

Dr. Kristen McMaster, Co-Advisor

December 2020

**Acknowledgements**

When I began the doctoral program in the Special Education Program, I aimed to finish in three years since I already had an Ed.S. from the School Psychology Program. Now, five years later, two years more than I anticipated, I am completing my doctorate and will be the first person in my extended family to complete a doctoral degree.

It has been a long and beautiful journey. I am grateful for the skills that I have gained and the relationships developed along the way. Many people have helped me in becoming a stronger researcher and a better human being. I wish to express my most sincere gratitude and appreciation to my advisor, Dr. Jennifer McComas. Without Dr. McComas' encouragement and confidence in me as a rising scholar, I may have never enrolled in the doctoral program in Special Education. I thank Dr. McComas for her friendship, mentorship and for seeing the scholar in students who come from nontraditional backgrounds.

I would also like to extend my gratitude to Ahmed Alghamdi, Dr. John Mouanoutoua, and Mohammed Almalki. I cannot thank them enough for helping me to critically reflect about my research and for their support as secondary coders. Also, thank you Kory Vue for always willing to engage in random or late-night conversations about statistics.

Lastly, I would like to thank my committee members. Thank you for putting in hours of support to ensure that I, not only engage in rigorous research, but also am successful.

## Dedication

This dissertation is dedicated to my family. I am especially grateful to my love of 19 years, my husband, Chakong Thao. You are my rock and I thank you for the unwavering support. We made it!

## Abstract

A meta-analysis of group studies and single-case design studies was conducted to examine the effectiveness of vocabulary instruction on vocabulary learning and reading comprehension for English Learners. Overall estimates indicate that vocabulary instruction promoted vocabulary learning and reading comprehension. The mean effect for vocabulary learning was $g$= 0.40 ($CI_{95}$= 0.26-0.54, $p$< .001), a small to moderate effect. The mean effect for reading comprehension was $g$= 0.26 ($CI_{95}$= 0.07-0.46, $p$= .01). Meta-regression was used to conduct moderator analyses, which indicated that differential effects were associated with methodological rigor, instructional programming, and outcome assessments at a statistically significantly level. Findings suggest that comprehensive interventions tend to produce larger effects, but that interventions do not require significant duration, frequency and intensity to produce positive effects. Direction for future research is suggested based on findings from moderator analyses.

# Table of Contents

**List of Tables**

**List of Figures**

# Chapter 1

## Introduction

The U.S. student population is becoming increasingly diverse, and this trend is unlikely to change in the coming decades. One student subpopulation that has seen significant growth is English Learners (ELs). ELs typically are students who have limited English proficiency, speak a language other than English at home, or are learning English as a second language (Collier, 1992; Slavin & Cheung, 2005). ELs represented 10% of the public school student population in 2017 (Zhang et al., 2020) and are projected to be the largest student subpopulation by 2025, representing 25% of all school-aged students (Cheung & Slavin, 2005; Fry, 2007).

ELs include students from a wide array of culturally and linguistically diverse backgrounds (National Clearinghouse for English Language Acquisition [NCELA], n.d.), and come from varying socio-economic backgrounds (Lindholm-Leary & Hernández, 2011). Students speak a multitude of languages with Spanish, Arabic and Chinese topping the list (McFarland et al., 2018; NCELA, n.d.). The diversity in home language and ethnic composition accentuates the heterogeneity of this group.

Unfortunately, data suggest that schools are failing to meet the unique needs of these students. ELs are confronted with the challenge of learning English while also acquiring content knowledge, and data from the National Assessment of Educational Progress (NAEP) highlights this challenge. Over the last 10 years, less than 10% of ELs have demonstrated proficiency on the NAEP reading assessment and have shown little to no growth over time (National Center for Education Statistics [NCES] & NAEP, 2017).

In 2019, results of the most recent administration of the NAEP reading assessment indicated only 10% of EL fourth graders and 4% of EL eighth graders met proficiency standards. The low proficiency rate is a significant contrast to that of non-EL students. Thirty-nine percent of fourth-grade and 36% of eighth-grade non-ELs met proficiency, and have shown nearly five percentage-point gains in proficiency rates since 2007. Consistently low proficiency rates and stagnant growth over the last decade for ELs on the NAEP suggest that schools are struggling with providing effective reading instruction for EL students or sufficient strategies to support academic achievement. Little progress has been made in closing the achievement gap between ELs and their non-EL peers (National Center for Education Statistics [NCES] & NAEP, 2017), and many ELs will continue to struggle throughout their school experience.

Research on reading instruction is extensive; however, much of the research has focused on monolingual English speakers. Experimental research focused on ELs is limited, leaving educators with little guidance on evidence-based strategies to promote the reading development of ELs. Despite limited evidence, some scholars advocate that supporting ELs' vocabulary growth will facilitate their reading achievement (Calderon et al., 2005; Edmonds et al., 2009; Lesaux, Crosson er al., 2010), which would likely promote their academic success.

**Recommended Vocabulary Instruction for English Learners**

Although ELs have always been a part of the U.S. student population, there continues to be a paucity of research available (Alber et al., 2009; August & Shanahan, 2006). To provide educators greater guidance on teaching academic content and literacy for ELs, the Institute of Education Sciences tasked a panel (referred to as IES Panel

hereafter) of scholars to compose a practice guide using existing research and expert knowledge (Baker et al., 2014). The panel found 15 studies that met its causal validity standards and put forth four recommendations. Recommendations one and two of the practice guide relate directly to developing ELs' vocabulary skills. Due to the scope of the current review, I expand upon Recommendations 1 and 2 in the following sections. Readers may refer to the IES practice guide for more information on recommendations 3 and 4 (c.f. Baker et al, 2014).

***IES Panel Recommendation 1***

The first recommendation was supported by six studies and rated as having a strong evidence base. The IES Panel recommended that academic words be taught across several days by using a variety of activities. The Panel emphasized teaching vocabulary depth by focusing on a limited number of words and integrating explicit instruction on word learning strategies. As such, explicit instruction must include clear and transparent explanations of concepts, strategies, and rules, frequent modeling provided by the teacher, and multiple guided and independent opportunities for students to apply their learning (NPR & NICHD, 2000).

The practice guide promoted teaching word learning strategies such as using context clues that encourage students to examine information surrounding the text to determine an unknown word's meaning, analyzing word parts which helps students to learn words by breaking down root words and affixes, and using cognates which support ELs with identifying similarities in words between English and a target language (e.g., home language, Latin). Words targeted for instruction should include words that are central to understanding the content (e.g., high utility words), appear across multiple

content areas, have multiple meanings (i.e., polysemy), and/or possess cross-language connections (e.g., cognates). Vocabulary words should be embedded in informational texts that are engaging and relevant to the content or unit. Learning should be facilitated by incorporating activities that elicit speaking, reading, writing, and listening, and provide opportunities to use and be exposed to target words in different contexts. In addition, scaffolding techniques such as graphic organizers, definitions, and providing examples and nonexamples of word meanings are recommended.

***IES Panel Recommendation 2***

The second recommendation was supported by five studies and classified as having a strong evidence base. The IES Panel suggested that oral language and writing instruction must play a role when teaching content areas such as science and social studies. The panel recommended that when developing vocabulary in content areas, words specific to a content area and/or general academic words foundational for comprehension be taught simultaneously with content-specific materials. Teaching words during content area instruction supports students in distinguishing how word meanings change depending on the context. The panel recommended that teachers use instructional tools such as videos, visuals and graphic organizers to support students with developing background knowledge and promoting active engagement.

The IES Panel provided a much-needed resource for schools and teachers who support ELs. However, the panel specified that recommendations were targeted at the elementary and middle grades since high school students' instructional needs differ substantially from younger students. Moreover, given the limited number of studies identified ($n$= 15) and the panel's limited focus in teasing out strategies that promote

vocabulary learning from those that promote overall language development and reading comprehension, questions remain as to whether the recommended strategies effectively promote vocabulary learning. Thus, there remains little clarity in how best to promote vocabulary learning for ELs and which strategies are most effective.

**Limitations of Existing Syntheses**

The most comprehensive synthesis on reading instruction for ELs was conducted in 2006 by the National Literacy Panel on Language-Minority Children and Youth (NLP; August & Shanahan, 2006). The NLP's international search for relevant vocabulary studies found only three studies that used experimental or quasi-experimental designs (Shanahan & Beck, 2006). This was a dramatically low number compared to the 47 studies identified by the National Reading Panel (NRP), which only included studies that focused on non-EL students (NRP & National Institute of Child Health and Human Development [NICHD], 2000). Similar to the NRP, the NLP did not conduct a meta-analysis because there were too few studies, with few similarities across studies, to appropriately conduct such an analysis.

A recent review on reading instruction for ELs was conducted by Snyder, Witmer and Schmitt (2017). Snyder and colleagues reviewed experimental and quasi-experimental studies conducted between 2003 and 2015 with demonstrated large effect sizes (i.e., $d$ >0.8, $\eta^2$>0.14-0.26) to understand the characteristics of reading interventions that produced such large effect sizes. A total of 10 articles met the inclusion criteria; however, only four studies specifically focused on vocabulary interventions. Snyder and colleagues relied only on descriptive analysis and could not identify or discuss specific

intervention components that contributed to the large effects that were observed in each study.

Both reviews mentioned above provide valuable insight into interventions that promote vocabulary development for ELs. However, the limited number of studies did not allow for more sophisticated analyses to understand the contribution of specific intervention components on vocabulary development. Additionally, the multicomponent and multifaceted design of vocabulary programs make it difficult to attribute positive outcomes of an intervention to any one specific strategy without sophisticated analytic techniques.

**Purpose of the Study**

The purpose of this study was to examine the relation between vocabulary instruction on vocabulary learning for ELs. Moreover, since reading comprehension is the fundamental objective of reading instruction and it is well known that comprehension and vocabulary are related (NRP & NICHD, 2000), it is important to examine the relation between vocabulary instruction and reading comprehension for ELs. To achieve the purpose of this review, I conducted a meta-analysis of the literature.

Meta-analyses are critical tools in the social sciences that allow for synthesizing effects across studies, and evaluating the magnitude and variability of those effects (Cooper et al., 2009; Gurevitch et al., 2018). The advancing technologies and methodologies of meta-analyses have enhanced researchers' abilities to generalize findings beyond the studies included in a meta-analysis (Cooper et al., 2009; Shadish et al., 2015), identify evidence-based practices (Goodwin & Ahn, 2010; Rolstad, 2005) and highlight research gaps (Chaffee et al., 2017; Reschly et al., 2009).

The major advantage that meta-analyses have over other synthesis methods is that meta-analyses provide a quantitative synthesis of the research. As a result, findings are calibrated to a common scale. By transforming findings from different studies to a common scale, the quantitative method allows researchers to meaningfully draw conclusions and recommendations across studies even though each study was conducted by different research teams, in different regions, or on different populations (Cooper et al., 2009; Lipsey, 2003).

**Significance of the Study**

It has been more than 10 years since the last comprehensive review on vocabulary interventions for ELs (August & Shanahan, 2006), yet there remains a lack of clarity on which instructional practices consistently promote vocabulary learning and reading comprehension for ELs. Furthermore, few studies have examined and synthesized the effects of vocabulary interventions on EL vocabulary learning, or the relation between vocabulary interventions and reading comprehension for ELs. In essence, these relations have not been comprehensively examined to date.

If school systems are to improve the academic success and outcomes of ELs, it is important to understand and identify effective strategies that promote learning. Conducting a meta-analysis of current research will shed light on practices that enhance the development of vocabulary skills that can facilitate reading comprehension. Until the field gains a deeper and better understanding of vocabulary interventions on reading development, recommending practices with limited supporting evidence may reinforce inadequate instructional practices that do not address the needs of ELs and further exacerbate the achievement gap between ELs and their non-EL peers.

## Research Questions

The current review adds to the literature by examining vocabulary instruction and its effect on vocabulary learning for ELs, and the relation between vocabulary instruction and reading comprehension. The research questions driving this review are:

1. What is the overall quality of the research?
2. To what extent are vocabulary programs and interventions effective in increasing vocabulary learning for English Learners?
    2a. What is the average or overall effect of vocabulary instruction on vocabulary learning for English Learners?
    2b. To what extent do methodological characteristics (e.g., participant demographics, study design characteristics, and measurement methods) moderate study outcomes?
3. To what extent are vocabulary programs and interventions effective in increasing reading comprehension for English Learners?

## Definitions of Key Terms

The following definitions will be adopted for the current review:

**English Learners** (ELs) are broadly conceptualized as students learning English as a second language, those who speak a language other than English at home, students who have limited proficiency in English, or those identified as language minority students (Collier, 1992; Slavin & Cheung, 2005).

**Explicit instruction** pertaining to vocabulary learning consists of providing definitions to words or other features of the word to be learned (NRP & NICHD, 2000). Explicit instruction consists of employing teaching practices that make rules to

deciphering word meanings (e.g., word parts) transparent or providing conspicuous clues and demonstrating to students how to use tools and strategies when learning word meanings and concepts (NRP & NICHD, 2000).

**Indirect instruction** is when individuals infer definitions or the meaning of words (NRP & NICHD, 2000) through natural exposure. Indirect instruction may occur in the context of students being exposed to a wide array of reading materials, encouraging students to engage in a wide of array of independent reading and word exploration, or reading aloud to students (Stahl & Nagy, 2006).

**Productive/expressive vocabulary** is the "set of words that an individual can use when writing or speaking" (Kamil & Hiebert, 2005, p. 2).

**Receptive vocabulary** is the "set of words for which an individual can assign meanings when listening or reading" (Kamil & Hiebert, 2005, p. 2). Namely, these are words that are understood through recognition in print or when listening to others talk.

**Vocabulary instruction/intervention** is broadly defined and consists of the teaching of word knowledge, meaning-making, and/or word learning (Beck et al., 2013; Kamil & Hiebert, 2005; Stahl & Nagy, 2006). In essence, instruction focuses on assigning meaning to words by teaching specific words or strategies to acquire new words. Incidental word learning is also part of vocabulary programs and may include wide reading, shared book reading, silent reading and reading aloud (Fukkink & Glopper, 1998).

# Chapter 2

## Literature Review

In 2009, Albers and colleagues conducted a widespread systematic review of journal coverage concerning the academic, mental health and social-emotional needs of ELs. The review examined 16 prominent journals in education (e.g., *School Psychology Review, Exceptional Children*) and found between 1995 and 2005 that nearly 6,000 articles were published about ELs. However, only 3% of sampled articles (*n*= 177) had a primary focus on ELs, and few publications offered sufficient evidence-based practices for educators when supporting ELs in schools (Albers et al., 2009). The authors noted that the lack of research available to inform practice left many educators unprepared to address the challenges ELs encounter, and thus, likely perpetuated EL students' difficulties in schools.

The above study (Albers et al., 2009) illustrates the value of systematic literature reviews. Researchers applied strict methodological rules related to a specific question, and presented data on the breadth of the research while identifying gaps in the research and implications for practice. Although systematic reviews do not lend themselves to quantitative syntheses like meta-analyses do, they offer meaningful contributions to the field in that they help to establish a foundation of the literature, allow researchers to evaluate the strengths and weaknesses of the current state of research, and set the stage for future research (Boote & Beile, 2005; Hart, 2018). Therefore, to gain better insight into EL research pertaining to vocabulary instruction and its effect on vocabulary development, it is important to review current literature on the topic, in addition to reviewing previous systematic reviews.

In this chapter, the first section centers on the role of vocabulary in reading comprehension and theories describing their relation. The second section provides an overview of previous syntheses conducted on vocabulary instruction for ELs. The last section presents the research contributions of the current study.

**The Relation Between Vocabulary and Reading Comprehension**

Vocabulary instruction is a foundational component of effective instruction in reading development, and vocabulary is identified as one of five critical pillars of reading (NRP & NICHD, 2000). Vocabulary may be a critical component of reading because of its correlational relation with reading comprehension (Beck et al., 1982; Joshi, 2005; Kieffer & Box, 2013). Several theories have been offered to explain the relation between vocabulary and reading comprehension. The first theory, the *instrumentalist hypothesis*, proposes that the more words a person knows, the greater their ability to comprehend text (Anderson & Freebody, 1979; Nagy, 2005). Hence, expanding a student's vocabulary should directly improve their comprehension. A second theory, the *aptitude hypothesis*, proposes that the relation between vocabulary and reading comprehension is moderated by verbal ability (Anderson & Freebody, 1979; Nagy, 2005). Students who have higher verbal skills should learn more words, enabling them to comprehend text.

Anderson and Freebody (1979) also proposed a third hypothesis, the *knowledge hypothesis*. The knowledge hypothesis suggests that the extent of a student's background knowledge enables reading comprehension (Anderson & Freebody, 1979; Nagy, 2005). Knowing the meaning of words is secondary to possessing knowledge about the content or topic, which is the essential factor in facilitating comprehension. Anderson and Freebody (1979) describe comprehension as a reflection of one's exposure to the content

because a schema has already been developed for understanding. Thus, comprehension has less to do with knowing a high volume of words, and instead may have more to do with extensive knowledge of the concepts for which words are used to label them. The last theory proposed is the *access hypothesis*, which suggests that comprehension is enabled when the student possesses fluency in both breadth and depth of word knowledge (Nagy, 2005). Hence, students must acquire a sufficient understanding of words and be able to extract the appropriate meaning when needed in order to comprehend text (Nagy, 2005).

Another theory concerning vocabulary and its contribution to reading comprehension is the Simple View of Reading (SVR), which was first discussed by Gough and Tunmer (1986). In the SVR, reading comprehension results from the multiplicative relation between decoding and linguistic comprehension [R= D x LC]. Decoding is broadly defined and is a function of efficient word recognition. Linguistic comprehension, also often referred to as language comprehension, is broadly defined as the ability to process and interpret meaning through speaking and listening. Vocabulary is said to live within the linguistic comprehension component (Gough & Tunmer, 1986). The multiplicative relation, [D x LC], emphasizes that decoding and linguistic comprehension are interdependent and necessary for reading comprehension. The equation underscores the notion that a weakness in decoding or linguistic comprehension compromises reading comprehension.

Although Gough and Tunmer titled their theory as simple, it is well understood that reading involves complex processes. Scarborough (2001) expanded upon the SVR to call attention to the complex processes involved in reading and the distinct strands that

comprise decoding and linguistic/language comprehension. Scarborough's illustration (see Figure 1) of the different processes is also known as the Reading Rope. In the Reading Rope, phonological awareness (e.g., phonemes), decoding (e.g., alphabetic knowledge), and sight recognition are three distinct strands woven together to form a decoding braid. Language comprehension consists of five core strands, background knowledge (e.g., facts), vocabulary (e.g., breadth), language structures (e.g., syntax), verbal reasoning (e.g., inference), and literacy knowledge (e.g., print concepts) that are woven together to form a second braid. It is the coordinated lacing of the language comprehension braid and decoding braid that result in skilled reading or reading comprehension.

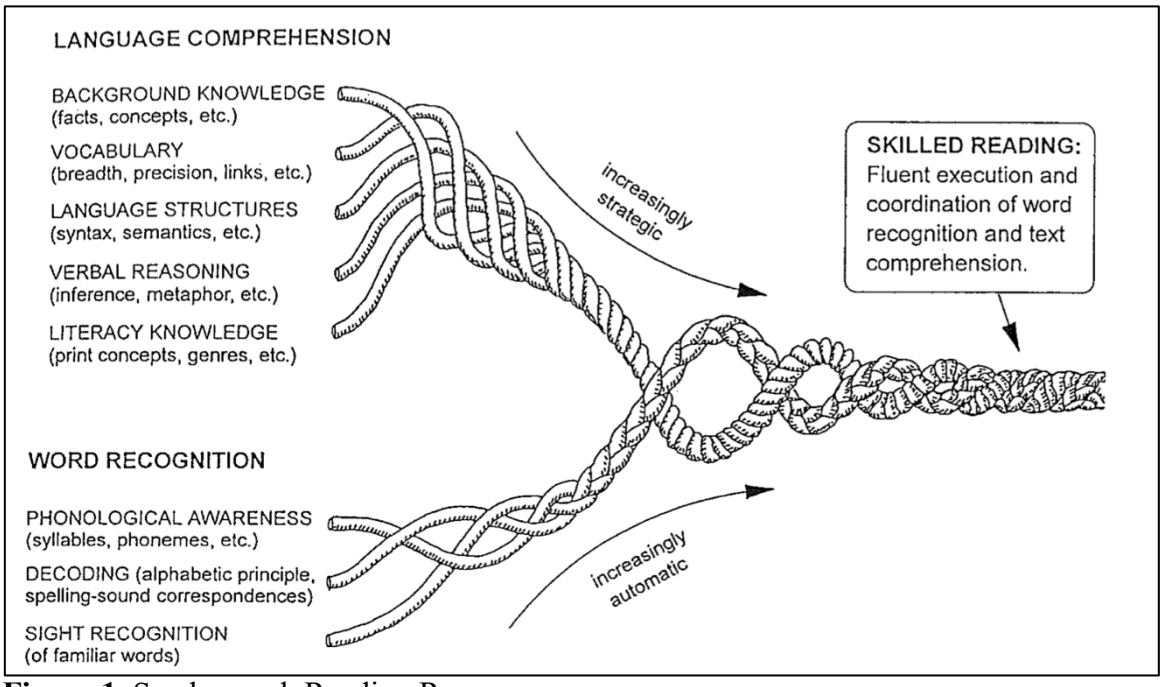



**Figure 1.** Scarborough Reading Rope.

Note. Scarborough, H.S. (2001). Connecting early language and literacy to later reding (dis)abilities: Evidence, theory, and practice. In Neuman, S., & Dickinson, D. (Eds). *Handbook of early literacy research* (pp. 97-110). New York: Guilford Press.

In this representation of the SVR, vocabulary is explicitly identified as a component within language comprehension that independently contributes to the facilitation of reading comprehension. However, ongoing discourse in the field debates the contribution that vocabulary makes to reading comprehension as an independent component of language comprehension. The evidence is mixed as to whether vocabulary is a unique component of language comprehension (Cho et al, 2019; Kim, 2017) or a unitary construct of listening comprehension (Braze et al., 2016; Tunmer & Chapman, 2012). Nevertheless, many continue to argue that vocabulary is critical to reading development and is a necessary component of reading instruction (NRP & NICHD, 2000).

Each of the above frameworks describe the relation between vocabulary and reading comprehension, but there is insufficient evidence to reliably support any one theory. There is uncertainty as to whether the relation between vocabulary and comprehension is causal (Butler et al., 2010; Joshi, 2005). However, the evidence is clear that vocabulary is critical for reading comprehension and achievement, and has a close relation with reading comprehension and achievement (Cain & Oakhill, 2011; Joshi, 2005; NRP & NICHD, 2000). Considering the close relation between vocabulary and reading comprehension, this signifies that vocabulary comprises an important aspect of reading development and requires further research to understand its specific influence on reading development.

## Syntheses and Literature Reviews of Vocabulary Instruction for English Learners

### *National Reading Panel Report*

In one of the most comprehensive reviews of reading research, the NRP (NRP & NICHD, 2000) found 47 experimental and quasi-experimental studies on vocabulary instruction. Findings indicated that vocabulary instruction should be integrated as part of reading instruction, and instruction focused on individual words and word learning strategies should actively engage students in the learning process. Furthermore, the NRP suggested that effective strategies for fostering vocabulary growth include: using a combination of explicit and implicit instruction; providing multiple and repeated opportunities to learn, see and use words; and employing computer-assisted technology.

One limitation of the NRP report was that studies focused on ELs were excluded (NRP & NICHD, 2000). Although the report was comprehensive, the exclusion of studies focused on ELs causes concerns when generalizing findings and trends to EL students. Caution must be exercised when generalizing strategies that have been effective for English-only (EO) students to ELs. It is possible that learning to read is different for ELs compared to monolingual speakers. Second language development theories indicate that ELs progress through a series of stages when acquiring a new language. Students move from a stage of absorbing language by listening and observing others talk to a final stage of being able to fluently discuss complex and sophisticated topics with others (Krashen & Terrell, 1995). ELs can spend from just one year to more than five years within each stage of development (IPEK, 2009). Monolingual speakers appear to move through similar stages of language acquisition, but tend to move through the stages at a quicker pace because they have been exposed to the language much of their lives starting at a

very young age (IPEK, 2009). Therefore, reading development in a second language may be similar to language acquisition in that ELs may require more time at each reading stage, which also may require more intensive instruction that differs from EO students. For these reasons, caution must be exercised, and further evidence gathered before suggesting that strategies effective for EO students are also effective for EL students.

***National Literacy Panel Report***

In response to findings from the NRP, a subsequent panel, the National Literacy Panel on Language-Minority Children and Youth (NLP), was formed to provide greater insight on research concerning ELs (August & Shanahan, 2006). The NLP conducted a national and international search for literacy research focused on ELs, producing, perhaps, the first and most comprehensive review for this special population of students. In a search for experimental and quasi-experimental studies on vocabulary instruction, the NLP found only three experimental and quasi-experimental studies that met inclusion criteria (Shanahan & Beck, 2006). This was a dramatically low number compared to the 47 studies identified by the NRP (NRP & NICHD, 2000).

Findings from the NLP (August & Shanahan, 2006) indicated that multiple exposures, in-depth word learning, and learning words in various contexts were effective in supporting vocabulary growth for ELs. These strategies were consistent with the NRP's findings for monolingual English speakers. Moreover, the NLP's review found that incorporating bilingual instruction when introducing new words supported vocabulary growth. Also similar to findings on EO students, the integration of vocabulary strategies as part of reading instruction resulted in positive effects on reading comprehension.

Results from the NRP and NLP seem to suggest that there may be overlapping vocabulary strategies that are effective for EO students and ELs. Unfortunately, given only three experimental studies identified by the NLP, it is difficult to make the broad generalization that effective strategies for EO students are also effective with EL students.

***Snyder and Colleagues' Review***

A recent review on reading instruction and ELs was conducted by Snyder, Witmer and Schmitt (2017). Snyder and colleagues reviewed experimental and quasi-experimental studies conducted between 2003 and 2015 with demonstrated large effect sizes ($d$ >0.8, $\eta^2$>0.14-0.26) so as to understand the characteristics of reading interventions that produce such large effect sizes. A total of 10 articles met the inclusion criteria. All 10 studies consisted of interventions examining at least one of the five pillars of reading identified by the NRP (i.e., phonemic awareness, phonics, fluency, vocabulary, and comprehension). While vocabulary was the most frequently cited reading component integrated as part of reading interventions, only four out of the 10 studies specifically examined vocabulary instruction.

Corroborating findings of past reviews, Snyder and colleagues found that reading comprehension was fostered by developing vocabulary skills, along with other reading skills such as fluency and phonics. Worth noting is that vocabulary growth was largest and prominent when vocabulary was the primary focus of the intervention. Although multicomponent interventions that combine fluency and phonics instruction can produce large gains in different areas of reading, the same cannot be said about vocabulary.

Snyder and colleagues concluded that when the aim of an intervention is to increase vocabulary skills, this is best achieved with vocabulary instruction being the focal point.

### *Xiong's Review*

As noted previously, systematic reviews support establishing a foundation of the literature, evaluating the state of research and setting the stage for future research (Boote & Beile, 2005; Hart, 2018). I conducted a systematic literature review to evaluate the landscape of current research on vocabulary instruction for ELs (Xiong, 2018)[*]. The search was conducted using three large databases (PsycInfo, Academic Search Premier, and Education Resources Information Center [ERIC]), screening only for peer-reviewed, experimental and quasi-experimental group and single-case design (SCD) studies conducted in the United States. The age range for studies was restricted to kindergarten to high school students. Studies targeting pre-kindergarten students were excluded from review because the primary focus of educational programs for early childhood is on oral language development and early literacy skills (Wilson, Dickinson, & Rowe, 2012), which differ in quality to that of vocabulary instruction programs delivered to elementary and secondary students. As such, studies focused on phonological awareness (Anthony et al., 2009), phonics (Vadasy & Sanders, 2011), speech (Kan & Sadagopan, 2015) and oral

[*] Xiong (unpublished manuscript) can be made available upon request

language development (Blom & Paradis, 2013) were beyond the scope of the review and excluded. The search resulted in 21 studies (*n*=14 group designs, *n*=7 SCDs).

Findings indicated that across these studies, over half of the studies (*n*=13) focused on elementary-age students. The majority of studies enrolled ELs from Spanish-speaking populations. Similar to reports from the NRP (NRP & NICHD, 2000) and NLP (August & Shanahan, 2006), there was immense variability across intervention programs with the majority of studies using researcher-developed assessments as opposed to commercialized assessments.

Research teams appeared to closely align intervention designs with recommendations made by the NRP (NRP & NICHD, 2000) and later echoed by the NLP (August & Shanahan, 2006) and IES panel (Baker et al., 2014). No one study used an isolated strategy to teach vocabulary, and most studies incorporated a combination of explicit and implicit instructional strategies (Guardino et al., 2014; Silverman et al., 2017; Spycher, 2009). Over half of the studies incorporated instructional strategies that consisted of explicitly defining vocabulary terms (Cannon et al., 2010; Silverman et al., 2017) and actively using vocabulary words in sentences (Kim & Linan-Thompson, 2013; Lesaux et al., 2010; Spycher, 2009), suggesting general consensus that the teaching of definitions and active use of vocabulary words were critical elements of intervention programs.

Regarding methods of word selection, research teams did not appear to use a common method in selecting words for instruction. Methods used to select words for instruction ranged from using Beck and colleagues' (Beck, McKeown & Kucan, 2013) recommendations in selecting high utility words (Cena et al., 2013; Lesaux et al., 2010;

Proctor et al., 2011) to words identified by teachers (Green et al., 2015). Despite the fact that teaching high-utility words was repeatedly recommended as crucial for vocabulary programs (Baker et al., 2014; Hairrell, Rupley, & Simmons, 2011; Marulis & Neuman, 2010), selecting high-utility words for instruction was not indicative of significant or large gains for ELs in the sampled studies (Carlo et al., 2004). Inconsistent effects were observed across studies, suggesting that for the present time, there is limited merit in prioritizing instruction on high-utility words. As such, it is important to continue examining instructional methods on how best to teach high-utility words if this remains the recommendation for promoting vocabulary growth among ELs.

Descriptive analyses of the 21 studies found no clear pattern to understand why instructional practices appeared effective in some studies and not others, or why some measures appeared more sensitive to vocabulary effects. Furthermore, some studies analyzed intervention effects by comparing intervention EL students to EO control (Proctor et al, 2011) or EO intervention students (Lesaux et al., 2010; Silverman et al., 2017), which made it difficult to consistently compare effects across studies or understand the gains intervention EL students made compared to their like-peers (i.e., control ELs).

The drawback of systematic reviews is that such methods cannot quantitatively examine the attributes that contribute to intervention effects. More sophisticated analytic approaches such as meta-analyses are needed in order to compare effects across studies, and identify attributes (e.g., intervention program, research design) responsible for positive outcomes observed in certain studies, or identify interaction effects and moderating variables that would explain the gains observed (Cooper et al., 2009).

None of the reviews discussed above have been able to quantitatively examine the degree to which vocabulary instruction leads to vocabulary learning for ELs, the attributes responsible for positive gains observed, or the relation between vocabulary learning and reading comprehension. It is important to answer these questions because resources in schools are becoming scarcer as the needs of students grow, and educators should not waste precious resources on practices that are inefficient and ineffective.

**Contributions of the Current Study**

The current study makes several contributions to the field. First, meta-analytic techniques provide a quantitative synthesis of the literature, which has not been done for previous reviews on vocabulary instruction for ELs. Sophisticated meta-analytic techniques help to evaluate the degree to which vocabulary instruction improves vocabulary learning of ELs, and examine the different variables that may be responsible for the positive effects. Second, past reviews have not traditionally included SCDs when evaluating the literature. SCDs are also experimental designs that lead to causal inferences and contribute unique knowledge about effective intervention practices (Horner et al., 2005). SCDs are included in the current study, and are essential to providing a comprehensive understanding of the effects of vocabulary instruction on vocabulary learning for ELs. Overall, examining these relations can shed light on effective practices and provide insight on the specific vocabulary components associated with improved vocabulary learning and comprehension skills.

# Chapter 3

# Method

## Literature Retrieval

### *Searching Digital Research Databases*

Literature retrieval and searches were conducted using the following large databases to extract published studies and unpublished manuscripts: Academic Search Premier (ASP), ProQuest Digital Dissertations, PsycInfo, Education Resources Information Center (ERIC), and Education Source.

### *Internet Search*

The websource Open Science PrePrints (OSF) was used to search for published and unpublished manuscripts related to vocabulary learning.

### *Reference List*

An ancestral search examining citations from previous EL vocabulary literature reviews (August & Shanahan, 2006; Baker et al, 2014; Snyder et al., 2017) were completed and these articles were included as part of the selection process (described below). Additionally, at the conclusion of Phase 3, the full-text screening, the reference lists of all eligible articles were reviewed in search of possible articles for inclusion.

Studies identified from a previous systematic review conducted by this author (Xiong, 2018) were included in the search process.

### *Contacting Researchers*

Researchers in the list below were contacted via email to inquire and retrieve unpublished manuscripts related to vocabulary interventions. Researchers were provided four weeks to respond to the inquiry-email with unpublished manuscripts. Researchers

selected for contacting (see Table 1) were the principal investigator of the NLP and authors of the NLP's review on vocabulary research (Shanahan & Beck, 2006). Furthermore, authors of the IES panel's practice guide on supporting literacy development for ELs (Baker et al., 2014) and authors of existing EL vocabulary literature review (Snyder et al., 2017) were included on the list. To promote comprehensive outreach, researchers who authored two or more studies from the sample of studies from a previous systematic literature review were also included ($n$ = 21; Xiong, 2018)[*].

**Table 1**

*List of Authors Contacted for Unpublished Manuscripts*

| Author's Name | Sources Used to Identify Authors |
|---|---|
| August, Diane | NLP |
| Baker, Scott | IES Panel |
| Beck, Isabel | NLP |
| Bravo, Marco | Authored two or more studies |
| Cannon, Joanna | Authored two or more studies |
| Cervetti, Gina | Authored two or more studies |
| Francis, David | Authored two or more studies |
| Geva, Esther | IES Panel |
| Kelley, Joan | Authored two or more studies |

[*] The unpublished manuscript can be made available upon request by contacting Ellina Xiong.

| | |
|---|---|
| Kieffer, Michael | IES Panel |
| Lesaux, Nonie | IES Panel |
| Linan-Thompson, Sylvia | IES Panel |
| Morris, Joan | IES Panel |
| Proctor, C. Patrick | IES Panel |
| Russell, Randi | IES Panel |
| Shanahan, Timothy | NLP |
| Snow, Catherine | Authored two or more studies |
| Snyder, Elizabet | Past literature review |

### *Search Terms*

Search strings consisted of variations of English Learner and vocabulary learning terms. Although it would be ideal to use exact search strings for all databases in order to maintain consistency in search procedures, the varying interfaces and search capabilities of the different databases did not allow for the use of exact search strings across databases. As such, search strings were adapted for each corresponding database or literature retrieval source. The search structure was informed by consulting social sciences librarian, Amy Riegelman. Below is an example of the search string that was used in the Open Science Framework Preprints (OSFPreprints) database. See Appendix A for detailed search strings for all other literature retrieval sources that were used for the current meta-analysis.

Source: OSFPreprints

("English Language Learner" or "English Learner" or "English as a second language" or "Limited English Proficiency" or "Language minority" or "Emergent

bilingual" or "Second Language Learner" or "Second language education" or "Second language acquisition" or Bilingual or Multilingual or "Linguistically diverse" or "Dual language" or "Dual Language Learner" or "Non-English speakers" or "English for speakers of other language") AND (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" or "vocabulary skills" or "vocabulary strategies" or "vocabulary building")

***Selection Process***

Studies identified for inclusion were screened through four phases. The screening process was implemented because a previous literature search of EL research (Xiong, 2018) revealed significant variability in research designs used to study vocabulary effects for ELs, and that group comparisons often compared ELs in treatment groups to EO peers in treatment or control conditions. The screening process promoted efficiency in excluding qualitative and correlational studies, and provided early flagging of studies that may require follow-up with authors to secure appropriate data for meta-analyzing. Each of these phases are described below.

a.) Phase 1 consisted of gathering studies from research databases, internet searches, ancestral searches, and researchers. Citations were imported into Zotero (Version 5.0.64), a citation manager program, and a Microsoft Excel (Version 16.23) spreadsheet for record keeping. Literature retrieval resulted in 1,203 studies. In this phase, using each study's full citation,

duplicates were removed (see Figure 2). This resulted in 1094 studies that were retained.

b.) Phase 2 consisted of reviewing manuscript titles and abstracts. Manuscripts were eliminated when they did not satisfy all screening criteria found on the Abstract Screening Checklist (see Appendix B). Any time a criterion could not be confirmed based on information in the title or abstract, the article was advanced to Phase 3 of the selection process. This resulted in 204 studies that were retained.

c.) Phase 3 consisted of a methods screening using the Methods Screening Checklist (see Appendix C) to verify that the inclusion criteria were addressed. Articles that satisfied all checklist components were advanced to Phase 4 of the selection process. Excluded studies, along with the corresponding reasons for exclusion were documented and reported. In this phase, unpublished studies (i.e., dissertations) were also reviewed for potential duplication as published studies that were extracted during the search process. When duplicates were identified, the published version was included in the meta-analysis, as this was considered the record of print (A. Riegelman, personal communication, February 14, 2019). During this phase, 25% of studies were randomly sampled and independently coded by secondary coders to ensure consistent screening of criteria established in Appendix C. IOA was 100%. This phase resulted in 68 articles that were retained.

d.) Phase 4 consisted of a reference list review. The reference lists of manuscripts that satisfied all inclusion criteria in Phase 3 were reviewed. When reviewing reference lists, all studies indicating reading or vocabulary interventions, or English Learners were identified and screened using the Methods Full-text Screening Checklist. This phase resulted in an additional two studies that would be advanced to full-text eligibility review. In total, 70 studies were advanced to full-text eligibility review.

### *Eligibility Criteria*

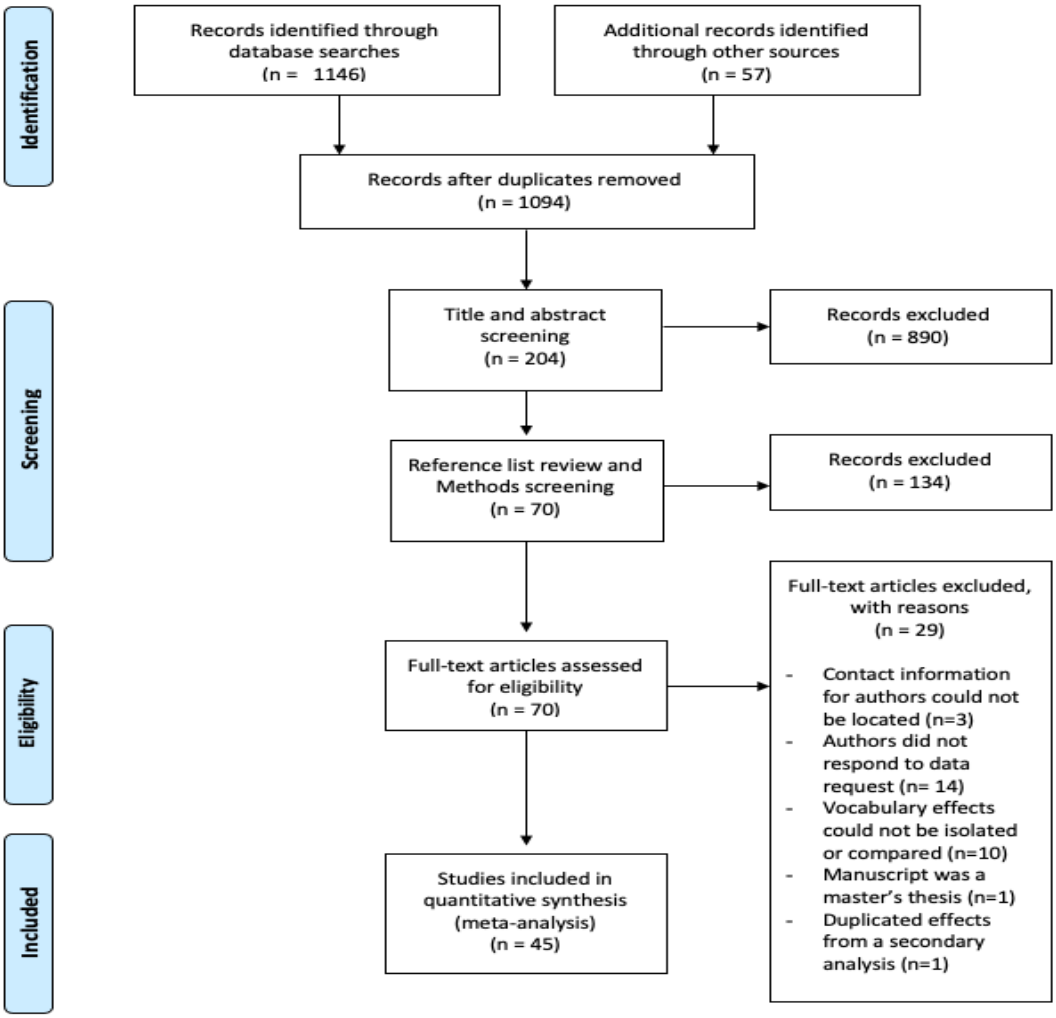
To reduce the potential for selection bias, a priori eligibility criteria were developed (Lipsey & Wilson, 2000; Wilson, 2009). The purpose of the eligibility criteria was to identify relevant studies to the topic of ELs and vocabulary learning, and limit data extraction to studies that employed experimental designs that permit causal inferences. The timeframe in which studies were made available or published was not restricted in order to maintain a comprehensive search of relevant literature. The following criteria were used to select relevant studies.

**Criterion 1**. Participants were English Learners (EL) in elementary or secondary schools (kindergarten–12th grade). ELs in the study were broadly conceptualized as students learning English as a second language, those who speak a language other than English at home, students who have limited proficiency in English, or those identified as language minority students. Studies that focused on English as a foreign language (EFL) were excluded because learning English is not intended for accessing the core curriculum as it is conceptualized in U.S. elementary and secondary schools (Anderson, 2004). In

reviewing past studies, it became clear that researchers were not always explicit on how ELs were defined or identified, therefore, the broad definition was used to be inclusive of studies focused on the EL experience.

**Figure 2**

*Flow Diagram of Selection Process*




**Criterion 2.** The study had to be conducted in the U.S., because standards shaping education across countries are highly variable (Husén, 1983; Stedman, 1997).

Focusing on only studies conducted in the U.S. should reduce the variability of educational standards in order to better support interpretation of effects. Moreover, studies were restricted to those reported in English.

**Criterion 3.** Vocabulary instruction was broadly described by authors as word knowledge, meaning-making, and/or word learning. This included both explicit (e.g., word learning strategies) and/implicit strategies (read aloud, silent reading, sustained reading). Vocabulary instruction had to be an independent variable that was manipulated, or a component of the independent variable that was manipulated. As such, vocabulary instruction could be a stand-alone intervention or integrated as part of a multi-component intervention.

**Criterion 4.** Studies included a control condition as an element of an experimental or quasi-experimental design. A control condition increases confidence that student outcomes are the result of the instructional strategy and not an extraneous variable. For group studies (GS), the study had to include an independent no-treatment, business-as-usual, or dose-equivalent active comparison group. For single-case designs (SCD), the study had to include a baseline/control condition before the implementation of the intervention, a comparison leg such as in multiple-baseline designs, or an alternating treatment. Therefore, studies had to employ experimental and empirical designs on vocabulary instruction to be considered. Qualitative and case studies were not included in the current review. Correlational design studies were also excluded because the objective of such studies is not focused on examining causal relations or inferences.

**Criterion 5.** Selected studies had to include at least one outcome variable that specifically measured the effect of vocabulary instruction, or presented data that allowed

vocabulary effects to be isolated. Therefore, studies had to include quantitative data on vocabulary measures to be considered. Studies that measured only comprehension effects were excluded.

**Criteria 6.** To be included in the meta-analysis, disaggregated data that isolated the effects of vocabulary outcomes or allowed data to be isolated to evaluate vocabulary effects must be reported for ELs for both control and treatment conditions.

When evaluating intervention effects for ELs, it is best to compare ELs in the intervention condition to ELs in the control condition (Rolstad et al., 2005; Slavin & Cheung, 2005). Such a comparison helps to reduce confounds associated with differing life experiences because of students' EL status (Rolstad et al., 2005; Slavin & Cheung, 2005). Therefore, group comparison studies must report descriptive and/or statistical data (e.g., *m*, *SD*, t-value) for ELs in the intervention and control condition. When studies satisfy all other criteria, but disaggregated data for ELs were unavailable to allow comparisons between intervention and control ELs, I contacted the first author or authors designated for correspondence for the disaggregated data. These authors were provided an Excel spreadsheet notating the requested data and were provided a three-week window for a response. When manuscripts did not include contact information for authors, I conducted a people/directory search via the author's institution/organization, Google or LinkedIn.

### *Exclusion Criteria*

The current review focused on outcomes for ELs who were provided reading vocabulary instruction or interventions. Studies focused on phonological awareness (Anthony et al., 2009), phonics (Vadasy & Sanders, 2011), speech (Kan & Sadagopan,

2015) and oral language development (Blom & Paradis, 2013) were beyond the scope of the current paper and excluded. Furthermore, studies focused on teacher professional development or teacher instructional performance without incorporating student outcomes were excluded (Rance-Roney, 2010). Studies that evaluated and validated assessment tools were also excluded (Goodwin et al., 2012).

Studies evaluating the effects of bilingual/EL/English as a Second Language (ESL) program models and studies that compared the differences between models were excluded. Evaluating the efficacy of bilingual/EL/ESL program models were beyond the scope of this paper. Readers interested in this topic may consult existing syntheses (Chueng & Slavin, 2005).

**Coding Procedures**

A total of 70 studies were eligible for full-text review and coding after Phase 4 of the screening process. All articles were double-coded. I was the primary coder, and secondary coding was completed by three individuals. The 70 studies were stratified by GS or SCD and randomly assigned to the three secondary coders (i.e., Set A, Set B, and Set C). The use of this method ensured that all secondary coders had an equal opportunity to code GS and SCD studies from a variety of journal outlets and search databases in order to reduce bias. Random assignment of articles to sets and secondary coders was conducted in R (R Core Team, 2013).

Coding procedures were broken into seven general areas: study information, regional demographics, participant demographics, study design characteristics, outcome measures, effect size information and methodological characteristics. Each of these areas

are briefly described below. Variables coded were used to support effect size computations, descriptive analyses, and meta-analysis.

A coding manual (see Appendix D) that included the variable name, response codes and operational definitions of each response option was developed to ensure systematic and consistent coding. The coding manual was developed based on previous work reviewing the literature on vocabulary interventions for ELs (Xiong, 2018). Findings in the previous review pertaining to highlights and concerns regarding measurement practices, participant backgrounds, and effect size calculations were incorporated as variables for the current review. The coding manual was finalized using an iterative process, which consisted of piloting the manual with secondary coders.

Piloting the manual was conducted using a random sampling of two GS and two SCD studies identified for full-text review, and coding the studies together with a secondary coder. All variables needing clarification were discussed until a consensus was reached and resulted in a refined operational definition for the variable when applicable. After the discussion, secondary coders and I recoded the four studies independently and repeated the process of discussion, refinement and recoding until an agreement rate of 80% (Ayers & Ledford, 2014) or higher was achieved. Once an agreement rate of 80% was achieved with the four studies, to further promote reliability, we coded a new set of GS and SCD study independently, and repeated the aforementioned process until an agreement rate of at least 80% was achieved.

Inter-rater agreement was calculated using agreement rate. For details on formulas, see the Inter-rater Reliability section below.

### *Training of Secondary Coders*

There were three secondary coders, all coding GSs and SCD studies. Two of the secondary coders were doctoral students in educational psychology, and the third coder was a recent doctoral graduate. I trained the secondary coders by familiarizing them with the coding manual and sharing the rationale behind the organizational structure of coding categories and variables. Using studies that have been excluded for review, I first modeled coding using one study. Second, we practiced coding the same study together and addressed any questions that emerged. Third, we coded a second, excluded study independently, compared and discussed our coding. This process was repeated until the coder felt confident with the coding process.

During Phase 3 of the selection process (as described above), I was in frequent contact with the secondary coders. We established regular meetings to check-in on the process and discussed any concerns that emerged. During these check-ins, we completed inter-rater agreement for completed studies by comparing coding responses (see the Inter-rater Reliability section below for more detail on inter-rater procedures).

A fourth secondary coder supported data extraction midway through coding. This secondary coder was my primary doctoral adviser and supported coding study information, and regional and participant demographic data when a team member became ill. I trained the coder by reviewing the coding handbook and then had her independently code two GS (i.e., peer-reviewed study and dissertation) for which secondary coding had already been completed. After independent coding, we discussed disagreements and clarified questions regarding coding procedures. This fourth secondary coder coded Set C group studies once an 80% or higher agreement rate was achieved during training.

### *Study information*

Study information was coded and used to support descriptive analysis and potential moderator analysis. Variables coded include (a) source of literature retrieval, (b) manuscript citation, (c) publication year, (d) journal name, (e) manuscript type (e.g., journal article, dissertation, unpublished manuscript, (f) publication status (e.g., peer-reviewed published, unpublished, (g) the date that coding was completed, and (h) the initials of the individual who coded the study.

### *Regional Demographics*

Information regarding where studies were conducted were coded based on information provided in the manuscript. Variables included: (a) U.S. state/region (e.g., Northeastern, Texas) and (b) geographic composition (e.g., urban, rural).

### *Participant Demographics*

Information pertaining to sample characteristics and procedures used to identify students as ELs were coded based on information provided in the manuscript. These variables included: (a) group sizes, (b) racial/ethnic composition, (c) gender composition, (d) age, (e) grade, and (f) students' home languages. Free/reduced lunch was also coded and used as a proxy to understand the socio-economic composition of students. The procedure used to identify students as ELs was also coded. Additionally, to gain a comprehensive understanding of participant demographics, the inclusion of students with disabilities was also coded (i.e., students with disabilities were included or excluded).

### *Study Design Characteristics*

Study design characteristics such as the use of random assignment or nonrandom assignment has been indicated to affect study outcomes (Newman et al., 2011; O'Keeffe

et al., 2012). Therefore, it was important to understand these factors and to include them in coding procedures. Variables that were coded included: (a) the type of research design (e.g., multiple baseline, posttest only group comparison, (b) mechanism used to assign participants to conditions (e.g., random, self-selection), (c) unit of assignment (e.g., school, student), (d) framework used to select target words, (e) the teaching of high utility words, (f) description of the control condition, and (g) instructional programming (e.g., explicit instruction, focus on breadth).

Contextual intervention information was also coded. These variables included: (i) intervention dosage, (j) intervention provider (e.g., teacher, paraprofessional), (k) types of vocabulary strategies implemented, (l) provision of professional development, (m) instructional setting (e.g., whole class), (n) language of intervention instruction, (o) total number of target words taught, (p) proportion of fidelity observations conducted, and (q) mean percent of fidelity of intervention implementation reported.

### *Study Outcome Measures*

All vocabulary and reading comprehension outcome measures were coded. Variables included: (a) production process of the measure (e.g., author created, commercially produced), (b) scope (e.g., broad, proximal), and reliability data (i.e., technical adequacy reporting, coefficients). For vocabulary measures, the type of vocabulary scale (e.g., productive, receptive), and subtypes of vocabulary constructs were coded (e.g., word identification, sentence construction).

### *Effect Size Information*

Whenever data were available, effect sizes were calculated based on raw scores provided in the manuscripts. When raw scores were unavailable, I reached out to authors

(see Analysis of Effects section for details on effect size conversions). Therefore, variables related to the calculation of effect sizes included: (a) pre- and posttest means and SDs of control and intervention groups, (b) significance test statistics, (c) effect size statistics reported by original authors, (d) confidence intervals reported by original authors, and (e) sample size and attrition rates.

### *Methodological Characteristics*

Past studies have indicated that methodological characteristics and rigor can affect the magnitude of effects observed (Carlo et al., 2004). Therefore, a series of variables were coded to understand how methodological characteristics and rigor may affect study outcomes. Variables developed and selected for coding methodological characteristics were guided by the work of Cook and colleagues (Cook et al., 2014) on quality indicators in determining the merits of a study or practice. As such, quality indicators proposed by Cook and colleagues that incorporated multiple components were re-worded to capture each unique element of the quality indicator (see Figure 3 for an example).

**Figure 3**

*Example of Cook et al. (2014) quality indicators transformed to capture each unique element.*

**Quality Indicator 4.1 (Cook et al., 2014):**

The study describes detailed intervention procedures (e.g., intervention components, instructional behaviors, critical or active elements, manualized or scripted procedures, dosage) and intervention agents' actions (e.g., prompts, verbalizations, physical behaviors, proximity) or cites one or more accessible sources that provide this information.

*Quality indicator transformed to:*

The study provided sufficient information to determine:

QIVstrat: the specific instructional strategies used during the intervention

QIdose: the overall dosage of intervention implemented

QImatS: the intervention materials (e.g., manipulatives, worksheets) used with students or cited at least one accessible source providing the information
QImatT: the intervention materials intervention providers used (e.g., teacher's manual) or cited at least one accessible source providing the information

To assess methodological rigor, studies were also coded using procedures outlined by the What Works Clearinghouse (WWC) standards for GSs (U.S. Department of Education [USDOE], Institute of Education Sciences [IES], WWC, 2017) and SCDs (Kratochwill et al., 2010). Studies were classified as meets standards without reservations, meets standards with reservations, or does not meet standards (see Appendix D for more detail). For GSs, meeting standards without reservations consists of (1) employing randomized assignment, (2) reporting acceptable attrition rates and (3) establishing equivalence at baseline. Due to the requirement of randomization, no quasi-experimental design can achieve a classification of meets standards without reservations. For SCDs, meeting standards consists of (1) the independent variable was systematically manipulated, (2) outcomes were measured across time by multiple assessors, (3) inter-rater agreement was collected for at least 20% of all sessions, (4) at least three demonstrations of an intervention effect were observed across timepoints or phases, and (5) the minimum number of data points in each phase was met.

To reduce bias and subjectivity, variables were coded based on reported information in studies confirming the presence or absence of the variable. Methodological variables were used for descriptive and moderator analyses.

Variables regarding methodological characteristics and rigor consist of: (a) setting (e.g., reporting of geographic location), (b) participants (e.g., reporting of participant ages or age range), (c) intervention agent (e.g., reporting of credentials), (d) intervention program/curriculum (e.g., reporting of materials), (e) fidelity of implementation (e.g., reporting of methods of data collection), (f) internal validity (e.g., reporting of assignment to conditions), (g) outcome measures (e.g., reporting of reliability coefficients), (h) data analyses (e.g., reporting of all effect size statistics), and (i) rigor (i.e., WWC standards).

**Analysis of Effects**

Meta-analyses are critical tools in the social sciences that allow for synthesizing effects across studies, and evaluating the magnitude and variability of those effects (Cooper et al., 2009; Gurevitch et al., 2018). The advancing technologies and methodologies of meta-analyses have enhanced researchers' ability to generalize findings beyond the studies included in a meta-analysis (Burns et al., 2010), identify evidence-based practices (Chaffee et al., 2017; Graham & Perin, 2007) and highlight research gaps (O'Keeffe et al., 2012). To capitalize on meta-analyses, findings from studies must be carefully coded and calibrated to a common scale in order to meaningfully draw conclusions and recommendations (Lipsey, 2003; Morris & DeShon, 2002). In this section, I discuss how effects were coded and how multiple outcomes within a study were treated. Next, I detail procedures on extracting data from GSs and SCD studies along with formulas used to calculate effect sizes for each respective research design. Last, I discuss the aggregation of effects using robust variance estimation to support meta-analyzing data.

To preserve the independence of effects across studies, all vocabulary and reading comprehension outcome measures were coded separately for each study. Manuscripts that consisted of more than one independent sample (e.g., study 1 and study 2) were treated as independent studies and all relevant outcome measures were coded separately. Time-series graphs in SCD studies were coded separately, and cases within a study were considered dependent.

For studies that reported multiple indices for a vocabulary outcome measure, such as a composite and subtest index, these indices were coded separately and used to evaluate effects. Coding subtest indices and composites separately allowed vocabulary constructs to remain differentiated to support overall analysis and moderator analyses. Thereby, effect sizes calculated for subtests within a study, and effect sizes calculated for multiple measures within a study were treated as dependent. Within each study, vocabulary outcome measures were categorized into one of two vocabulary scales, specific word knowledge and word learning strategies (see Table 2). This approach was taken to account for and understand the multicomponent nature of vocabulary programs in which current standards recommend the teaching of both specific word knowledge and word learning strategies as part of a comprehensive vocabulary program (Baker at el., 2014; Stahl & Nagy, 2006). Effect sizes were assigned a positive sign (+) to indicate that effects favored EL intervention students, or a negative sign (-) to indicate that effects favored ELs in the control/comparison group.

**Table 2**

*Example of Vocabulary Outcome Measures Classified as Specific Word Knowledge or Word Learning Strategy.*

| Vocabulary Taxonomy of Instruction | Vocabulary Outcome Measures |
|---|---|
| Specific word knowledge | ● Multiple-choice word identification<br>● Fill-in-the-blank definitions<br>● Pairing words with definitions |
| Word learning strategy | ● Word analysis test<br>● Polysemy production test<br>● Morphological decomposition test |

For reading comprehension measures, only composite scores were used and subtest scores were ignored. Studies that reported more than one reading comprehension outcome were treated as dependent. Distinctions between content mastery and global reading comprehension measures were not made because I was interested in the general acquisition of knowledge and understanding of text through reading.

***Data Extraction and Effect Size Calculation for Group Designs***

To support with data management, Qualtrics (2019) and Comprehensive Meta-Analysis (CMA, Version 3), were used to extract and digitally record coded variables as described in the coding manual. After all studies were coded, I exported the data from Qualtrics and CMA into a comma separated values file (csv) to complete analyses in R (R Core Team, 2013).

Effect sizes were calculated using Hedges' *g* (Hedges, 1981) on posttest means of intervention and control groups. Hedges' *g* is a standardized mean difference index that facilitates understanding the magnitude of effects between participants who received an

intervention and those in the control group. It is important to restrict effect sizes to a common metric to ensure comparability and facilitate meaningful inferences. Hence, posttest raw scores were used to calculate Hedges' *g*. When original authors reported pre- and posttest change scores, the scores were converted into group mean differences using procedures described by Morris and DeShon (2002). This conversion was done to ensure that scores were restricted to a common metric to support appropriate interpretations across studies. When authors provided pre- and posttest correlations, those correlations were used. When pre- and posttest correlations were not provided, *r*= .70 was used as a conservative correlation, which is a standard practice (Rosenthal, 1991). Pre- and posttest correlations were available for 11% of outcomes.

Hedges' *g* is a modification of Cohen's *d*, as it corrects for bias that can occur when sample sizes are small (Borenstein, 2009). Hence, it is best to use *g* when sample sizes are small (Borenstein, 2009). Since Hedges' *g* is adapted from Cohen's *d*, *d* must be calculated first and then converted into Hedges' *g*. The formula for Cohen's *d* is as follows:

$$d_i = \frac{\underline{Y}_{E_i} - \underline{Y}_{C_i}}{S_i} \qquad [\ 1\ ]$$

where,

$\underline{Y}_{E_i}$ = mean of the intervention group for the i$^{th}$ study

$\underline{Y}_{C_i}$ = mean of the control group for the i$^{th}$ study

$S_i$ = within group pooled standard deviation for the i$^{th}$ study

such that $S_i$ is calculated as:

$$S_i = \sqrt{\frac{\left(n_{E_i}-1\right)S_{E_i}^2+\left(n_{C_i}-1\right)S_{C_i}^2}{n_{E_i}+n_{C_i}-2}} \qquad [\,2\,]$$

$n_{E_i}$ = number of participants in the intervention group for the i$^{th}$ study

$n_{C_i}$ = number of participants in the control group for the i$^{th}$ study

$S_{E_i}^2$ = the variance of the intervention group for the i$^{th}$ study

$S_{C_i}^2$ = the variance of the control group for the i$^{th}$ study.

To correct for bias in $d_i, d_i$ was converted into *g* by way of the *J* correction factor, which is as follows:

$$J = 1 - \frac{3}{4df-1} \qquad [\,3\,]$$

where,

df = degrees of freedom used to estimate $S_i$ ($n_{E_i}$- $n_{C_i}$ -2).

This resulted in a final equation for Hedges' *g:*

$$g_i = J\ x\ d_i. \qquad [\,4\,]$$

***Data Extraction and Effect Size Calculation for Single-Case Studies***

When evaluating intervention effects for SCDs, it is recommended that visual analysis is paired with quantitative analysis (Harrington & Velicer, 2015; Manolov & Moeyaert, 2017). For SCDs, visual analyses support interpretations of social significance (Harrington & Velicer, 2015), which is similar to effect sizes accompanying significance testing for group designs when evaluating social significance. Visual analysis supports with evaluating the practical significance of the behavior change, and that the change occurred as a result of the independent variable (Lane & Gast, 2014). As a result, effect

size calculations and visual analysis procedures were used to code SCD outcomes for all studies that met inclusion criteria.

In order to calculate effect sizes, raw data from time-series graphs pertaining to vocabulary and reading comprehension outcomes were extracted using the free software, Webplot Digitizer (v. 4.1; Rohatigi, 2018). Webplot Digitizer has been recommended as a user-friendly and reliable tool in extracting data from SCD graphs for the purposes of meta-analyses (Moeyaert et al., 2016). The software allows for time-series graphs to be recreated. This was done by setting the boundaries of the graph in the software program and assigning approximate X- and Y-values using a point-click method. All data points were rounded to the nearest whole number. Secondary coders independently digitized all time-series graphs to ensure accurate data extraction. Agreement rate (see formula in Inter-rater Reliability section below) was used to assess reliability of data extraction on a point-by-point basis. An agreement window of 1 integer (for outcomes with counts) and 2% (for outcomes with percentage) were used to determine agreements; this is a common practice employed when digitizing time-series graphs (Moeyaert et al., 2016).

For SCD studies, the unit of interest for analysis was comparisons between baseline and intervention, hence, data from baseline and the adjacent intervention phase (labeled AB contrast hereafter) were extracted. To maintain the unit of analysis, conventions recommended by Scruggs and Mastropieri (1994) regarding effect size calculations for specific research designs were applied. For studies that employed multiple treatment design, the AB contrast consisted of the most effective treatment and its corresponding baseline. An effect size was calculated for each participant or case within a study. For studies that employed a multiple baseline/probe design across

behaviors, AB contrasts were calculated for each tier and then averaged, using the number of tiers as the denominator. For studies that employed a multiple baseline/probe design across participants, AB contrasts were calculated for each participant. Since maintenance phases were not consistently implemented in all intervention studies, to support interpretation of findings across studies, maintenance phases were not included in the current review and thus, data were not extracted.

Two statistics were calculated to evaluate intervention effects, Baseline Corrected Tau (*BCTau* hereafter; Tarlow, 2017) and between-case effect size (*BCES* hereafter; Pustejovsky et al., 2014). *BCTau* is a nonoverlap index that uses pairwise comparisons to assess for concordance or discordance (Brossart et al., 2018; Zimmerman et al., 2018) of observations between baseline and intervention phases. Procedures proposed by Tarlow (2017) are selected for this review because calculations account for baseline trends that if ignored, can affect erroneous inferences (Parker et al., 2011). *BCTau* accounts for baseline trend by performing a Theil-Sen regression, which removes unexplained variance from the linear trend (Tarlow, 2017). The Theil-Sen regression would only be performed when a statistically significant baseline trend was detected. Procedures that were used are as follows:

1. Confirming that a baseline trend existed by calculating Kendall's Tau (Zimmerman, et al., 2018) for baseline observations.
2. If a trend was detected, a Theil-Sen regression was performed. If no trend was detected, skipped the Theil-Sen regression and advanced to step 3.
3. Calculated Kendall's Tau (Zimmerman, et al., 2018) with the corrected data or original data based on results from step 2.

$$Kendall's\ Tau\ = \frac{N_c - N_d}{N_c + N_d} \qquad [5]$$

where,

$N_c$= number of concordance observations

$N_d$= number of discordance observations.

To promote accurate calculations, *BCTau* was calculated using an online software designed by Tarlow (2016). Due to differences in parametric assumptions and the fact that *BCTau* and group design effect size indices are inequivalent metrics, it is inappropriate to combine *BCTau* with GSs (Shadish et al., 2015; Zimmerman et al., 2018). Therefore, *BCTau* will be meta-analyzed and interpreted only for SCDs.

The *BCES* was proposed by Pustejovsky and colleagues (2014) to mimic effect sizes used in group designs (Odom et al., 2018; Shadish et al., 2015). The advantage of using *BCES* compared to other SCD between-case effect sizes is its ability to account for trend in both baseline and intervention. *BCES* uses multilevel modeling to calculate effects and capture linear changes in outcomes each time the behavior is measured (Pustejovsky et al., 2014). The index has been applied and found useful in interpreting findings in multiple systematic reviews (Maggin et al., 2017; Peterson et al., 2019). However, the application of *BCES* is restricted to multiple baseline/probe designs across participants and reversal designs, with a minimum requirement of at least three participants within each study. Currently, there is not a between-case effect index for alternating treatment designs (Odom et al., 2018).

*BCES* was calculated based on the procedures described by Pustejovsky and colleagues (2014), using software developed by Pustejovsky (2016). This practice

supports accurate calculations as the software visually models the data for inspection and calculates the *BCES* index for its respective research design. As mentioned previously, *BCES* was developed to align with effect size indices used for GSs. As such, *BCES* was combined with Hedges' *g*, and meta-analyzed with GSs. This practice is encouraged (Hedges et al., 2013; Pustejovsky et al., 2014), has been demonstrated to be valid (Zelinsky & Shadish, 2018), and applied successfully in published literature (Petersen-Brown et al., 2019).

### *Determining Functional Relations for SCDs*

As mentioned previously, an essential aspect of visual analysis is to confirm a functional relation (Lane & Gast, 2014), which refers to the systematic demonstration of changes in the behavior covarying with manipulations of the intervention (Gast & Spriggs, 2014; Ledford et al., 2018). For this review, determining a functional relation consisted of assessing (1) trend, such that the data path follows in the desired and expected direction, and (2) level, such that a mean change is observed from baseline to intervention in the desired and expected direction (See Appendix D for operational definitions).

**Aggregation of effect sizes.** During the National Reading Panel's (NRP & NICHD, 2000) extensive review on reading research, the NRP reported that vocabulary studies varied so greatly that it was challenging to devise a taxonomy to classify studies. This was anticipated for the current study because vocabulary interventions are multicomponent and multifaceted (Hiebert & Kamil, 2005; Stahl & Nagy, 2006). The NRP (NRP & NICHD, 2000) also found that vocabulary assessments varied widely across studies. Hence, it was hypothesized that there would be widespread variability in

sample sizes, instructional strategies, and outcome measures used for vocabulary learning and to evaluate vocabulary effects. In fact, it is common practice for researchers to develop their own outcome measures in evaluating vocabulary learning (Marulis & Neuman, 2010; NRP & NICHD, 2000; Shannon & Beck, 2006). Given these reasons, a random effects model (REM) was selected for the current meta-analysis.

The goal of a REM is to estimate the average effect of a sample of studies under the assumption that effect sizes are randomly distributed across studies (Borenstein 2008; Borenstein et al., 2010). As such, the application of REM implies that there is not one true effect size, but rather, a distribution of effects (Borenstein, 2008; Hedges et al., 2010). This is an advantage of a REM. Additionally, REM allows the opportunity to make inferences that extend beyond studies that have been included in the current review (Borenstein, 2008).

There are multiple approaches to meta-analyze data and arrive at summary statistics. Many approaches require assumptions of independence in effects, which can be problematic for studies that report multiple outcomes because dependence arises from correlated errors. Methods to address such dependency have involved taking the average of the effects, selecting the outcome that is most representative of the study, or randomly selecting one effect (Borenstein, 2008). Unfortunately, these methods may result in a loss of information, because not all effects and their true contributions can be accounted for or estimated.

Hedges, Tipton, and Johnson (2010) proposed a method that circumvents problems when multiple outcomes are reported in a study. Robust variance estimation (RVE) uses a mathematical process to model dependencies by accounting for the

correlated relation between effect sizes within a study and effects across studies. This is done by assuming that the correlation between pairs of effects sizes are constant across all studies and assuming that sampling variances within studies are nearly equal (Hedges et al., 2010; Tanner-Smith & Tipton, 2014). As such, RVE's weighting matrix integrates the covariance and variance between pairs of effect sizes, allowing the use of all effect sizes in a meta-analysis. The benefit of using all relevant effect sizes in a meta-analysis promotes precision with better estimates of variance and estimates of the mean effect size (Hedges et al., 2010; Tanner-Smith & Tipton, 2014).

RVE has been applied reliably in multiple meta-analyses spanning various fields such as education (Dietrich, 2009; Gardella et al., 2017), healthcare (Shields et al., 2016), and social sciences (Bediou et al., 2018; Klingbeil et al., 2017). These studies indicate the dynamic use of RVE for meta-analyses.

The advantages of RVE and its appeal are that, since the core of its processes is to adjust standard errors of regression coefficients, it does not require normality in distribution or conditions for weights. Therefore, RVE can be used with any weight scheme and dependence type such as correlated effects (e.g., multiple outcomes reported in one study) or hierarchical effects (e.g., multiple studies with the same research team; Hedges et al, 2010; Tanner-Smith & Tipton, 2014). The authors emphasize that weights in RVE are intended for statistical efficiency and are not used to indicate precision of a study (Hedges et al, 2010; Tanner-Smith & Tipton, 2014). Given that different dependence types may exist across studies, it is recommended that the weighting scheme employed be based on the most common dependence type represented in a meta-analysis (Tanner-Smith & Tipton, 2014).

From previous work, the majority of studies in the area of EL vocabulary interventions appeared to report multiple outcomes within a study and few studies overlapped in research teams (Xiong, 2018); the same could be said about the current sample of studies. The majority of studies retrieved reported multiple outcomes, and thus it was best to use methods that address correlated effects. Analyses were conducted according to procedures presented by Hedges et al. (2010) and Tanner-Smith and Tipton (2014), with the following regression model to address correlated effects:

$$T_{ij} = \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_2 X_{2ij} + \varepsilon_{ij} \quad [6]$$

Where,

$T_{ij}$ = effect size $i^{th}$ in study $j^{th}$

$\beta$ = weighted least-squares estimates

X = value of the covariate for the $i^{th}$ effect size in the $j^{th}$ study

$\varepsilon$ = vector of residuals.

The equation for weights is as follow:

$$W_{ij}^{*} = \frac{1}{\{(V_{\cdot j} + \tau^2)\,[1+(k_j-1)\rho]\}} \quad [7]$$

where,

$v_{\bullet i}$ = average of the within-study variances for study *j*

$\tau^2$ = estimate of the between-study variance

$k_i$ = number of effect sizes within study *j*

$\rho$ = correlation between all pairs of observed effect sizes.

Since RVE assumes that the correlation between pairs of effect sizes is constant, a value for ρ was set to 0.80 given the likelihood that outcomes would be highly correlated.

To ensure that $\rho$= 0.80 was appropriate, a sensitivity test to evaluate values between 0 and 1 was conducted on the estimated variance. Although a sensitivity test was conducted to increase confidence that $\rho$= 0.80 is appropriate, empirical data (Wilson, Tanner-Smith, Lipsey, Steinka-Fry, & Morrison, 2011) and statistical modeling (Tanner-Smith & Tipton, 2014) suggests that the value of $\rho$ is not likely to affect the validity of confidence intervals or statistical inferences, even when an inappropriate value has been selected. Readers interested in proofs and details of methods of RVE or small sample corrections to the method of moments estimator may refer to Hedges et al. (2010) and Tipton (2015), respectively. Analyses were supported by software packages in R (R Core Team, 2013), namely robumeta (Fisher & Tipton, 2015).

Heterogeneity of effects was assessed by calculating $I^2$, which represents the proportion of variance due to systematic differences (Shadish & Haddock, 2009). Significant levels of $I^2$ can prompt further investigations to explain systematic differences such as moderator analyses. The formula for $I^2$ is as follow:

$$I^2 = 100\% \left(\frac{Q-(n-1)}{Q}\right) \qquad [8]$$

where,

$Q$ = homogeneity test statistic

$n$ = number of studies.

***Overall Analysis***

The current paper aimed to meta-analyze GSs and SCD studies to evaluate the effects of vocabulary instruction on vocabulary learning and reading comprehension.

Table 3 summarizes the analysis of effects and synthesis of data with the corresponding research question.

**Table 3**

*Research Questions and Corresponding Analysis Plan*

| Research Questions | Analysis of Effects |
|---|---|
| 1. What is the overall quality of the research? | Descriptive analysis using quality indicators and methodological standards guided by Cook et al., (2008), WWC (USDOE, IES, WWC, 2017), and Kratochwill (2010). |
| 2. To what extent are vocabulary programs and interventions effective in increasing vocabulary learning for English Learners? | |
| 2a. What is the average or overall effect of vocabulary instruction on vocabulary learning for English Learners? | The combination of Hedges' g (GSs) and BCES (SCDs) of vocabulary outcome measures will be meta-analyzed and synthesized using RVE to estimate effects.<br><br>TauBC (SCDs) from vocabulary outcome measures will be meta-analyzed separately using RVE to provide a secondary estimate of effects. |
| 2b. To what extent do methodological characteristics (e.g., participant demographics, study design characteristics, and measurement methods) moderate study outcomes? | Meta-regression will be used to evaluate moderators affecting vocabulary learning effects. |
| 3. To what extent are vocabulary programs and interventions effective in increasing reading comprehension for English Learners? | |
| | Hedges' g (GSs) of vocabulary outcome measures will be meta-analyzed and synthesized using RVE to estimate effects. |

***Moderator Analysis***

Understanding fully that vocabulary programs will vary significantly from one study to another (NRP & NICHD, 2000), heterogeneity in effects was likely. To understand systematic unexplained variance, moderator analyses were used to examine variables that appeared to correlate with results (Hall & Rosenthal, 1991). Variables of interest may be theoretically or empirically identified (Wood & Eagly, 2009). Overlooking the opportunity to investigate potential reasons for unexplained variance in a meta-analysis could lead to inappropriate interpretations and threaten the validity of construct and statistical conclusions (Hall & Rosenthal, 1991; Wood & Eagly, 2009). Moderator analyses can add a great deal to a meta-analysis by providing clues as to whom particular interventions may be more effective for, or conditions that promote large effects (Bloch, 2014).

Meta-regression has been recommended to examine moderators to allow for the examination of multiple predictors in a model (Lipsey & Wilson, 2001; Pigott, 2012). Given that past syntheses on vocabulary research for ELs could not address moderators of effects, and the possibility of a small sample size that can affect power, moderators of interest for the current review were identified and prioritized based on findings from monolingual syntheses (Marulis & Neuman, 2010; NRP & NICHD, 2000) and my previous work (Xiong, 2018).

For the current review, moderator analyses consisted of examining the association between intervention effects and participant demographics, specifically, grade level groupings (i.e., K-5, 6-8, 9-12). Interest in this variable was based on past studies that have found differences among effects based on age (NRP & NICHD, 2000; Swanborn &

de Glopper, 1999). The association between intervention effects and study design characteristics were also examined. Specifically, intervention dosage (frequency, intensity, duration), intervention provider, and domain of target vocabulary words were examined. Past studies have indicated that intervention dosage and intervention provider can affect the magnitude of intervention outcomes and lead to significant implications for practice (Marulis & Neuman, 2010; NRP & NICHD, 2000; Swanborn & de Glopper, 1999). Target word domain was identified for moderator analysis because there is limited agreement on how words should be selected for instruction and which words should be used for instruction (Baker et al., 2014; Beck et al., 2013; Hairrell et al., 2001). Given various recommendations for selecting words for instruction, it was hypothesized that differences observed in effects may be a function of target word domains. Current research in this area is limited, therefore this variable necessitated further examination.

Due to past studies suggesting that different vocabulary measures exhibit varying sensitivities in detecting effects (NRP & NICHD, 2000; Swanborn & de Glopper, 1999) and that students can have varying levels of knowledge across different vocabulary measures (Kuhn & Stahl, 1998), it is important to examine intervention effects on measurement methods. Specific interest was focused on whether differences in effects can be explained by how measures were produced (i.e., author created assessments), the type of vocabulary scale being measured (i.e., productive, receptive, or mixed), and the taxonomy of instruction being assessed (i.e., specific word knowledge or word learning strategy).

Lastly, methodological rigor was assessed using the standards of practice proposed by Kratochwill and colleagues (2010) and WWC (USDOE, IES, WWC, 2017).

Methodological rigor is likely to account for a portion of systematic differences in effects. Past studies in the social sciences and education have observed inconsistencies regarding the influence of methodological rigor on the magnitude of effects. Studies with high rigor have been observed to produce both large effects (Maggin et al., 2017) and small effects (Klingbeil et al., 2017) when compared to low rigor studies. For these reasons, it was critical to assess methodological rigor in order to draw appropriate and accurate inferences from results. A summary of moderators identified and prioritized for analysis can be found in Table 4.

**Table 4**

*Variables of Interest for Moderator Analyses*

| | Potential Moderator Variables |
|---|---|
| Participant demographics | ● Grade level groupings |
| Study design characteristics | ● Intervention provider<br>● Intervention dosage<br>● Target word domains |
| Measurement methods | ● Production of the measure<br>● Type of vocabulary scale<br>● Taxonomy of instruction |
| Methodological rigor | ● Standards of practice |

## Inter-rater Reliability

The use of reliability checks increases confidence that results are accurate and reliable (Gast & Ledford, 2014; Gersten et al., 2005). Agreement rate (AR) was used as the inter-rater reliability index.

### *Agreement Rate*

Agreement rate (AR) was calculated to address both continuous and categorical variables for an overall inter-rater reliability index. The index was selected because it is commonly used in the field and supports easy interpretation (Orwin & Vevea, 2009). Agreement rate was calculated on an item-by-item basis with the following equation:

$$AR = \frac{number\ of\ agreements}{number\ of\ agreements + disagreements} * 100 \qquad [9]$$

# Chapter 4

## Results

The purpose of the current review was to synthesize experimental studies to understand the overall effects of vocabulary instruction and interventions on vocabulary learning and reading comprehension for ELs. Effects were examined across group studies (GS) and single-case design studies (SCD) to gain a comprehensive understanding of overall effects.

This chapter presents results from the current study. The chapter discusses descriptive results, followed by presenting meta-analyzed data for GSs and SCDs, and concluding with results from moderator analyses. For reference, *n* represents the sample size of studies and *k* represents the sample size of effects.

### Descriptive Results

#### *Study Information*

Literature retrieval was conducted in April 2019 following procedures outlined in Chapter 3. The literature retrieval resulted in 1,203 papers. After the removal of duplicates and initial screening, 70 individual papers remained for further review of eligibility (see Figure 2 for flow diagram of study selection process). Twenty-one papers (30%) required a request for additional data for which four authors responded to data requests. One author indicated that data for vocabulary subscales could not be provided and only full reading scale scores were available. As such, the Cassady et al. (2018) study was excluded. Another author provided data for vocabulary measures, however, vocabulary outcomes were only administered to students in the intervention condition. Therefore, the O'Connor et al. (2017) study was excluded because vocabulary effects

could not be examined between students in the intervention and control conditions. Data provided for the Wanzek et al. (2017) and Neuman & Kaefer (2018) papers met inclusion criteria and were included in the current study. Hence, data provided by these authors were used for analysis (Neuman & Kaefer, 2018; Wanzek et al., 2017).

Taking into consideration the Wanzek et al. (2017) and Neuman & Kaefer (2018) papers, initial screening of appropriate papers for inclusion resulted in 70 individual papers. Of the 70 papers, 41 papers met inclusion criteria for a total of 45 studies ($n$= 10 SCDs, $n$= 35 GSs) and 184 effect sizes ($k$= 119 SCDs, $k$ = 65 GSs). Four of the 41 papers consisted of two independent studies (Graves et al., 2011; Tong et al., 2014; Vaughn et al., 2006; Vaughn et al., 2009), therefore resulting in the final count of 45 total studies.

Studies were published between 1996 and 2018 (See Table 5 for study characteristics). More than a third of studies were unpublished dissertations ($n$= 17) while remaining studies were peer-reviewed published studies (62%). For GSs, 63% of studies used random assignment, 26% used nonrandom assignment (e.g., systematic, self-selection), and in 9% of studies, assignment of participants to conditions could not be determined. For SCDs, the majority of studies implemented a multiple baseline/probe design (90%).

***Inter-rater Agreements***

Agreement rate (AR) was used as an inter-rater reliability index for coding extracted data. AR was calculated for data-series extraction, categorical variables and continuous variables noted in Chapter 3. The overall mean AR was 98% (*range*= 88%-

**Table 5**

*Study Characteristics*

| Citation | Journal Name | Manuscript Type | *n* | Grade | Race/ Ethnicity | Students with Disabilities | Geographic Area | U.S. Region |
|---|---|---|---|---|---|---|---|---|
| Alison et al. 2017 | Journal of Special Education Technology | Peer-Reviewed | 3 | Elementary | 66% Hispanic 33% African American | Y | Suburban | Southeast |
| Anderson 2014 | Dissertation | Dissertation | 6 | Elementary | 100% Hispanic | Y | Urban | Southeast |
| August et al. 2009 | Journal of Research on Educational Effectiveness | Peer-Reviewed | 562 | HS | NR | | Urban | Southwest |
| Avila & Sadoski 1996 | Language Learning | Peer-Reviewed | 63 | Elementary | 100% Hispanic | | Urban | Southwest |
| Benoit 2017 | Dissertation | Dissertation | 5 | MS | NR | | Suburban | Southeast |
| Bravo & Cervetti 2014 | Equity & Excellence in Education, | Peer-Reviewed | 115 | Elementary | NR | Y | Rural & Suburban | West |
| Burns 2001 | Dissertation | Dissertation | 78 | Elementary | NR | | NR | Northwest |
| Cannon et al. 2010 | Communication Disorders Quarterly | Peer-Reviewed | 4 | Elementary | NR | | Urban | Southeast |
| Cena et al. 2013 | Reading and Writing | Peer-Reviewed | 50 | Elementary | NR | | NR | Northwest |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Cervetti et al. 2015 | Contemporary Educational Psychology | Peer-Reviewed | 147 | Elementary | NR | | NR | Mid-Atlantic |
| Crevecoeur et al. 2014 | Reading & Writing Quarterly | Peer-Reviewed | 4 | Elementary | 100% Hispanic | | Urban | Northeast |
| Crum 2017 | Dissertation | Dissertation | 99 | Elementary | NR | | NR | Southeast |
| Cruz-Cruz 2005 | Dissertation | Dissertation | 28 | Elementary | 93% Hispanic<br>4% African American<br>4% White | | NR | Southwest |
| Dack 1996 | Dissertation | Dissertation | 53 | MS | 94% Hispanic<br>6% Asian | | Urban | Southwest |
| Denton et al. 2008 | Learning Disabilities Research & Practice | Peer-Reviewed | 22 | MS | NR | Y | Urban | Southwest |
| Frasco 2008 | Dissertation | Dissertation | 34 | Elementary | NR | | Rural | Central |
| Graves et al. 2011 | The Elementary School Journal | Peer-Reviewed | NA | MS | NR | Y | Urban | West |
| Graves et al. 2011 | The Elementary School Journal | Peer-Reviewed | 50 | MS | NR | Y | Urban | West |
| Green et al. 2015 | Contemporary Issues in Communication Science and Disorders | Peer-Reviewed | 2 | Elementary | 100% Hispanic | Y | Urban | Southwest |
| Guardino et al. 2014 | Communication Disorders Quarterly | Peer-Reviewed | 3 | MS, HS | 67% Hispanic<br>33% white | Y | NR | Southeast |
| Helman 2015 | Dissertation | Dissertation | 4 | HS | 100% Hispanic | Y | Urban | NR |
| Helman et al. 2015 | Learning Disability Quarterly | Peer-Reviewed | 3 | HS | 100% Hispanic | Y | Urban | Mid-Atlantic |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Hinrichs 2008 | Dissertation | Dissertation | 5 | Elementary | NR | | Suburban | Midwest |
| Kieffer et al. 2012 | The Elementary School Journal | Peer-Reviewed | 349 | MS | NR | | Urban | West |
| Kim & Linan-Thompson 2013 | Remedial and Special Education | Peer-Reviewed | 4 | Elementary | 100% Hispanic | | NR | Southwest |
| Kittley-Koshenina 2009 | Dissertation | Dissertation | 15 | Elementary | 100% Hispanic | | Urban | Southwest |
| Lawrence et al. 2012 | Bilingualism: Language and Cognition | Peer-Reviewed | 117 | MS | NR | | Urban | Northeast |
| Lia 2010 | Dissertation | Dissertation | 4 | Elementary | 75% Asian<br>25% Hispanic | | Suburban | Midwest |
| McBroom 2009 | Dissertation | Dissertation | 4 | Elementary | NR | | NR | Southeast |
| Mieure 2014 | Dissertation | Dissertation | 73 | Elementary | 99% Hispanic<br>1% Asian | | Urban | West |
| Nelson et al. 2011 | Journal of Literacy Research | Peer-Reviewed | 3 | Elementary | 100% Hispanic | | NR | Central |
| Neuman & Kaefer 2018 | Contemporary Educational Psychology | Peer-Reviewed | 84 | Elementary | NR | Y | Urban | NR |
| Proctor et al. 2011 | Reading and Writing | Peer-Reviewed | 4 | Elementary | 49% Hispanic | | Urban | Northeast |
| Stevens 2018 | Dissertation | Dissertation | 6 | MS | NR | | NR | Appalachia |
| Tong et al. 2014 | The Journal of Educational Research | Peer-Reviewed | 56 | Elementary | 100% Hispanic | | Urban | Southwest |
| Tong et al. 2015 | The Journal of Educational Research | Peer-Reviewed | 56 | Elementary | 100% Hispanic | | Urban | Southwest |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ulanoff & Pucci 1999 | Bilingual Research Journal | Peer-Reviewed | 60 | Elementary | 100% Hispanic | | Urban | West |
| Vang 2004 | Dissertation | Dissertation | NA | Elementary | NR | | NR | West |
| Vaughn et al. 2006 | American Educational Research Journal | Peer-Reviewed | 89 | Elementary | 100% Hispanic | | Urban | Southwest |
| Vaughn et al. 2006 | American Educational Research Journal | Peer-Reviewed | 94 | Elementary | 100% Hispanic | | Urban | Southwest |
| Vaughn et al. 2009 | Journal of Research on Educational Effectiveness, | Peer-Reviewed | 97 | MS | NR | | NR | Southwest |
| Vaughn et al. 2009 | Journal of Research on Educational Effectiveness, | Peer-Reviewed | 106 | MS | NR | | NR | Southwest |
| Wanzek et al. 2017 | Journal of Educational Psychology, | Peer-Reviewed | 60 | Elementary | NR | Y | Rural & Urban | NR |
| Weitz 2003 | Dissertation | Dissertation | 120 | HS | NR | | NR | West |
| Yang 2015 | Dissertation | Dissertation | 64 | Elementary | NR | | Suburban & Urban | Southwest |

Note. n= sample size of EL students. NR= not reported. Y= yes

100%). The mean AR for data-series extraction was 96% (*range*= 83%-100%). The low AR rate of 83% was the result of extracting data from a study in which data-series graphs were rotated. The rotated graphs increased disagreements on a point-by-point basis. Nonetheless, an AR of 83% still fell in an acceptable range for reliability (Orwin & Vevea, 2009).

***Assessing for Outliers***

The dataset was assessed for extreme outliers. Extreme outliers were defined as effects that were ±3.0 *SD* larger than the overall estimate (Greenhouse & Iyengar, 2009; Marulis & Neuman, 2010). Outliers for GS vocabulary effects were examined; no effects fell beyond the range of extreme values.

Examining the confidence intervals of effects and their overlap with the estimate (i.e., $g = 0.36$, $SE = 0.07$, $CI_{95} = 0.23\text{-}0.50$, $p < .001$ of the GS vocabulary dataset) was also conducted to further assess for extreme outliers (Harrer et al., 2019). Effects that had confidence intervals outside of the confidence intervals of the estimate may be considered extreme. This process was aided by using R (R Core Team, 2013) and the dmetar package (Harrer et al., 2019). The analysis indicated that 12 studies signified potential extreme values because confidence intervals did not overlap with the confidence interval of the estimate (See Appendix E for analysis results). The analysis generated a model scenario that set the weights of the 12 effects across 10 studies to a zero-value and provided a new overall estimate of *g*. This resulted in $g = 0.36$ ($CI_{95} = 0.27\text{-}0.45$, $p < .0001$) with $T^2 = 0.05$ and $I^2 = 47\%$. Given that the differences in the estimate did not change significantly, the overall estimate remained statistically significant, and no effects

were ±3.0 *SD* from the estimate, all studies and their effects were retained for meta-analyzing.

Although *BCTau*, a nonoverlap index, was not the primary effect size of interest for meta-analyzing in the full dataset, outlier analysis was still conducted as described above. No extreme outliers were observed that fell in the ±3.0 *SD* range. Using the dmetar package, eight effects across four studies were identified as potential outliers (See Appendix E for analysis results). The model scenario set the weights of these effects to zero, resulting in *BCTau*= 0.80 ($CI_{95}$ = 0.79-0.82, $p$ <.0001) with $T^2$= 0 and $I^2$= 0%. Given that the differences in the estimate did not change significantly, the overall estimate remained statistically significant, and no effects were ±3.0 *SD* from the estimate, all studies and their effects were retained for the *BCTau* dataset.

The SCD dataset for *BCES* was not examined for outliers. Using *BCES* as a group-design comparable effect size for meta-analyses remains in its early stages (Hedges et al., 2013; Shadish & Haddock, 2009; Shadish et al, 2015) and limited guidance has been developed for application in meta-analyses for the effect size. To remove or adjust effects with limited theoretical support may compromise results that could overcorrect or undercorrect inappropriately (Cooper et al., 2009; Greenhouse & Iyengar, 2009), resulting in loss of valuable data. Given that *BCES* will contribute only a limited number of effects to the final model ($k$ = 13), accounting for less than 20% of all effect sizes, no study was removed or adjusted. In summary, the full dataset was retained without making adjustments.

***Publication Bias***

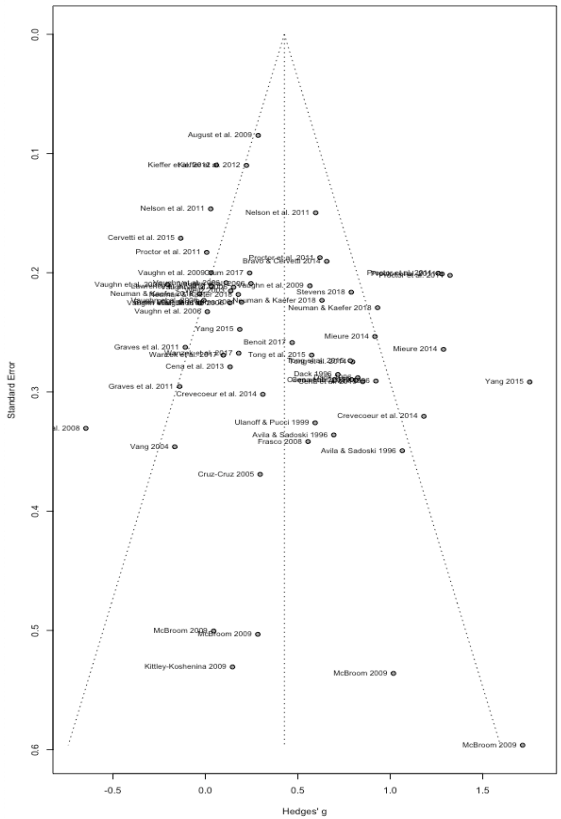
When conducting meta-analyses, it is important to assess for publication bias, which is the relation between the probability of publishing and study results that are statistically significant (Sterne & Harbord, 2004). Failing to consider the influence of publication bias may lead to inappropriate and potentially inflated results (Sterne & Harbord, 2004; Sutton, 2009).

A funnel plot was used to examine publication bias. Funnel plots are scatterplots that graph effect sizes in relation to their standard errors or a statistic of precision (see Figure 4). As such, effects plotted populate smaller studies with less precision near the bottom of the graph and larger studies with more precision near the top of the graph. The line running through the middle of the graph represents the overall estimate. Publication bias may be present when data fall outside of the funnel shape and show asymmetry, particularly with few published studies populated in the lower right area of the graph. However, other factors besides publication bias may explain asymmetry observed on a funnel plot. Factors such as poor rigor, using measures that lack sensitivity in detecting effects, or true differences in studies (Egger et al. 1997; Sutton, 2009) may produce asymmetry.

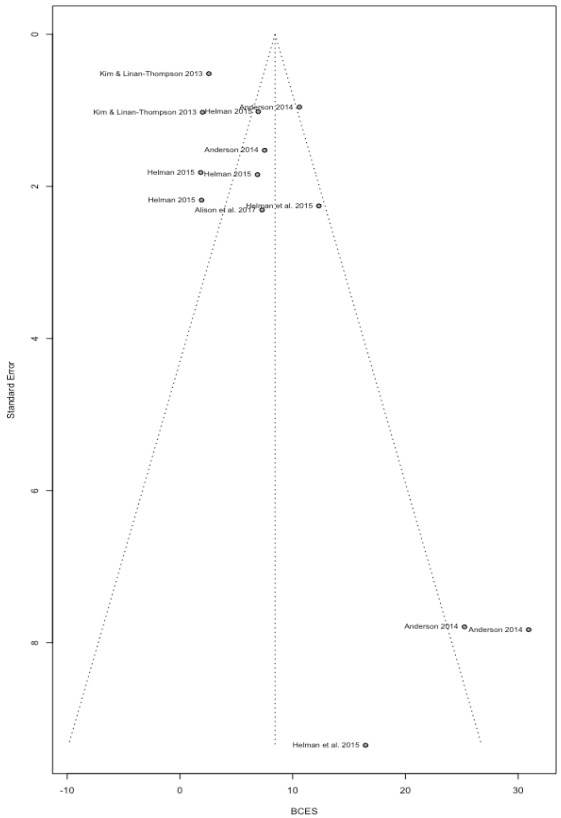
Guidelines have not been developed for RVE and publication bias. Therefore, effects were fitted using a random effects model (REM) using unweighted effects. Furthermore, due to the significant variability of effects across the three effect sizes of interest (i.e., *g*, *BCES, BCTau*), data for each set of effect sizes were plotted separately to support meaningful analysis. Fitting the model and graphing the funnel plots were completed using R (R Core Team, 2013) and the metaphor package (Viechtbauer, 2010).

**Figure 4**

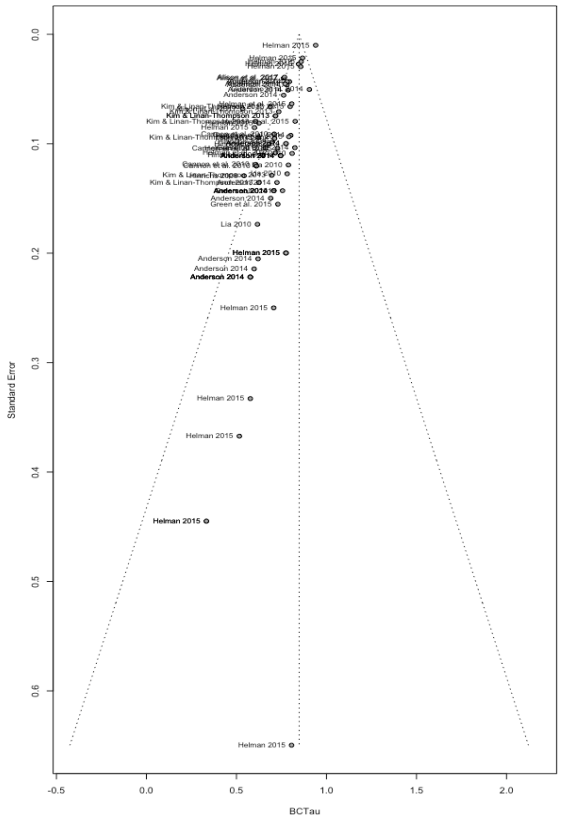
*Funnel Plots of Vocabulary Effects*




Panel A




Panel B




Panel C

Visual inspection of the three funnel plots suggested asymmetry. Outcomes for *g* (*k* = 65) did not disperse evenly along the funnel shape with effects more likely to occupy the upper left area of the graph and only unpublished studies populating the lower region of the graph (See Figure 4, Panel A). For outcomes of *BCES* (*k*= 13), effects tended to occupy the upper left region, falling outside of the funnel shape and few published studies populated the lower region (See Figure 4, Panel B). For outcomes of *BCTau* (*k*= 75), although effects appeared to follow a funnel shape, effects populated mainly on the left side of the graph with few effects populating the right side of the graph (See Figure 4, Panel C). Egger's regression tests (Egger et al., 1997) were used as a statistical method to test for bias and asymmetry. Egger's regression tests were conducted in R (R Core Team, 2013) using the dmetar package (Harrer et al., 2019). Regression results for all three funnel plots were statistically significant (*g*= 0.43, $p < .001$; *BCES*= 8.44, $p$=.002; *BCTau*= 0.81, $p$ <.0001), indicating that asymmetry and potential publication bias were detected. These results not only suggested that asymmetry was observed, but also that effects observed in larger studies were statistically significantly different from effects observed in smaller studies across all three effect size indices of interest.

To understand the degree to which publication bias may affect overall results of the meta-analysis, the *Fail-safe N* statistic was calculated. The *Fail-safe N* provides a statistic that assesses the stability of study findings (Carson, 1990; Sutton, 2009). The statistic communicates the number of additional studies, particularly unpublished and/or nonsignificant studies that may be stored away, that if retrieved, would nullify results. *Fail-safe N* was calculated separately for all three effect sizes, in addition to the combined model for *g* and *BCES*. Calculations were completed in R (R Core Team,

2013) using the metaphor package (Viechtbauer, 2010). For GSs, the *Fail-safe N* noted that an additional 4,699 studies were needed to nullify findings. The *Fail-safe N* was 12,404 for the *BCTau* dataset and 1,023 for the *BCES* dataset. For the combined model, the *Fail-safe N* indicated that 10,165 studies were needed to nullify findings of the current study. Given the limited research in the area of English Learners, and specifically, vocabulary instruction for English Learners, it is unlikely that a significant number of unpublished studies exist and the substantial size of 10,165 additional studies needed to nullify results suggest that findings from the current study would unlikely change in the event of newly discovered studies (Sutton, 2009). Nonetheless, caution is still warranted when interpreting results because no study, regardless of its level of rigor, is free from limitations.

***Participants and Settings***

Across the 45 studies, there were 2,805 ELs. These students represented as few as 7% of the original study sample to 100% of the study sample ($n$= 22 GSs with 100% of ELs in the original sample; $n$= 10 SCDs with 100% of ELs in the original sample). Less than half of studies (47%) reported racial/ethnic demographics for ELs. In the studies that reported racial/ethnic demographics, Hispanic students were included in all studies. Hispanic students represented between 25% to 100% of the EL sample in the original studies. Almost 20% of studies did not provide sufficient information to understand EL students' home language. Of the studies that reported home languages, ELs spoke 27 different languages with Spanish as the most commonly reported home language ($n$= 37), followed by Vietnamese as the second most cited language spoken by ELs ($n$= 4). Only 27% of studies reported including students with disabilities in the sample. Students with

disabilities included those with Autism, hearing impairment, learning disability and speech or language impairment. Two-thirds of studies were conducted with elementary students and 7% of studies were conducted with high school students.

The majority of studies were conducted in urban areas (51%) with almost 30% of studies failing to provide sufficient information to determine regional demographics (e.g., urban, rural, suburban). Regarding U.S. regions, as designated by IES (IES, n.d.), the top three regions where studies were most likely to be conducted were the Southwestern (40%), Western (19%) and Southeastern U.S. (17%). Studies conducted in the southwest were all conducted in Texas.

***Intervention Characteristics***

Of the 45 studies, almost half of studies (47%) consisted of vocabulary interventions that provided explicit vocabulary instruction and 31% of studies provided a combination of explicit (e.g., teaching word parts, providing definitions for target words) and incidental instruction (e.g., read aloud, independent reading). Approximately 18% of studies ($n$= 8) did not provide sufficient information to determine specific vocabulary instructional strategies that were implemented. The most frequently cited vocabulary instructional strategies used consisted of visual/picture pairings to teach target words (Lia, 2010; Vaughn et al., 2009), providing multiple and repeated exposures to target words (Crevecoeur et al., 2014; Graves et al., 2011), teaching definitions of target words (Guardino et al., 2014; Tong et al., 2014), engaging students in discussions focused on target words (Denton et al., 2008; Hinrichs, 2008), and using graphic organizers to teach target words (Mieure, 2014; Stevens, 2018). Some of the least cited strategies included acting out the meaning of target words (McBroom, 2009) and pre-teaching target words

(Ulanoff & Pucci, 1999). Eight studies provided Spanish translations of instructional materials to students (e.g., August et al., 2009; Vaughn et al., 2009). There were also eight studies that integrated the teaching of cognates (e.g., Cervetti et al., 2015; Proctor et al. 2011).

Target words that were used for instruction were categorized into four domains: basic/function, general academic, academic content-specific, and mixed-method (see Table 6 for descriptive summaries). Twelve studies (27%), provided instruction on general academic words, which were words that were drawn from the curriculum, were frequently used across multiple disciplines and typically need to be learned in order to access more complex topics (e.g., establish, personality, dreaming, inspire, generate). More than a fifth of studies (22%) provided instruction on academic content-specific words, or words that were specific to academic disciplines (e.g., science, social studies) and words with technical definitions (e.g., nomads, roots, segregation, veto). Twenty percent ($n$= 9) of studies provided instruction on a combination of general academic and academic content-specific words. Across the 45 studies, the mean number of target words taught was 70 words (*range*= 6-383), with the mean number of words in GSs being 82 (*range*= 10-383) and 43 words for SCD studies (*range*= 6-120). Almost half of studies (47%; $n$= 20 GSs, $n$= 1 SCD) did not report the number of words taught during the intervention period.

The majority of interventions were delivered to the whole class in GSs (57%), followed by small groups of five or fewer students (17%). For SCD studies, the majority of interventions were delivered in small groups of five or fewer students (50%), followed by one-on-one instruction (30%). Approximately 11% of all studies did not provide

sufficient information to determine the group format in which interventions were delivered. Intervention providers were most likely to be classroom teachers for GSs (60%) and research teams/first authors for SCD studies (70%). Thirteen percent of studies (*n*= 5 GSs, *n*= 1 SCDs) utilized a mix of intervention providers, specifically a combination of research teams and classroom teachers, and classroom teachers and paraprofessionals.

Intervention dosage as it pertained to duration (i.e., total hours), frequency (i.e., total sessions), and intensity (i.e., minutes per session) revealed wide variability across the 45 studies. The total hours that interventions were implemented ranged from one hour to 210 hr with a median of 30 hr. Almost a third of studies (31%) implemented interventions for more than 30 hr (*n*= 13 GSs, *n*= 1 SCD). A little over a fifth of studies (22%) did not provide sufficient information to determine the total number of hours interventions were implemented.

Regarding frequency, interventions were delivered one day per week to daily, resulting in a range of three total sessions to 250 sessions, with a median of 40 sessions. Less than half of studies (47%) delivered interventions for more than 40 sessions (*n*= 20 GSs, *n*= 1 SCD). Approximately 18% of studies did not provide sufficient information to determine the total number of sessions students received vocabulary interventions.

Intervention sessions were conducted between 6 min and 90 min with a median of 30 min. Over 60% of studies provided more than 20 min of vocabulary interventions during one instructional session (*m* = 21 GSs, *m* = 8 SCDs). Approximately 13% of studies did not provide sufficient information to determine the number of minutes vocabulary interventions were implemented per session.

**Table 6**

*Descriptive Summary of Intervention Characteristics*

| Intervention Characteristics | Descriptors | *n* | % of Total | % of GS | % of SCD |
|---|---|---|---|---|---|
| Instructional programming | Explicit | 21 | 47% | 43% | 60% |
| | Implicit | 2 | 4% | 6% | |
| | Combined | 14 | 31% | 29% | 40% |
| | NR | 4 | 18% | 23% | |
| Target Word Domain | Basic/Functional | 2 | 4% | | 20% |
| | General academic | 12 | 27% | 26% | 30% |
| | Content-specific | 10 | 22% | 20% | 30% |
| | Mixed method | 9 | 20% | 20% | 20% |
| | NR | 12 | 27% | 34% | |
| Group Format | small group 5 or fewer | 11 | 24% | 17% | 50% |
| | Large group 6-10 | 4 | 9% | 11% | |
| | whole class | 20 | 44% | 57% | |
| | 1 on 1 | 3 | 7% | | 30% |
| | combination | 1 | 2% | 3% | |
| | self-administered | 1 | 2% | 3% | |
| | NR | 5 | 11% | 9% | 20% |
| Intervention Dosage Duration | 10 hr or less | 11 | 24% | 11% | 70% |
| | 11 to 30 hr | 10 | 22% | 26% | 10% |
| | More than 30 hr | 14 | 31% | 37% | 10% |
| | NR | 10 | 22% | 26% | 10% |
| Frequency | 5 sessions or fewer | 1 | 2% | 3% | |
| | 6 to 40 sessions | 15 | 33% | 20% | 80% |
| | more than 40 sessions | 21 | 47% | 57% | 10% |
| | NR | 8 | 18% | 20% | 10% |
| Intensity | 20 min or less | 10 | 22% | 22% | 20% |
| | more than 20 min | 29 | 64% | 64% | 80% |
| | NR | 6 | 13% | 13% | |

Note. *n*= sample size of studies. NR= not reported. hr= hour. min= minutes

***Outcome Measures***

There were 84 unique vocabulary outcome measures across the 45 studies. GSs contributed 65 unique measures, while SCD studies contributed 19 measures. Vocabulary outcome measures tended to be author-created (55% of GS outcomes, 100% of SCD outcomes). Of the 35 GSs, 26% ($n$= 9) used both author-created and standardized measures to examine vocabulary effects. It should be noted that two SCD studies used a combination of author-created and standardized measures to examine vocabulary effects (Helman, 2015; Helman et al., 2015). Scores from these standardized measures were pre- and post-test scores, but were inappropriate to meta-analyze with GS scores and SCD time-series data. As such, these measures were excluded and were not included in the total count of the 84 unique vocabulary measures.

Sixteen GSs (46% of the 35 GSs) included reading comprehension measures, yielding 22 effect sizes. Over half of measures were standardized measures (70%). Two studies used a combination of standardized and author-created measures (August et al., 2009; Burns, 2001). Only one SCD study included time-series data on reading comprehension (Alison et al., 2017). Due to the small sample size of SCD effects in this area, it was deemed inappropriate to meta-analyze with GS comprehension effects and thus, was not included in the total count of the 23 unique reading comprehension measures used for analysis. Wanzek and colleagues (2017) also included two reading comprehension measures, however, these outcomes were excluded since improving reading comprehension was the primary focus of the study and the integration of vocabulary instruction was intended as a supplemental component of the intervention. Including these two measures would bias outcomes since intervention components in the

Wanzek et al. study were designed to directly promote reading comprehension. For these reasons, the two reading comprehension measures from the Wanzek et al. study were excluded in the descriptive summaries and meta-analysis.

### *Methodological Characteristics*

To answer the research question, What is the overall quality of the research?, standards of practice in the field of education were used to examine methodological rigor. The quality of study methodology was informed by the CEC Quality Indicators (Cook et al., 2014) and WWC Standards for GSs and SCDs (USDOE, IES, WWC, 2017; Kratochwill et al., 2010).

For CEC Quality Indicators, studies were classified as met 80% of indicators (Met80) or did not meet 80% of indicators (DNM80). The cut-point of 80% was determined arbitrarily after observing that no study demonstrated evidence of meeting all quality indicators. The cut-point was created to provide a meaningful way to analyze data and to define rigor as demonstrating evidence on the majority of quality indicators. Using this classification, 11% of studies met 80% of quality indicators ($n$= 1 GS; $n$= 4 SCD). SCD studies often failed to meet quality standards because social validity data were not collected or studies did not provide sufficient information to understand participant characteristics. GSs often failed to meet quality standards because intervention adherence data were not collected using direct observational methods, reliability evidence for outcome measures were not reported, or studies did not provide sufficient information to understand the participant sample.

For WWC standards, studies were classified as met standards without reservations (Met), met standards with reservations (MWR), or does not meet standards (DNM).

Almost a third of studies (38%, *n*= 17) met WWC standards of practice, 22% designated as MWR, and 40% of studies did not meet WWC standards. Studies often did not meet WWC standards because those implementing SCDs did not provide evidence of collecting at least 20% of inter-observer data or missed to collect sufficient data for each phase (e.g., at least five data points for each phase). For GSs, most did not meet standards because attrition was high or baseline equivalence was not established for quasi-experimental designs.

Overall, descriptive data suggest that the majority of studies would not be considered rigorous in methodological design based on WWC and CEC standards of practice.

**Overall Estimate of Vocabulary Effects**

To answer the primary research question, To what extent are vocabulary programs and interventions effective in increasing vocabulary learning for English Learners?, effect sizes for vocabulary outcomes were meta-analyzed. Effect sizes were aggregated according to procedures outlined in Chapter 3. The computer software, Comprehensive Meta-Analysis (CMA, Version 3), was used to convert Cohen's *d* into Hedge's *g,* and to generate standard error and variance for GS outcomes. Online software programs for *BCTau* (Tarlow, 2016) and *BCES* (Pustejovsky et al., 2020) were used to support accurate calculations of effect sizes.

Effect sizes for GSs (i.e., Hedge's *g*), and SCDs (*BCTau* and *BCES*) were meta-analyzed separately, and in combination to answer the sub-research question: What is the average or overall effect of vocabulary instruction on vocabulary learning for ELs? The following sections discuss the overall effects for GSs, followed by SCD studies, and

conclude with overall effects when effects sizes from GSs and SCD studies were combined.

***Overall Effects of Group Studies***

Means and standard deviations for pre- and post-test scores for intervention and control groups were extracted for 80% of GS vocabulary outcomes. Almost a quarter of outcomes (24%) reported post-test only scores. The Kittley-Koshenina (2009) study was the only study to report change scores for intervention and control groups. Using formulas from Morris and Deshon (2002) as outlined in Chapter 3, the change scores were converted into a raw score metric, and then used to compute Cohen's *d*. Completing this conversion ensured that synthesized effect sizes were derived from the same scale.

A random effects model (REM) with robust variance estimation (RVE) of inversed weights was used to meta-analyze 65 vocabulary effect sizes across 35 studies. As noted in Chapter 3, $\rho$ was set to 0.8, resulting in an estimated effect of $g$= 0.36 ($SE$= .07, $CI_{95}$ = 0.23 - 0.50, $p$< .001) for group studies. Effect sizes ranged between -0.65 to 1.75. A sensitivity test was conducted for different values of $\rho$, however no differences in the estimate or $T^2$ were observed when $\rho$ was set to .2, .4, or .5.

Between-study heterogeneity was $T^2$= 0.11 with 70% of the variability representing systematic differences. Using suggested guidelines in interpreting $I^2$ (25%-small heterogeneity, 50%-medium heterogeneity, and 75%-large heterogeneity; Shadish & Haddock, 2009), $I^2$= 70% was substantial. This level of variability due to true differences fell in the medium range. As such, across the 35 GSs, vocabulary instruction

improved vocabulary learning for ELs in the intervention group by an average of 0.36 *SD* compared to peers in the control condition.

### *Overall Effects of Single Case Design Studies*

**Baseline Corrected Tau.** A web-based software was used to calculate *BCTau* (Tarlow, 2016). Data series from all 10 SCD studies were entered into the software. The software allowed data to be tested for baseline trend prior to calculating the final effect size. Baseline trend was not detected in any data set ($p > .05$) and thus, no adjustments were made to correct for trend. Four studies (Cannon et al., 2010; Guardino et al., 2014; Hinrichs, 2008; Lia, 2010) used multiple baseline designs across behaviors to examine vocabulary effects. For these studies, effect sizes were aggregated across behaviors for each case using the MAd package in R (Del re & Hoyt, 2018). The Borenstein et al. (2009) method was selected in pooling test statistics. The within-study correlation was set to .75, which was the mean correlation estimated from calculating the correlation of data-series for each of the four aforementioned studies. Aggregating cases using this method allowed for the calculation of an effect size and its variance for each case.

Using REM with RVE and inversed weights, 10 SCD studies with 75 vocabulary effect sizes were meta-analyzed. Rho was set to 0.8, which resulted in an estimated effect of *BCTau* = 0.72 (*SE* = .02, $CI_{95}$ = 0.67-0.77, $p < .001$). Effect sizes ranged between 0.33 to 0.94. A sensitivity test was conducted for different values of $\rho$, however no differences in the estimate or $T^2$ were observed when $\rho$ was set to .2, .4, or .5.

Results indicated that the *BCTau* estimated effect size of 0.72 was considered to be a moderate to large effect. Although the effect size was statistically significant at $p <$

.001 level, indicating that the average effect was different from zero, $I^2$ *and* $T^2$ both resulted in a zero-value. This was an unusual event in the use of REM and RVE with *BCTau*. A zero-value for $I^2$ *and* $T^2$ in the event of a statistically significant estimate is an area that will require further research since calculating effect sizes for SCDs for the purpose of meta-analyses is less developed than GSs (Hedges et al., 2013; Shadish & Hadock, 2009; Shadish et al, 2015). Nonetheless, *BCTau*=0.72 suggests that ELs made moderate to large gains in vocabulary learning compared to the counterfactual (i.e., baseline condition).

**Between Case Effect Size.** A web-based software was used to accurately calculate *BCES* (Pustejovsky et al., 2020). For the current study, *BCES* can only be calculated for multiple-baseline designs with three or more participants. Currently, there are no methods to calculate *BCES* for multiple-baseline designs across behaviors. Therefore, only 13 effect sizes across five studies (Alison et al., 2017; Anderson, 2014; Helman, 2015; Helman et al., 2015; Kim & Linan-Thompson 2013) were calculated and synthesized.

Data were uploaded onto the web-based platform for each study. The web-based platform's point-and-click format allowed users to make various model specifications in order to best fit data. In fitting data across all cases to support consistency in interpretation, a restricted maximum likelihood model was selected. To fit baseline data, I held level fixed and allowed slopes to vary randomly effects because the average outcome across studies could not be assumed to be zero. To fit data in intervention phases, fixed and random effects models were selected signifying variability observed

throughout cases, and linear trend was selected to model the general positive direction of intervention effects across cases.

Using REM with RVE and inversed weights, 13 vocabulary effect sizes across five SCD studies were meta-analyzed. Rho was set to 0.8, which resulted in an estimated effect of *BCES*= 6.92 (*SE*= 2.39, $CI_{95}$= -0.37-14.2, *p*=.057). Effect sizes ranged between 1.85 to 30.93. A sensitivity test was conducted for the different values of $\rho$, however differences in the estimate (.01 difference) and $T^2$ (.11 difference) were minor when $\rho$ was set to .2, .4, or .6.

It is worth noting that the between-study heterogeneity was $T^2 = 17.9$ and proportion of heterogeneity was $I^2 = 79.78$. Although the estimated effect size was not statistically significant, further illustrated by the confidence interval including zero, this level of heterogeneity would be considered large, although effects were not statistically significant. Results from the *BCES* dataset suggested that vocabulary interventions did not significantly improve vocabulary learning for ELs compared to the counterfactual (i.e., baseline condition).

*BCES* guidance on the magnitude of an effect considered to be meaningful or large is limited. A magnitude of *ES*= 2.0 has been proposed to be considered a large effect in the context of SCD studies (Jenson et al., 2007). Such guidance would suggest that an effect of 6.92 would be considered large, although still not statistically significant. Due to the limited number of studies and effects available for synthesizing, the nonsignificant finding may be because the sample was too small to have sufficient power to detect an effect (Cooper et al., 2009).

### *Overall Effects*

Effect sizes from GSs and SCD studies were combined to understand the overall effects of vocabulary instruction on vocabulary learning (see Table 7). Specifically, the 65 Hedge's *g* effect sizes from GSs ($n$= 35) and 13 *BCES* from SCD studies ($n$= 5) were combined, resulting in 78 effect sizes. REM with RVE and inversed weights was used to meta-analyze the 78 effect sizes across 40 studies. The overall model yielded a mean effect size of 0.40 ($SE$ = 0.07, $CI_{95}$ = 0.26 to 0.54, $p$< .001). Between-study heterogeneity was $T^2$ = 0.16 and proportion of heterogeneity was $I^2$ = 74.68. This indicated that 75% of the variability represented true differences and considered to be a large level of heterogeneity.

Results suggested that in general, vocabulary interventions and instruction improved vocabulary learning for ELs by 0.40 standard deviation compared to the counterfactual (i.e., EL peers in the control condition, or baseline condition). This effect was a moderate and meaningful effect.

### *Overall Effects of Reading Comprehension*

To answer the research question, To what extent are vocabulary instruction and interventions effective in increasing reading comprehension?, reading comprehension measures from GSs were meta-analyzed. Reading comprehension effect sizes were meta-analyzed similarly to vocabulary outcomes for GSs. Means and standard deviations for pre- and posttests were extracted for a majority of outcomes (80%), with the remaining outcomes reporting posttest-only scores. No studies used pretest posttest change scores, therefore, converting from a change score metric to a raw score metric was not required.

**Table 7**

*Vocabulary Effects and Methodological Rigor*

| Citation | Outcome Measure | *g* | *SE* | $CI_{95}$ L | $CI_{95}$U | WWC | CEC |
|---|---|---|---|---|---|---|---|
| **Group Studies** | | | | | | | |
| August et al. 2009 | Science Vocabulary | 0.29 | 0.08 | 0.12 | 0.45 | Meets | DNM80 |
| Avila & Sadoski 1996 | Cued Recall Test | 0.69 | 0.34 | 0.04 | 1.35 | DNM | DNM80 |
| | Sentence Completion | 1.06 | 0.35 | 0.38 | 1.75 | | |
| Benoit 2017 | Measure of Academic Vocabulary | 0.47 | 0.26 | -0.04 | 0.98 | DNM | DNM80 |
| Bravo & Cervetti 2014 | Science Vocabulary | 0.66 | 0.19 | 0.28 | 1.03 | Meets | DNM80 |
| Burns 2001 | GRADE-WM | -0.13 | 0.22 | -0.57 | 0.31 | DNM | DNM80 |
| | Depth of Knowledge | 0.20 | 0.22 | -0.24 | 0.64 | | |
| Cena et al. 2013 | Depth of Knowledge- Definition | 0.79 | 0.29 | 0.22 | 1.36 | Meets | DNM80 |
| | Depth of Knowledge- Total | 0.85 | 0.29 | 0.28 | 1.42 | | |
| | Depth of knowledge- Usage | 0.81 | 0.29 | 0.25 | 1.38 | | |
| | Test de Vocabulario en Imagenes Peabody | 0.13 | 0.28 | -0.41 | 0.68 | | |
| Cervetti et al. 2015 | Vocabulary Assessment | -0.13 | 0.17 | -0.47 | 0.20 | Meets | DNM80 |
| Crevecoeur et al. 2014 | Peabody Picture Vocabulary Test | 0.31 | 0.30 | -0.28 | 0.90 | MWR | DNM80 |
| | Target Word Knowledge | 1.18 | 0.32 | 0.55 | 1.81 | | |
| Crum 2017 | Social Studies Vocabulary Test | 0.24 | 0.20 | -0.15 | 0.63 | MWR | DNM80 |
| Cruz-Cruz 2005 | Vocabulary Test | 0.30 | 0.37 | -0.43 | 1.02 | DNM | DNM80 |
| Dack 1996 | Dack Vocabulary Assessment in Content Areas Battery- Science | 0.72 | 0.29 | 0.16 | 1.28 | DNM | DNM80 |
| | Dack Vocabulary Assessment in Content Areas Battery- Composite | 0.92 | 0.29 | 0.35 | 1.49 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Dack Vocabulary Assessment in Content Areas Battery- Social studies | 0.82 | 0.29 | 0.26 | 1.39 | | |
| Denton et al. 2008 | Peabody Picture Vocabulary Test | -0.65 | 0.33 | -1.29 | 0.00 | DNM | DNM80 |
| Frasco 2008 | Peabody Picture Vocabulary Test | 0.56 | 0.34 | -0.11 | 1.22 | DNM | DNM80 |
| Graves et al. 2011 (Study 1) | Vocabulary test | -0.11 | 0.26 | -0.62 | 0.41 | Meets | DNM80 |
| Graves et al. 2011 (Study 2) | Vocabulary test | -0.14 | 0.30 | -0.72 | 0.44 | Meets | DNM80 |
| Kieffer et al. 2012 | Nonword Morphological Derivation task | 0.22 | 0.11 | 0.01 | 0.44 | MWR | Met80 |
| | Real Word Morphological Decomposition | 0.06 | 0.11 | -0.16 | 0.27 | | |
| Kittley-Koshenina 2009 | Science vocabulary test | 0.15 | 0.53 | -0.89 | 1.19 | MWR | DNM80 |
| Lawrence et al. 2012 | Vocabulary Multiple choice | 0.03 | 0.21 | -0.38 | 0.45 | DNM | DNM80 |
| McBroom 2009 | Expressive Vocabulary Test | 0.28 | 0.50 | -0.70 | 1.27 | DNM | DNM80 |
| | Peabody Picture Vocabulary Test | 0.04 | 0.50 | -0.94 | 1.03 | | |
| | Word Context Vocabulary Test | 1.02 | 0.54 | -0.03 | 2.07 | | |
| | Word Knowledge Vocabulary Test | 1.72 | 0.60 | 0.55 | 2.88 | | |
| Mieure 2014 | Mastery Test | 1.29 | 0.26 | 0.77 | 1.81 | Meets | DNM80 |
| | Weekly Assessments | 0.92 | 0.25 | 0.42 | 1.41 | | |
| Nelson et al. 2011 | Root Word Vocabulary | 0.60 | 0.15 | 0.30 | 0.89 | Meets | DNM80 |
| | Woodcock Reading Mastery Test- Word comprehension | 0.03 | 0.15 | -0.26 | 0.32 | | |
| Neuman & Kaefer 2018 | Expressive One-Word Picture Vocabulary Test | 0.18 | 0.22 | -0.25 | 0.61 | MWR | DNM80 |
| | Science Vocabulary | 0.63 | 0.22 | 0.19 | 1.07 | | |
| | Peabody Picture Vocabulary Test | -0.03 | 0.22 | -0.46 | 0.40 | | |
| | Science Vocabulary | 0.93 | 0.23 | 0.48 | 1.38 | | |
| Proctor et al. 2011 | Gates MacGinitie Reading Achievement Test- Vocabulary Subtest | 0.01 | 0.18 | -0.35 | 0.37 | MWR | DNM80 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Vocabulary Breadth Test | 0.62 | 0.19 | 0.25 | 0.99 | | |
| | Vocabulary Depth Test- Caption | 1.26 | 0.20 | 0.87 | 1.65 | | |
| | Vocabulary Depth Test- Definition | 1.28 | 0.20 | 0.89 | 1.67 | | |
| | Vocabulary Depth Test- Total | 1.32 | 0.20 | 0.93 | 1.72 | | |
| Stevens 2018 | Post Unit Social Studies Vocabulary Test | 0.79 | 0.22 | 0.37 | 1.21 | DNM | DNM80 |
| Tong et al. 2014 (Study 1) | Woodcock Language Proficiency- Oral Vocabulary | 0.80 | 0.27 | 0.26 | 1.33 | Meets | DNM80 |
| Tong et al. 2014 (Study 2) | Woodcock Language Proficiency Battery Oral Vocabulary | 0.78 | 0.27 | 0.25 | 1.32 | Meets | DNM80 |
| | Woodcock Language Proficiency Battery- Verbal Analogies | 0.57 | 0.27 | 0.05 | 1.10 | | |
| Ulanoff & Pucci 1999 | Vocabulary Test | 0.59 | 0.33 | -0.05 | 1.23 | MWR | DNM80 |
| Vang 2004 | California Achievement Test- Vocabulary Subtest | -0.17 | 0.35 | -0.84 | 0.51 | DNM | DNM80 |
| Vaughn et al. 2006 (Study 1) | Woodcock Language Proficiency Battery- Picture Vocabulary (English) | 0.01 | 0.23 | -0.45 | 0.47 | Meets | DNM80 |
| | Woodcock Language Proficiency Battery- Picture Vocabulary (Spanish) | -0.01 | 0.22 | -0.45 | 0.43 | | |
| | Woodcock Language Proficiency Battery- Picture Vocabulary (English) | -0.03 | 0.23 | -0.47 | 0.41 | | |
| | Woodcock Language Proficiency Battery- Picture Vocabulary (Spanish) | 0.13 | 0.23 | -0.31 | 0.57 | | |
| Vaughn et al. 2006 (study 2) | Woodcock Language Proficiency Battery- Picture Vocabulary (English) | 0.15 | 0.21 | -0.26 | 0.57 | Meets | DNM80 |
| | Woodcock Language Proficiency Battery- Picture Vocabulary (Spanish) | -0.20 | 0.21 | -0.62 | 0.21 | | |
| | Woodcock Language Proficiency Battery- Verbal Analogies (English) | 0.11 | 0.21 | -0.30 | 0.52 | | |
| | Woodcock Language Proficiency Battery- Verbal Analogies (Spanish) | 0.25 | 0.21 | -0.16 | 0.66 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Vaughn et al. 2009 (Study 1) | Social Studies Vocabulary Test | 0.57 | 0.21 | 0.15 | 0.98 | Meets | DNM80 |
| Vaughn et al. 2009 (study 2) | Social Studies Vocabulary Test | 0.03 | 0.20 | -0.36 | 0.42 | Meets | DNM80 |
| Wanzek et al. 2017 | Gates MacGinitie Reading Achievement Test-Vocabulary Subtest | 0.18 | 0.27 | -0.34 | 0.70 | Meets | DNM80 |
| | Woodcock-Johnson Tests of Achievement-Picture Vocabulary | 0.10 | 0.27 | -0.43 | 0.62 | | |
| Weitz 2003 | Gates MacGinitie Reading Achievement Test-Vocabulary Subtest | 0.14 | 0.22 | -0.28 | 0.56 | DNM | DNM80 |
| Yang 2015 | STELLA Vocabulary Fluency | 1.75 | 0.29 | 1.18 | 2.33 | DNM | DNM80 |
| | Woodcock Language Proficiency Battery-Picture Vocabulary | 0.19 | 0.25 | -0.30 | 0.67 | | |
| **Single Case Design Studies** | | | | | | | |
| Alison et al. 2017 | WH pairings | 7.29 | 2.31 | 2.76 | 11.81 | Meets | DNM80 |
| Anderson 2014 | Academic Vocabulary Knowledge | 10.60 | 0.96 | 8.72 | 12.47 | MWR | Met80 |
| | Academic Vocabulary Generalization | 7.51 | 1.53 | 4.52 | 10.50 | | |
| | Culturally Relevant Probes | 25.24 | 7.79 | 9.97 | 40.51 | | |
| | Non-Culturally Relevant Probes | 30.93 | 7.83 | 15.59 | 46.28 | | |
| Helman 2015 | Strategy Use | 6.95 | 1.02 | 4.95 | 8.94 | Meets | DNM80 |
| | Strategy Knowledge | 6.88 | 1.85 | 3.26 | 10.50 | | |
| | CLUES Probes Generalization-Controlled | 1.85 | 1.82 | -1.72 | 5.41 | | |
| | CLUES Probes Generalization-Uncontrolled | 1.92 | 2.18 | -2.36 | 6.19 | | |
| Helman et al. 2015 | Strategy Use | 12.33 | 2.26 | 7.90 | 16.75 | DNM | Met80 |
| | Strategy Knowledge | 16.46 | 9.35 | -1.86 | 34.78 | | |
| Kim & Linan-Thompson 2013 | Receptive Vocabulary | 2.01 | 1.03 | 0.00 | 4.02 | DNM | DNM80 |
| | Expressive Vocabulary Test | 2.57 | 0.52 | 1.55 | 3.58 | | |

| | | | | |
|---|---|---|---|---|
| **Overall Estimate** | **0.40** | **0.07** | **0.26** | **0.54** |

Note. Overall estimate $I^2$= 74%, $T^2$= 16%, $p< .001$. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit. WWC= What Works Clearinghouse standards of practice; Meets= meets standards without reservation; MWR= meets standards with reservation; DNM= does not meet standards. CEC= Council for Exceptional Children standards of practice; DNM80= did not meet 80% of standards; Met80= met 80% of standards.

Using REM with RVE and inversed weights ($\rho = .80$), 22 effect sizes across 16 studies yielded an overall estimate of 0.26 ($SE$ = 0.09, $CI_{95}$ = 0.07 to 0.46, $p$= .01, see Table 8). Effect sizes ranged between -0.42 to 1.22. A sensitivity test was conducted for different values of $\rho$, however no differences in the estimate were observed, and only a 0.0001 difference was observed for $T^2$ when $\rho$ was set to .2, .4, or .5.

Between-study heterogeneity was $T^2$ = 0.08 and the proportion of heterogeneity was $I^2$= 67.11. This indicated that 67% of the variability represented systematic differences, which was considered a medium to large level of heterogeneity. These results suggested that vocabulary instruction promoted reading comprehension gains for ELs, and ELs in the intervention condition outperformed ELs in the control condition, on average, by 0.26 *SD*, a significant and meaningful difference (Lipsey & Wilson, 1993).

**Table 8**

*Group Study Comprehension Effects*

| Citation | Outcome | *g* | *SE* | *CI*$_{95}$ *L* | *CI*$_{95}$*U* |
|---|---|---|---|---|---|
| Proctor et al. 2011 | The Gates MacGinitie Reading Achievement Test- Comprehension Subtest | 0.01 | 0.18 | -0.35 | 0.37 |
| Denton et al. 2008 | Woodcock Johnson- Passage Comprehension | 0.04 | 0.32 | -0.59 | 0.66 |
| Vaughn et al. 2006 (Study 1) | Woodcock Language Proficiency Battery– Passage Comprehension (English) | 0.04 | 0.23 | -0.41 | 0.49 |
| | Woodcock Language Proficiency Battery– Passage Comprehension (Spanish) | 0.32 | 0.23 | -0.12 | 0.77 |
| Vaughn et al. 2006 (Study 2) | Woodcock Language Proficiency Battery– Passage Comprehension (English) | 0.13 | 0.21 | -0.28 | 0.54 |
| | Woodcock Language Proficiency Battery– Passage Comprehension (Spanish) | -0.04 | 0.21 | -0.45 | 0.37 |

| | | | | | |
|---|---|---|---|---|---|
| Vaughn et al. 2009 (Study 1) | Social Studies Comprehension Test | 0.71 | 0.21 | 0.29 | 1.12 |
| Vaughn et al. 2009 (Study 2) | Social Studies Comprehension Test | 0.69 | 0.21 | 0.29 | 1.09 |
| Frasco 2008 | Gray Oral Reading Tests (GORT) | -0.02 | 0.33 | -0.67 | 0.64 |
| Stevens 2018 | Social Studies Content Knowledge | 0.74 | 0.22 | 0.32 | 1.16 |
| | Social Studies Reading Comprehension | 0.35 | 0.21 | -0.07 | 0.76 |
| Weitz 2003 | The Gates MacGinitie Reading Achievement Test- Comprehension Subtest | 0.05 | 0.22 | -0.37 | 0.47 |
| Burns 2001 | Standford Achievement Test- Comprehension Subtest | 0.02 | 0.22 | -0.42 | 0.46 |
| | Stanford Achievement Test- Sentence Reading Subtest | -0.42 | 0.23 | -0.87 | 0.02 |
| Graves et al. 2011 (study 1) | Maze | 0.12 | 0.27 | -0.42 | 0.65 |
| Graves et al. 2011 (study 2) | Maze | 1.22 | 0.33 | 0.57 | 1.87 |
| | Woodcock Reading Mastery Test- Passage Comprehension | 0.50 | 0.29 | -0.06 | 1.07 |
| August et al. 2009 | Science Knowledge | 0.17 | 0.08 | 0.00 | 0.33 |
| Bravo & Cervetti 2014 | Science Reading | 0.49 | 0.19 | 0.12 | 0.86 |
| | Science Understanding | 0.63 | 0.19 | 0.26 | 1.00 |
| Cervetti et al. 2015 | Science Knowledge Assessment | -0.27 | 0.17 | -0.61 | 0.06 |
| Tong et al. 2015 | Woodcock Language Proficiency Battery– Passage Comprehension | 0.89 | 0.28 | 0.35 | 1.43 |
| **Overall Estimate** | | **0.26** | **0.09** | **0.07** | **0.46** |

Note. Overall estimate $I^2$= 67%, $T^2$= 8%, $p$= .01. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit.

## Moderator Analysis

To answer the research question, To what extent do methodological characteristics moderate study outcomes?, a series of moderator analyses using meta-regression were completed. A priori meta-regression moderator analyses were prioritized and identified (i.e., participant demographics, study design characteristics, outcome

characteristics, and methodological rigor; see Chapter 3) based on previous meta-analyses and systematic reviews (i.e., August & Shanahan, 2006; Baker et al., 2014; Elleman et al., 2009), which have suggested that these variables of interest differentially affect vocabulary learning. All studies and their respective effect sizes were included in the moderator analyses unless noted. Several studies did not provide sufficient information to code variables of interest, and the level of missing data was highly variable. The range of missing data consisted of 18% of studies missing to provide sufficient information on the type of vocabulary programming (e.g., explicit, incidental, combination) to 46% of studies missing to report on the number of words taught. It is important to understand whether the variability in missing data or insufficient data influences effect sizes, hence studies with missing data were coded as not reported (NR) and included in analyses.

In conducting meta-regressions for each moderators of interest, the vocabulary effect size was included as the dependent variable while the moderator was set as the independent variable. REM with RVE using inversed weights was used to support moderator analyses. Results are discussed in the following order: participant demographics (e.g., grade level groupings), study design characteristics (e.g., instructional programming, intervention provider, intervention dosage, target word domains), outcome characteristics (e.g., type of measurement production, type of vocabulary scale, taxonomy of outcome), and methodological rigor (i.e., standards of practice). See Tables 9-12 for a summary of moderator analyses.

### *Participant Demographics*

Grade level groups were categorized as of elementary (kindergarten-fifth grade), middle school ($6^{th}$-$8^{th}$ grade) and high school students ($9^{th}$-$10^{th}$ grade). In conducting the meta-regression, elementary was used as the intercept. Results indicated that there was a significant difference in effect as a function of grade level groupings. There was a positive and significant effect size for elementary students ($g$= .50, $CI_{95}$ = 0.32 – 0.67, $p$ = < .001), a negative but nonsignificant effect for middle school students ($g$= - 0.28, $CI_{95}$ = -0.58 – 0.02), and a negative and nonsignificant effect for high school students ($g$= -0.04, $CI_{95}$ = -11.83 – 11.75). In other words, middle school students appeared to perform 0.28 *SD* lower than elementary students, and high school students performed 0.04 *SD* lower than elementary students.

**Table 9**

*Relation Between Grade Level Groupings and Vocabulary Effects*

| | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $T^2$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|
| Elementary[a] | 57 | 0.50 | 0.08 | <.001 | 0.32 | 0.67 | 0.17 | 74.83 |
| Middle School | 14 | -0.28 | 0.15 | 0.07 | -0.58 | 0.02 | | |
| High School | 7 | -0.04 | 1.20 | 0.98 | -11.83 | 11.75 | | |

Note. a = variable used as the intercept for comparisons. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit.

### *Study Design Characteristics*

**Instructional Programming.** Moderator analyses were conducted to examine differences that may arise in the type of instructional programming that was delivered (e.g., explicit, incidental, combination), the interventionist (e.g., classroom teacher,

paraprofessional), intervention dosage (e.g., frequency, duration, intensity) and the type of words that were taught.

Meta-regression results for instructional programming as a moderator was statistically significant, indicating that effects were associated with the type of programming EL students received. Programs that implemented a combination of explicit and implicit vocabulary instruction (i.e., combination programs) was used as the intercept. Results indicated a significant and positive effect when programs implemented a combination of explicit and incidental strategies ($g$= .52, $p$ = .003, $CI_{95}$ = 0.32 – 0.67, $p$= .003). There was a negative and nonsignificant effect for programs that implemented only explicit instruction, those implementing only incidental strategies, and those that did not report sufficient information to determine instructional strategies (i.e., not reported or NR). Programs using only incidental instructional strategies produced outcomes that were more than half a standard deviation ($g$= -0.52 *SD*) lower than combination programs.

**Intervention Provider.** Interventions provided by research teams was used as the intercept in examining differences across intervention providers. Vocabulary effects appeared to be related to the moderator of intervention providers. Results indicated that all providers had a positive but nonsignificant effect. However, peers implementing vocabulary interventions had a positive and statistically significant effect ($g$= 18.27, $CI_{45}$= 17.71-18.82, $p$< .001). This finding should be interpreted with caution. Although four effect sizes contributed to the moderator analysis, all four effect sizes were derived from one study (Anderson, 2014). The study conducted by Anderson (2014) was also a SCD study in which calculated effect sizes (i.e., *BCES*) were larger than conventional effect sizes (i.e., *g*) and may skew differences since it was the only study in this category

of intervention providers. When effects from the Anderson (2014) study were removed, differences based on intervention providers were no longer statistically significant ($p$= .22, $CI_{95}$= -0.25-0.85).

**Intervention Dosage.** Understanding differences in dosage of instruction prompts a need to examine frequency, duration and intensity of the intervention (Marulis & Neuman, 2013). Using guidance from Marulis and Neuman (2013), frequency was examined based on the number of sessions students received. Total sessions were calculated by multiplying the number of days per week by the number of weeks interventions were implemented. The median total sessions conducted was 40 sessions (range = 3-250 sessions). The median number of sessions was used to create categories of 40 or fewer sessions implemented, more than 40 sessions implemented and NR. The intercept was 40 or fewer sessions, and results indicated that differences in effects were associated with intervention frequency. Specifically, programs implementing 40 or fewer sessions had a statistically significant and positive effect ($g$ = 0.40, $p$ <.05, $CI_{45}$ = 0.01-0.80). Programs implemented longer than 40 sessions had a negative, but nonsignificant effect. Programs that did not provide sufficient information to determine frequency of intervention had a positive but nonsignificant effect.

Duration of an intervention was conceptualized as the total hours an intervention was implemented. Total hours were calculated by multiplying the number of minutes the intervention was implemented per session by frequency, and dividing by 60. The median of total hours implemented was 30 hr (range = 1-210 hr) and used to create the categories, 30 hr or less, more than 30 hr and NR. The intercept was 30 hr or less, and meta-regression results indicated a significant difference in the effect size as a function of

duration. Interventions implemented for 30 hr or less had a positive and statistically significant effect ($g$ = .43, $CI_{94}$ = 0.17-0.68, $p$ = .003). Programs implemented for more than 30 hr had a smaller, but nonsignificant effect compared to programs implemented for 30 hr of less. Programs that did not provide sufficient information to determine duration had a positive, but nonsignificant effect.

Intensity was conceptualized as the number of minutes conducted for each session. Minutes conducted per session was based on information authors reported. The median value for minutes conducted per session was 30 min (*range*= 6-90 min) and was used to create the categories 20 min or less, more than 20 min, and NR. The intercept was the category 20 min or less. Results indicated that differences in effects were associated with intensity. Interventions conducted for 20 minutes had a positive, statistically significant effect ($g$ = .24, $CI_{95}$ = 0.02-0.47, $p$ = .04,). Programs conducted for more than 20 minutes and those that did not provide sufficient information to determine intervention intensity had a positive, but nonsignificant effect.

**Target Word Domain.** Given the various word lists that exist (e.g., Academic Word List, General Service List) and the different models to prioritize words to teach (e.g., Beck's Tiered words), it was important to examine whether intervention effects were affected by the types of target words chosen for instruction. In other words, do target word domains moderate effects? Target word domains were informed by research resulting in the following categories for analysis: general academic, content-specific, mixed-method and NR. The domain, basic/functional words was used during coding and discussed in the descriptive summaries section separately, however, given that the domain only represented one effect size (Alison et al., 2017), the domain was combined

with the general academic domain. Merging the two domains seemed appropriate since basic/functional words, such as words that are learned to communicate basic needs, do not differ in quality drastically from general academic words (i.e., words that are needed to access more complex concepts).

Meta-regression was conducted using the domain, academic content-specific as the intercept. Results indicated that target word domain was associated with differences in effects. Programs that taught content-specific words had a positive and statistically significant effect ($g$=0.48, $CI_{95}$= 0.06-0.89, $p$= .03). Programs that taught general academic words produced effects that were 0.12 *SD* higher than those teaching content-specific words, however, this difference was not statistically significant. Instructional programs that did not report the domain of words used for instruction and programs using mixed methods had negative, but nonsignificant effects.

**Table 10**

*Relation Between Study Characteristics and Vocabulary Effects*

| | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $T^2$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|
| Instructional Programming | | | | | | | | |
| Combination[a] | 26 | 0.52 | 0.13 | 0.003 | 0.23 | 0.81 | 0.18 | 76.06 |
| Explicit Instruction | 32 | -0.11 | 0.18 | 0.53 | -0.49 | 0.26 | | |
| Incidental Instruction | 2 | -0.52 | 0.20 | 0.17 | -1.79 | 0.76 | | |
| NR | 18 | -0.17 | 0.16 | 0.32 | -0.51 | 0.18 | | |
| Intervention Provider | | | | | | | | |
| Research Team[a] | 13 | 0.30 | 0.21 | 0.22 | -0.26 | 0.86 | 0.142 | 72.31 |
| Classroom Teacher | 35 | 0.01 | 0.22 | 0.96 | -0.53 | 0.55 | | |
| Combination of Providers | 16 | 0.28 | 0.28 | 0.34 | -0.36 | 0.92 | | |
| Peers | 4 | 18.27 | 0.21 | <.001 | 17.71 | 18.82 | | |
| Paraprofessional | 4 | 0.30 | 0.39 | 0.52 | -1.27 | 1.87 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Self-administered | 6 | 0.54 | 0.22 | 0.12 | -0.33 | 1.42 | | |
| Intervention Dosage | | | | | | | | |
| Frequency of Training | | | | | | | | |
| 40 or fewer sessions[a] | 24 | 0.40 | 0.17 | 0.05 | 0.01 | 0.80 | 0.16 | 74.77 |
| more than 40 sessions | 14 | -0.08 | 0.19 | 0.70 | -0.49 | 0.34 | | |
| NR | 40 | 0.22 | 0.22 | 0.32 | -0.24 | 0.69 | | |
| Duration | | | | | | | | |
| 30 hours or less[a] | 38 | 0.43 | 0.12 | 0.003 | 0.17 | 0.68 | 0.18 | 75.63 |
| More than 30 hours | 24 | -0.11 | 0.16 | 0.49 | -0.45 | 0.22 | | |
| NR | 16 | 0.09 | 0.18 | 0.62 | -0.29 | 0.48 | | |
| Intensity | | | | | | | | |
| 20 minutes or less[a] | 15 | 0.24 | 0.09 | 0.04 | 0.02 | 0.47 | 0.17 | 75.58 |
| more than 20 minutes | 51 | 0.18 | 0.13 | 0.20 | -0.11 | 0.47 | | |
| NR | 12 | 0.29 | 0.20 | 0.17 | -0.15 | 0.72 | | |
| Target Word Domain | | | | | | | | |
| Content-specific[a] | 15 | 0.48 | 0.17 | 0.03 | 0.06 | 0.89 | 0.19 | 75.64 |
| General academic | 28 | 0.12 | 0.22 | 0.59 | -0.35 | 0.60 | | |
| Insufficient information | 21 | -0.30 | 0.21 | 0.19 | -0.75 | 0.16 | | |
| Mixed method | 14 | -0.03 | 0.20 | 0.90 | -0.47 | 0.41 | | |

Note. a = variable used as the intercept for comparisons. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit. Not reported.

### *Outcome Characteristics*

Three moderators regarding the characteristics of vocabulary outcomes were examined. Specifically, how measures were produced (i.e., measurement production), the vocabulary scale the outcomes intended to assess (i.e., productive, expressive), and the vocabulary taxonomy (i.e., word learning strategies, word knowledge) the measure intended to assess were examined for differential effects.

**Measurement Production.** Past studies have reported that differential effects are observed on vocabulary measures that were author-created compared to standardized measures (Elleman et al., 2009; Marulis & Neuman, 2013). Using author-created measures as the intercept, meta-regression results indicated and confirmed that effects were statistically significantly associated with measurement production. Author-created measures had a positive and significant effect, producing larger effects ($g$= 0.52, $CI_{95}$= 0.32-0.72, $p$<.001) compared to standardized measures ($g$= -0.31, $CI_{95}$= -0.58- -0.04, $p$= .02), which had a negative and statistically significant effect.

**Vocabulary Scale.** Students learn and can demonstrate their understanding of words receptively or expressively. The vocabulary scale (i.e., expressive or receptive) was examined as a moderator to understand whether one scale produced larger effects compared to the other scale. Three categories were used for analysis: receptive measures, expressive measures, and outcome measures that probed both expressive and receptive knowledge (i.e., mixed-method). Using mixed-method as the intercept, meta-regression results were not statistically significant, indicating that vocabulary scale may not moderate effects ($p$>.05.). Mixed-method and productive outcomes had positive, but nonsignificant effects. Receptive measures had a negative, but nonsignificant effect.

**Vocabulary Taxonomy.** It was important to understand whether vocabulary measures designed to detect word knowledge skills would function differently from measures designed to detect word learning strategies. Three categories were created, word knowledge, word strategies, and NR, which was used as the intercept. Results indicated that vocabulary measures classified based on the vocabulary taxonomy of instruction did not appear to moderate effects (p> .05.). Given the small sample size of

NR (*k*= 2), removing the data and refitting the model still did not result in statistically significant effects.

**Table 11**

*Relation Between Outcome Characteristics and Vocabulary Effects*

| | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $T^2$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|
| Measurement Production | | | | | | | | |
| Author-Created[a] | 49 | 0.52 | 0.10 | <.001 | 0.32 | 0.72 | 0.16 | 74.19 |
| Standardized | 29 | -0.31 | 0.13 | 0.02 | -0.58 | -0.04 | | |
| Vocabulary Scale | | | | | | | | |
| Combination[a] | 2 | 0.46 | 0.19 | 0.24 | -1.89 | 2.81 | 0.18 | 74.80 |
| Expressive | 32 | 0.18 | 0.22 | 0.52 | -1.12 | 1.49 | | |
| Receptive | 44 | -0.17 | 0.20 | 0.54 | -1.85 | 1.51 | | |
| Taxonomy Scale | | | | | | | | |
| NR[a] | 2 | 0.17 | 0.126 | 0.41 | -1.44 | 1.78 | 0.19 | 75.46 |
| Word Knowledge | 46 | 0.189 | 0.153 | 0.41 | -1.11 | 1.49 | | |
| Word Learning | 30 | 0.421 | 0.18 | 0.18 | -0.60 | 1.44 | | |

Note. a = variable used as the intercept for comparisons. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit. NR= Not reported.

### *Methodological Rigor*

Methodological rigor has been shown to moderate effects with studies implementing more rigorous research designs producing different effects compared to studies lacking in rigor (Elleman et al., 2009). Two standards of practice were used to examine methodological rigor, those established by the WWC (Kratochwill et al., 2010; USDOE, IES, WWC, 2017), and those informed by the CEC quality indicators (Cook et al., 2014).

For the WWC standards of practice, three categories were used, met all standards (i.e., Met), met standards with reservations (i.e., MWR), and does not meet standards (DNM). DNM was used as the intercept. Results indicated a significant difference in

effects as a function of methodological rigor based on WWC standards. Studies deemed to be less rigorous (i.e., DNM) had a positive and statistically significantly effect ($g$= 0.42, $CI_{95}$= 0.12-0.72, $p$= .02). Studies with rigorous designs (i.e., Met) had a negative, but nonsignificant effect. Studies identified as MWR had a positive, but nonsignificant effect.

As for CEC standards of practice, no study met all quality indicators outlined. To create categories for analysis, 80% was arbitrarily selected to define a study as satisfying the majority of quality standards and considered to be methodically rigorous. Hence, two categories were created: met 80% or more of quality indicators (i.e., Met80), and did not meet 80% of quality indicators (i.e., DNM80). Results indicated that methodological rigor informed by CEC standards statistically significantly moderated effects. Studies with low rigor had a positive and statistically significant effect ($g$= 0.41, $CI_{95}$= 0.26-0.55, $p$< .001). Rigorous studies had a negative, but nonsignificant effect and tended to produce effects that were .10 *SD* smaller than less rigorous studies.

**Table 12**

*Relation Between Methodological Rigor and Vocabulary Effects*

| | | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $T^2$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|---|
| WWC | | | | | | | | | |
| | DNM[a] | 25 | 0.42 | 0.14 | 0.01 | 0.12 | 0.72 | 0.18 | 75.75 |
| | Met | 33 | -0.06 | 0.17 | 0.72 | -0.41 | 0.29 | | |
| | MWR | 20 | 0.06 | 0.18 | 0.75 | -0.34 | 0.46 | | |
| CEC | | | | | | | | | |
| | DNM80[a] | 70 | 0.41 | 0.07 | <.001 | 0.26 | 0.55 | 0.17 | 74.78 |
| | Met80 | 8 | -0.10 | 1.65 | 0.96 | -17.38 | 17.17 | | |

Note. a = variable used as the intercept for comparisons. $CI_{95}$ L= 95% confidence interval lower limit; $CI_{95}$ U= 95% confidence interval upper limit. DNM= Does not meet standards;

MWR= Meets with reservation. DNM80= Did not meet 80% of standards; Met80= Met 80% or more of standards.

# Chapter 5

## Discussion

It has been more than 10 years since the last comprehensive review on vocabulary interventions for ELs and the pursuit of a meta-analysis (August & Shanahan, 2006). The current meta-analysis was conducted to understand the effects of vocabulary instruction on vocabulary learning and reading comprehension for ELs. In particular, there was a need to understand and quantify the degree to which vocabulary instruction promoted vocabulary learning for ELs, and the effect of interventions on reading comprehension.

To answer research questions pertaining to the effectiveness of vocabulary instruction and interventions on vocabulary learning and reading comprehension, the current study examined outcomes for ELs in kindergarten to 12$^{th}$ grade (K-12). In synthesizing GSs and SCD studies, the random effects overall estimate was $g$= 0.40 ($CI_{94}$= 0.26-0.54, $p$< .001). This estimate indicates that when provided vocabulary instruction and interventions, the gains that ELs in the intervention condition made were, on average, 0.40 *SD* higher than the counterfactual (i.e., EL peers in the control condition, or baseline condition). This is considered a small to moderate effect with statistical and practical significance (Lipsey & Wilson, 1993). The overall effect of vocabulary instruction on reading comprehension was $g$= 0.26 ($CI_{95}$= 0.07-0.46, $p$= .01). This estimate was derived only from GSs and is considered a small but meaningful effect (Lipsey & Wilson, 1993). These results indicate that vocabulary interventions had an overall positive effect on vocabulary learning and reading comprehension for ELs.

One goal of meta-analyses is to compare effects to other reviews and syntheses in order to better understand effects in context (Cooper et al., 2009). Unfortunately, such a

comparison is difficult to complete. No quantitative synthesis exists in examining vocabulary effects for ELs. The handful of syntheses that focus on the general student population may provide some meaningful context for comparison of effects. However, these syntheses had a different scope of focus in regards to examining only incidental word learning (Fukkink & Glopper, 1998), examining effects on specific vocabulary instruction (Stahl & Fairbanks, 1986), and examining effects on passage-level comprehension (Elleman et al., 2009). Therefore, caution is advised in interpreting and comparing the overall effects of the current study to the aforementioned syntheses. For these reasons, effects will not be compared with depth.

Fukkink and Glopper (1998) focused on examining the effects of incidental vocabulary programs on vocabulary learning for elementary to high school students. The overall effect for vocabulary learning was *ES*= 0.43. Stahl and Fairbanks (1986) aimed to understand the impact of vocabulary instruction on vocabulary learning and reading comprehension of K-12 students, and identifying the most effective vocabulary program. Overall effects on global vocabulary measures was *ES*= 0.26 ($p < .01$) and reading comprehension was *ES*= 0.97 ($p < .01$). Elleman and colleagues (2009) conducted a synthesis focused on the effect of vocabulary interventions on passage-level comprehension for K-12 students whose primary language was English. The overall estimates of vocabulary instruction for vocabulary learning was *d*= 0.29 ($p < .01$) on standardized vocabulary measures and *d*= 0.79 on author-created measures ($p < .01$). The overall estimate for reading comprehension was *d*= 0.10 ($p = .08$).

From these syntheses, it appears that the overall effect of vocabulary instruction on vocabulary learning appears to have a small to moderate effect. Effects for reading

comprehension range from small to large in magnitude. The overall estimate for vocabulary learning ($g$= 0.40) and reading comprehension ($g$= 0.36) for the current study appears consistent with past vocabulary syntheses. Generally, these estimates provide evidence that vocabulary interventions do promote vocabulary learning and reading comprehension.

Another research question guiding the current study was to understand the overall quality of the research. The WWC and CEC standards of practice were used to examine methodological rigor. Overall, the current state of the quality of the research is best characterized as low rigor with only 38% of studies meeting WWC standards without reservation and 11% of studies meeting the majority of CEC standards. The characterization of low rigor is made despite the fact that more than half of studies in the sample were from peer-reviewed journals. Furthermore, the current review found that studies with less rigor appeared to produce larger effects compared to studies that were more methodologically sound.

Valentine (2009) makes a distinction between study quality (i.e., design factors that were not considered) and reporting quality (i.e., critical information that was not reported) for which components of both constructs are represented in WWC and CEC standards of practice. Given the complexities of ELs who are a heterogeneous group, further highlighted by the 27 different native languages reported for students in studies of the current sample, reporting quality and study quality are equally critical and important. Accurate descriptions of who study participants are and the context of their environment (e.g., urban, low SES, high school students) are necessary to interpret effects and understand the principal objective for whom an intervention is effective for and under

what conditions promote positive effects. In fact, studies in the current sample overwhelmingly reported that ELs were Hispanic students. Given the large representation of Hispanic students and the limited representation of ELs from other racial/ethnic backgrounds, it begs the question of whether current findings extend to ELs from non-Hispanic and non-Spanish-speaking backgrounds. Hence, reporting quality matters for interpreting effects in context. In terms of study quality, confidence that effects are reliable and real (i.e., not due to chance) hinges on and is conditional on the rigor of research methodology.

The observation of methodological rigor moderating effects is not a peculiar finding. Syntheses conducted in other research areas have also observed the association of methodological rigor and study effects (Camilli et al, 2010; Hattie et al., 1996; Maggin et al., 2017). The current finding of low rigor studies producing larger effects compared to high rigor studies can complicate the identification of effective vocabulary practices. Consumers of research would rightfully be reluctant in trusting intervention effects because the effects may be grounded in questionable evidence due to poor methodologies. Therefore, it is important that future research in the area of vocabulary and ELs address methodological rigor as a priority. Future research must be designed with sufficient rigor and report sufficient information in order to better advance the field. For journal publications related to ELs, perhaps special requests with editors may be needed so that researchers have sufficient space to report all necessary information.

A key point from examining the quality of the research is that meta-analyses should not universally exclude studies based on methodological rigor. The decision to exclude studies based methodological rigor should be driven by the research question

(Cooper et al., 2009). For the current study, it was important not only to understand the overall quality of the research, but also to understand how methodological rigor influences intervention effects. Due to the small sample size of the current study, statistical procedures were not conducted to control for methodological rigor. Future research may consider pursuing this endeavor as new research becomes available that results in a larger sample size and appropriately allows for statistical adjustments and modeling.

**Implications for Practice and Research**

Moderator analyses can inform practice and provide insight into variables that appear to differentially impact vocabulary effects. Moderator analyses are not recommended to be conducted with small sample sizes (Cooper et al., 2009; Pincus et al., 2011) and the heterogeneity of vocabulary research (Ellemen et al., 2009; NRP & NICHD, 2000; Wright & Cervetti, 2016). However, arguments can be made to explore moderators based on theory or previous findings. Several moderators were prioritized based on past syntheses for the current study to understand their effects on study outcomes. Findings indicated several statistically significant moderators (i.e., grade level group, intervention dosage, target word domain, and instructional programming).

Specifically, vocabulary interventions tended to produce higher effects for elementary students compared to older students (middle and high schoolers). Although the NRP found a similar trend in their review, it remains unclear if this is an artifact of limited studies focused on older students or if less effective strategies are used with older students. Approximately 30% of studies in the current study included older students, with only 7% of studies aimed at high school students. This level of disparity in available

research for different student age groups is not unique to research on the EL population and vocabulary research (August & Shanahan, 2006; NRP & NICHD, 2000). The struggles of learning English as a second language are life-long (August & Shanahan, 2006; Baker et al., 2014). There is a need for more research to be conducted with older students to understand the differential effects of greater gains for younger students compared to older students and to better promote older students' academic growth.

Regarding instructional practices, programs that used a combination of explicit and incidental vocabulary strategies appeared to produce higher effects. In addition, interventions that integrated the use of content-specific words also observed larger effects. These findings provide supporting evidence for vocabulary instructional models (Baker et al., 2014; Beck et al., 2013; NRP & NICHD, 2000) that call for the use of comprehensive strategies that teach specific word meanings and provide opportunities for students to learn and be exposed to words in various contexts. The findings also highlight the importance of teaching content-specific words for ELs, especially since ELs are confronted with the unique challenge of learning English while also acquiring content knowledge.

It is also worth highlighting that intervention dosage (i.e., frequency, intensity and duration) was a statistically significant moderator. Results suggest that vocabulary interventions do not need to be substantially long in duration (e.g., less than 30 hr total), high in frequency (e.g., 40 sessions or fewer) and high in intensity (e.g., 20 minutes or less per session) to produce meaningful and positive effects. This finding is consistent with other studies that have observed positive effects with modestly implemented vocabulary interventions (Marulis & Neuman, 2010; Marulis & Neuman, 2013). Given

that resources in schools are becoming scarcer as the needs of students grow, highly efficient interventions that result in meaningful gains are vital.

**Limitations**

The current study has several limitations. Results should be interpreted with caution since effects from GSs and SCDs were combined and meta-analyzed. Although syntheses have combined effects and used similar methods as those described in the current study (Hedges, Pustejovsky, & Shadish, 2013; Petersen-Brown et al., 2019; Zelinsky & Shadish, 2018), limited guidance has been developed to inform practice and interpretation of findings. Moreover, due to limited guidance of interpreting effects for *BCES*, the *BCES* dataset was not analyzed for outliers. More research is needed in this area to create guidance in interpreting the magnitude of *BCES* and how to appropriately address outliers.

Another limitation is, the current synthesis did not account for effects that reside in nested designs. Specifically, statistical adjustments were not conducted to account for correlated effects that may manifest in GSs that assign participants to conditions by clusters (e.g., school, classroom). As more research becomes available on combining GSs and SCD study effects, future syntheses of this topic may consider re-analyzing the current dataset or accounting for correlated effects of clusters with new datasets. At present time, results from the current study should be interpreted with caution.

**Future Considerations for Research and Practice**

To my knowledge, this is the first meta-analysis to examine the effects of vocabulary instruction on vocabulary learning and reading comprehension for ELs. In the grand scheme of meta-analytic research and reading research, a sample of 45 studies after

screening more than 1,000 articles, is considered a small sample size (Cooper et al., 2009). Some vocabulary researchers have cautioned the use of moderator analyses because too few studies exist to sufficiently represent each construct (Ellement et al., 2009; Wright & Cervetti, 2016). Once again, caution in interpreting current findings is emphasized.

Nevertheless, the current review underscores the fact that research on ELs is scarce, and that vocabulary research is highly variable. This should not discourage efforts on synthesizing the literature. As noted in the introduction of this study, meta-analyses not only support with generalizing findings and identifying evidence-based practices, meta-analyses also highlight research gaps. We cannot fully understand gaps in research or practice if attempts at quantitatively synthesizing the literature are never conducted.

As such, a gap highlighted from current findings is the debate regarding instructional approaches consisting of teaching word learning strategies or word knowledge. The NRP (NRP & NICHD, 2000), IES (Baker et al., 2014) and several leading scholars (Beck et al., 1982; Graves, Schneider, & Ringstaff, 2018; Nagy, 2005) have recommended the instruction of word learning strategies for vocabulary learning. Word learning strategies consist of instruction around using context clues, understanding word parts and using the dictionary (Graves et al., 2018). The essence of word learning strategies is to promote independent word learning (Graves et al., 2018), hence, providing students with skills to learn words beyond the classroom. Results from the current review suggest that vocabulary measures designed to assess word learning strategies seemed to produce larger effects compared to word knowledge, however, these differences were not statistically significant. Hence, there is limited evidence to suggest that there is an

advantage of implementing one strategy over another when providing vocabulary instruction to ELs. In a systematic review by Wright and Cervetti (2016), the authors also did not find supporting evidence noting the advantage of one instructional approach over the other. This is an area that needs further attention. It is unclear if the nonsignificant findings in the current review indicate that both instructional approaches are equally effective or important for vocabulary instruction. It would also be worthwhile to understand whether the instructional approaches produce variable effects as a function of students' age. For older students, given the demands of having to learn about multiple content areas simultaneously, it would be important to know whether teaching word learning strategies in upper grades produce substantial effects on student outcomes.

Another area requiring more research is methods used for target word selection. Results from the current study found that instructional programming focused on content-specific words were more likely to observe larger effects compared to other target word domains. The difference was statistically significant. This adds supporting evidence for which domain of words to teach for greater effects, however, it provides limited understanding of how to select words for instruction. In the current study, I did not code methods that original authors used for selecting target words for instruction. Future research should consider coding for this characteristic as the field remains at odds regarding which method to use when selecting words for instruction (Baker et al, 2014; Beck et al., 2013). Perhaps some people may suggest that teachers should rely on words highlighted in textbooks and curricula for instruction. In a recent curriculum review of the four most commonly used kindergarten reading curricula (Wright & Neuman, 2018), the authors found that methods used to identify vocabulary words were obscure.

Moreover, target words highlighted for instruction were not challenging, sophisticated, or of high utility (i.e., words central to understanding, words used in various contexts). The majority of target words selected in these curricula did not require direct instruction to learn their meanings. This finding indicates that teachers should not rely on textbooks or curricula when identifying important vocabulary words for instruction. The findings also call attention to the need to deliberately plan for vocabulary instruction. In order to support deliberate efforts of vocabulary instruction, more research is needed, and specifically, more research in the area of methods used to select target words for instruction. It is important to continue research in these areas so that recommendations made for designing instructional programs are supported by sound and rigorous evidence shown to produce positive student outcomes.

**Conclusion**

The current meta-analysis makes major contributions to the field. Perhaps, for the first time, the effects of vocabulary learning are quantified for ELs and are shown to be statistically significant and meaningful. Additionally, the review was comprehensive in nature, examining effects of published and unpublished studies and GSs and SCD studies. Furthermore, these findings examined how ELs in intervention conditions performed compared to like-peers or the counterfactual (i.e., baseline condition). Conventional standards seem to suggest that intervention effects for ELs be compared between ELs in intervention conditions and English-only (EO) peers in intervention conditions (e.g., Lesaux et al., 2010; Silverman et al., 2017) or control conditions (e.g., Cena et al., 2023; Crevecoeur et al., 2014). Although it is important to understand the extent to which the vocabulary gap closes between ELs and EOs, it is equally important to understand how

ELs in intervention conditions compare to their like-peers in control conditions; an achievement of the current synthesis.

The current study set out to understand the effects of vocabulary instruction on vocabulary learning and reading comprehension for ELs. These effects were considered to be of small to moderate magnitude, and meaningful and statistically significant. Results provided insight into effective instructional practices and direction for future research. It is important that research in this area continue to expand in order to continue supporting the academic growth of ELs and close the achievement gap between ELs and their English-only peers.

## References

*References marked with asterisks indicate studies included in the meta-analysis

Albers, C. A., Hoffman, A. J., & Lundahl, A. A. (2009). Journal coverage of issues related to English language learners across student-service professions. *School Psychology Review, 38(1), 121–134.* https://doi.org/10.1080/02796015.2009.12087853

*Alison, C., Root, J. R., Browder, D. M., & Wood, L. (2017). Technology-based shared story reading for students with Autism who are English-Language Learners. *Journal of Special Education Technology*, 32(2), 91. http://doi.org/10.1177/0162643417690606

*Anderson, A. L. (2014). *Effects of a culturally relevant peer-delivered computer-assisted intervention on academic vocabulary acquisition and generalization of Latino English learners with disabilities* [Doctoral dissertation, University of North Carolina]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1614202872?accountid=14586

Anderson, R. C., & Freebody, P. (1979). *Vocabulary knowledge* (Technical Report No. 136). Washington, DC: National Institute of Education.

Anderson, R. C., & Nagy, W. E. (1993). The vocabulary conundrum. *Center for the Study of Reading Technical Report; no. 570.*

Anthony, J. L., Solari, E. J., Williams, J. M., Schoger, K. D., Zhang, Z., Branum-Martin, L., & Francis, D. J. (2009). Development of bilingual phonological awareness in Spanish-speaking English language learners: The roles of vocabulary, letter

knowledge, and prior phonological awareness. *Scientific Studies of Reading*, 13(6), 535–564. https://doi.org/10.1080/10888430903034770

*Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, 46, 379-395. https://doi.org/10.1111/j.1467-1770.1996.tb01241.x

*August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2009). The impact of an instructional intervention on the science and language learning of middle grades English language learners. *Journal of Research on Educational Effectiveness*, 2(8), 345-376. https://doi.org/10.1080/19345740903217623

August, S. & Shanahan, T. (Eds). (2006). *Developing literacy in second-language learners*: *Report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). Teaching academic content and literacy to English Language Learners in elementary and middle School. What Works Clearinghouse. http://ies.ed.gov/ncee/wwc/publications_reviews.aspx.

Baker, S. K., Simmons, D.C., Kameenui, E.J. (1995). Vocabulary Acquisition: Synthesis of the Research. Technical Report No. 13. Retrieved from https://files.eric.ed.gov/fulltext/ED386860.pdf

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506–521. https://doi.org/10.1037/0022-0663.74.4.506

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77. https://doi.org/10.1037/bul0000130

*Benoit, J. M. (2017). The effect of game-based learning on vocabulary acquisition for middle school English language learners [Doctoral dissertation, Liberty University]. ProQuest Dissertations & Theses. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1889535768?accountid=14586

Bloch, M. H. (2014). Meta-analysis and moderator analysis: Can the field develop further? *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(2), 135–137. https://doi.org/10.1016/j.jaac.2013.12.001

Blom, E., & Paradis, J. (2013). Past tense production by English Second Language Learners with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 56(1), 281–294. https://doi.org/10.1044/1092-4388(2012/11-0112)

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, *34*(6), 3-15. https://doi.org/10.3102/0013189X034006003

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221-236).

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. https://doi.org/10.1002/jrsm.12

*Bravo, M. A., & Cervetti, G. N. (2014). Attending to the Language and Literacy Needs of English Learners in Science. *Equity & Excellence in Education*, 47(2), 230-245. https://doi.org/10.1080/10665684.2014.900418

Brossart, D. F., Laird, V. C., & Armstrong, T. W. (2018). Interpreting Kendall's Tau and Tau-U for single-case experimental designs. *Cogent Psychology*, 5(1), 1–26. https://doi.org/10.1080/23311908.2018.1518687

*Burns, D. A. (2011). *Examining the effect of an overt transition intervention on the reading development of at-risk English-language learners in first grade* [Doctoral dissertation, University of Oregon]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/884788286?accountid=14586

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, 39(1), 69–83. https://doi.org/10.1080/02796015.2010.12087791

Butler, S., Urrutia, K., Buenger, A., Gonzalez, N., Hunt, M., & Eisenhart, C. (2010). A review of the current research on vocabulary instruction. *National Reading Technical Assistance Center, RMC Research Corporation*.

Cain, K., & Oakhill, J. (2011). Matthew Effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, 44(5), 431–443. https://doi.org/10.1177/0022219411410042

Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*, 579-620.

*Cannon, J. E., Fredrick, L. D., & Easterbrooks, S. R. (2010). Vocabulary instruction through books read in American Sign Language for English-Language Learners with hearing loss. *Communication Disorders Quarterly*, 31(2), 98–112. https://doi.org/10.1177/1525740109332832

Carlo, M. S., August, D., Mclaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39(2), 188–215. https://doi.org/10.1598/RRQ.39.2.3

Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the" fail-safe" statistic in meta-analysis. *Educational and Psychological Measurement*, 50(2), 233-243. https://doi.org/10.1177/0013164490502001

*Cena, J., Baker, D. L., Kame'enui, E. J., Baker, S. K., Park, Y., & Smolkowski, K. (2013). The impact of a systematic and explicit vocabulary intervention in

Spanish with Spanish-speaking English learners in first grade. *Reading and Writing*, 26(8), 1289–1316. https://doi.org/10.1007/s11145-012-9419-y

*Cervetti, G. N., Kulikowich, J. M., & Bravo, M. A. (2015). The effects of educative curriculum materials on teachers' use of instructional strategies for English Language Learners in science and on student learning. *Contemporary Educational Psychology*, 40, 86-98. http://doi.org/10.1016/j.cedpsych.2014.10.005

Chaffee, R. K., Johnson, A. H., & Volpe, R. J. (2017). A meta-analysis of class-wide interventions for supporting student behavior. *School Psychology Review*, 46(2), 149–164. https://doi.org/10.17105/SPR-2017-0015.V46-2

Cho, E., Capin, P., Roberts, G., Roberts, G. J., & Vaughn, S. (2019). Examining sources and mechanisms of reading comprehension difficulties: Comparing English learners and non-English learners within the simple view of reading. *Journal of Educational Psychology, 111*(6), 982–1000. https://doi.org/10.1037/edu0000332

Collier, V. P. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal*, *16*(1-2), 187-212. https://doi.org/10.1080/15235882.1992.10162633

Comprehensive Meta-Analysis (Version 3) [Computer software]. Englewood, NJ: Biostat.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T., McWilliam, R., Tankersley, M., Test, D., & Council for Exceptional Children. (2014). Council for Exceptional Children Standards for evidence-based practices in special education. *Teaching Exceptional Children*, 46(4), 206–212. https://doi.org/10.1177/0040059914531389

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis (*2nd ed). Russell Sage Foundation.

Crevecoeur, Y. C., Coyne, M. D., & McCoach, D. B. (2014). English Language Learners and English-Only Learners' response to direct vocabulary instruction. *Reading and Writing Quarterly*, *30*(1), 51–78. http://doi.org/10.1080/10573569.2013.758943

*Crum, C. E. (2017). *Influence of technology on English Language Learners' vocabulary, reading, and comprehension* [Doctoral dissertation, Walden University]. ProQuest Dissertation & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1860921172?accountid=14586

* Cruz-Cruz, M. L. (2005). *The effects of selected music and songs on teaching grammar and vocabulary to second-grade English Language Learners* [Doctoral dissertation, Texas A&M University – Kingsville]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/305393216?accountid=14586

* Dack, G. H. (1996). Effects of a content-based curriculum on beginning middle school ESL students: An examination of gender differences and vocabulary in science and social studies [Doctoral dissertation, University of Houston]. ProQuest Dissertations & Theses.

http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/304303446?accountid=14586

Del Re, A. C. & Hoyt, W. T. (2014). MAd: Meta-analysis with mean differences. R package version 0.8-2. URL https://cran.r-project.org/package=MAd

*Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research & Practice,* 23(2), 79-89. https://doi:10.1111/j.1540-5826.2008.00266.x

Dietrich, S. M. (2008). *Effects of an explicit, systematic vocabulary intervention on first - grade English language learners* [Doctoral dissertation, Walden University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://www-proquest-com.ezp1.lib.umn.edu/docview/304391455?accountid=14586

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1-44. https://doi.org/10.1080/19345740802539200

Egger, M., G. Davey Smith, M. Schneider, and C. Minder. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. Retrieved from: https://arxiv.org/abs/1503.02220

*Frasco, R. D. (2008). *Effectiveness of reading first for English language learners: Comparison of two programs* [Doctoral dissertation, Walden University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/304396723?accountid=14586

Fry, R. (2007). How far behind in math and reading are English Language Learners? *Hispanic*, 1–32.

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of educational research*, *68*(4), 450-469.

Gardella, J. H., Fisher, B. W., & Teurbe-Tolon, A. R. (2017). A systematic review and meta-analysis of cyber-victimization and educational outcomes for adolescents. *Review of Educational Research*, 87(2), 283–308. https://doi.org/10.3102/0034654316689136

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences*. Routledge.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149–164. https://doi.org/10.1177/001440290507100202

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, 60(2), 183–208. https://doi.org/10.1007/s11881-010-0041-x

Goodwin, A. P., Huggins, A. C., Carlo, M., Malabonga, V., Kenyon, D., Louguit, M., & August, D. (2012). Development and validation of extract the base: An English Derivational Morphology Test for third through fifth grade monolingual students and Spanish-speaking English language learners. *Language Testing*, 29(2), 265–289. https://doi.org/10.1177/0265532211419827

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445

* Graves, A. W., Duesbery, L., Pyle, N. B., Brandon, R. R., & McIntosh, A. S. (2011). Two studies of tier II literacy development: Throwing sixth graders a lifeline. *The Elementary School Journal*, 111(4), 641-661. http://doi.org/10.1086/659036

*Green, L., Cearley, J., Stockholm, M., & Sheffield-Anderson, L. (2015). Direct vocabulary instruction with two 5th-grade English-Language Learners with language-learning disabilities: A treatment study. *Contemporary Issues in Communication Science and Disorders*, 42, 191–202. https://doi.org/10.1044/cicsd_42_F_191

*Guardino, C., Cannon, J. E., & Eberst, K. (2014). Building the evidence-base of effective reading strategies to use with deaf English-language learners. Communication Disorders Quarterly, 35(2), 59–73. https://doi.org/10.1177/1525740113506932

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175–182. https://doi.org/10.1038/nature25753

Hairrell, A., Rupley, W., & Simmons, D. (2011). The state of vocabulary research. Literacy Research and Instruction, 50(4). https://doi.org/10.1080/19388071.2010.514036

Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communications Monographs*, *58*(4), 437-448. https://doi.org/10.1016/S0149-2063(99)80074-9

Harrer, M., Cuijpers, P., Furukawa, T.A, & Ebert, D. D. (2019). *Doing Meta-Analysis in R: A Hands-on Guide*. https/doi.org/10.5281/zenodo.2551803

Harrington, M., & Velicer, W. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research*, 50(2), 162–183. https://doi.org/10.1080/00273171.2014.973989

Hart, C. (2018). Doing a literature review: Releasing the social science research imagination. Sage. https://doi.org/10.1080/09500790.2011.588012

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review of Educational Research*, *66*(2), 99–136. https://doi.org/10.3102/00346543066002099

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and related estimators. Journal of Educational Statistics, 6(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, 4(4), 324–341. https://doi.org/10.1002/jrsm.1086

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(2), 164–165. https://doi.org/10.1002/jrsm.17

*Helman, A. (2015). *The CLUES Strategy: Improving science vocabulary acquisition for Secondary English Language Learners with reading disabilities* [Doctoral dissertation, Lehigh University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1706911761?accountid=14586

*Helman, A. L., Calhoon, M. B., & Kern, L. (2015). Improving science vocabulary of high school English Language Learners with reading disabilities. *Learning Disability Quarterly*, 38(1), 40-52. http://doi.org/10.1177/0731948714539769

Hiebert, E. H., & Kamil, M. L. (2005). *Teaching and learning vocabulary: Bringing research to practice*. Routledge.

*Hinrichs, S. R. (2008). *An analysis of vocabulary and comprehension knowledge growth of first -grade English-Language Learners using an instructional package* [Doctoral dissertation, Northern Illinois University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/304542122?accountid=14586

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165–179. https://doi.org/10.1177/001440290507100203

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Bullock Mann, F., Barmer, A., and Dilig, R. (2020). The Condition of Education 2020 (NCES 2020-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020144

Institute of Education Sciences [IES] (n.d.). The regional educational laboratory program: About Us. Retrieved from https://ies.ed.gov/ncee/edlabs/about/

IPEK, H. (2009). Comparing and contrasting first and second language acquisition: Implications for language teachers. *English Language Teaching*, *2*(2), 155. https://doi.org/10.5539/elt.v2n2p155

Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta- analyses, and single subject designs. *Psychology in the Schools*, 44(*5*), 483–493. https://doi.org/10.1002/pits.20240

Joshi, R. M. (2005). Vocabulary: A critical component of comprehension. *Reading & Writing Quarterly*, 21(3), 209–219. https://doi.org/10.1080/10573560590949278

Kan, P. F., & Sadagopan, N. (2015). Speech practice effects on bilingual children's fast mapping performance. *Seminars in Speech and Language*, 36(2), 109–119. https://doi.org/10.1055/s-0035-1549106

Kieffer, M. J., & Box, C. D. F. (2013). Derivational morphological awareness, academic vocabulary, and reading comprehension in linguistically diverse sixth graders. *Learning and Individual Differences*, 24, 168–175. https://doi.org/10.1016/j.lindif.2012.12.017

*Kieffer, M. J., & Lesaux, N. K. (2012). Effects of academic language instruction on relational and syntactic aspects of morphological awareness for sixth graders from linguistically diverse backgrounds. *The Elementary School Journal*, 112(3), 519-545. https://doi:10.1086/663299

*Kim, W., & Linan-Thompson, S. (2013). The effects of self-regulation on science vocabulary acquisition of English Language Learners with learning difficulties. *Remedial and Special Education*, 34(4), 225–236. https://doi.org/10.1177/0741932513476956

*Kittley-Koshenina, C. W. (2009). The effects of video instruction on science vocabulary development of the English Language Learners in elementary education (Master's Thesis, University of Houston-Clear Lake). ProQuest Dissertations and Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/305156279?accountid=14586

Klingbeil, D. A., Renshaw, T. L., Willenbrink, J. B., Copek, R. A., Chan, K. T., Haddock, A., Yassine, J., & Clifton, J. (2017). Mindfulness-based interventions with youth: A comprehensive meta-analysis of group-design studies. *Journal of School Psychology*, 63, 77–103. https://doi.org/10.1016/j.jsp.2017.03.006

Krashen, S. D., & Terrell, T. D. (1995). Implications of second language acquisition theory for the classroom. In S.D. Krashen & T.D. Terrell, *The natural approach: Language acquisition in the classroom* (pp. 53-62). Prentice Hall Europe.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation. What Works Clearing House, June, 1–34. https://doi.org/10.1037/e578392011-004

Kuhn, M. R., & Stahl, S. A. (1998). Teaching children to learn word meanings from context: a synthesis and some questions. *Journal of Literacy Research*, 30(1), 119–138. https://doi.org/10.1080/10862969809547983

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, 24(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

*Lawrence, J. F., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. E. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition*, 15(3), 437-451. https://doi.org/10.1017/S1366728911000393

Ledford, J. R., & Gast, D. L. (2014). *Single case research methodology: Applications in special education and behavioral sciences*. Routledge.

Ledford, J. R., Lane, J. D., & Severini, K. E. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, 19(01), 4–17. https://doi.org/10.1017/BrImp.2017.16

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, 45(2), 196–228. https://doi.org/10.1598/RRQ.45.2.3

*Lia, M. P. (2010). *The effects of vocabulary instruction on the fluency and comprehension of fifth-grade nonnative English speakers* [Doctoral dissertation, Northern Illinois University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/527746876?accountid=14586

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, 13(1), 43-60. https://doi.org/10.1080/13670050902777546

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The ANNALS of the American Academy of Political and Social Science*, 587(1), 69–81. https://doi.org/10.1177/0002716202250791

Lipsey, M.W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (pp. 147-158).

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from a meta-analysis. *American Psychologist*, 48, 1181–1209. https://doi.org/10.1037/0003-066X.48.12.1181

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

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education*, 38(6), 353–370. https://doi.org/10.1177/0741932517716900

Manolov, R., & Moeyaert, M. (2017). How can single-case data be analyzed? software resources, tutorial, and reflections on analysis. *Behavior Modification*, 41(2), 179–228. https://doi.org/10.1177/0145445516664307

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. https://doi.org/10.3102/0034654310377087

Marulis, L.M. & Neuman, S.B. (2013) How vocabulary interventions affect young children at risk: A meta-analytic review, *Journal of Research on Educational Effectiveness*, 6:3, 223-262. https://doi.org/10.1080/19345747.2012.755591

*McBroom, D. B. (2009). Developing the expressive and productive academic language of Limited English Proficient learners [Doctoral dissertation, Walden University). ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/305090480?accountid=14586

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., and Bullock Mann, F. (2018). *The condition of education 2018* (NCES 2018-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2018/2018144.pdf

Microsoft Excel (Version 16.23) [computer software]. New Mexico: Microsoft.

*Mieure, D. B. (2014). *An exploratory study of purposeful and strategic communicative techniques to teach vocabulary from core reading programs to English Learners* [Doctoral dissertation, Utah State University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1658570091?accountid=14586

Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, 40(6), 874–900. https://doi.org/10.1177/0145445516645763

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105. https://doi.org/10.1037/1082-989X.7.1.105

Nagy, W.(2005). Why vocabulary instruction needs to be long-term and comprehensive. In E.H. Hiebert & M.L. Kamil (Eds), *Teaching and learning vocabulary: Bringing research to practice* (pp. 27-44). Lawrence Erlbaum Associates, Inc.

National Center for Education Statistics (NCES) & National Assessment of Educational Progress (NAEP). (2017). 2017 mathematics & reading assessments [Data file]. Retrieved from https://www.nationsreportcard.gov/reading_2017/nation/achievement?grade=4

National Clearinghouse for English Language Acquisition and Language Instruction Education Programs (NCELA). (n.d.). Fast facts. Retrieved from https://ncela.ed.gov/fast-facts

National Reading Panel (NRP) & National Institute of Child Health, & Human Development (NICHD). (2000). *Report of the national reading panel: Teaching children to read: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office. Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

* Nelson, J., Vadasy, P., & Sanders, E. (2011). Efficacy of a tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish- speaking English Learners. *Journal of Literacy Research*, 43(2), 184-211. https://doi.org/10.1177/1086296X11403088

* Neuman, S. B., & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology*, 52, 15-24. https://doi.org/10.1016/j.cedpsych.2017.12.001

Newman, M. G., Castonguay, L. G., Borkovec, T. D., Fisher, A. J., Boswell, J. F., Szkodny, L. E., & Nordberg, S. S. (2011). A randomized controlled trial of cognitive-behavioral therapy for generalized anxiety disorder with integrated techniques from emotion-focused and interpersonal therapies. *Journal of Consulting and Clinical Psychology*, 79(2), 171–181. https://doi.org/10.1037/a0022489

Odom, S. L., Barton, E. E., Reichow, B., Swaminathan, H., & Pustejovsky, J. E. (2018). Between-case standardized effect size analysis of single case designs: Examination of the two methods. *Research in Developmental Disabilities*, 79, 88–96. https://doi.org/10.1016/j.ridd.2018.05.009

O'Keeffe, B. V., Slocum, T. a., Burlingame, C., Snyder, K., & Bundock, K. (2012). Comparing results of systematic reviews: Parallel reviews of research on repeated reading. *Education and Treatment of Children*, 35(2), 333–366. https://doi.org/10.1353/etc.2012.0006

Orwin, R.G., & Vevea, J.L. (2009). Evaluating coding decisions. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (pp. 177-204).

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. Behavior Modification, 35(4), 303–322. https://doi.org/10.1177/0145445511399147

Petersen-Brown, S. M., Henze, E. E., Klingbeil, D. A., Reynolds, J. L., Weber, R. C., & Codding, R. S. (2019). The use of touch devices for enhancing academic achievement: A meta-analysis. *Psychology in the Schools*, 56(7), 1187-1206. https://doi.org/10.1002/pits.22225

Pigott, T. (2012). *Advances in meta-analysis*. Springer Science & Business Media.

Pincus, T., Miles, C., Froud, R., Underwood, M., Carnes, D., & Taylor, S. J. (2011). Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology*, *11*, 14. https://doi.org/10.1186/1471-2288-11-14

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 and Writing*, 24(5), 517–544. https://doi.org/10.1007/s11145-009-9218-2

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistic*s, 39(5), 368–393. https://doi.org/10.3102/1076998614547577

Qualtrics (2019) [Computer Software]. Retrieved from www.qualtrics.com

R Core Team (2013). R: A language and environment for statistical computing. Retrieved from: http://www.R-project.org

Rance-Roney, J. (2010). Jump-starting language and schema for English-Language Learners: Teacher-composed digital jumpstarts for academic reading. *Journal of Adolescent & Adult Literacy International Reading Association*, 53(5), 386–395. https://doi.org/10.1598/JA

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47(6), 427–469. https://doi.org/10.1016/j.jsp.2009.07.001

Rohatgi, A. (2018). WebPlotDigitizer (version 4.1). Retrieved from: http://arohatgi.info/WebPlotDigitizer

Rolstad, K. (2005). The Big Picture: A meta-analysis of program effectiveness research on English Language Learners. *Educational Policy*, 19(4), 572–594. https://doi.org/10.1177/0895904805278067

Scarborough, H.S. (2001). Connecting early language and literacy to later reding (dis)abilities: Evidence, theory, and practice. In Neuman, S., & Dickinson, D. (Eds). Handbook of early literacy research (pp. 97-110). Guilford Press.

Shadish, W.R. & Haddock, C.K. (2009). Combining estimates of effect size. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (pp. 257-278). Russell Sage Foundation.

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). The role of between-case effect size in conducting, interpreting, and summarizing single-case research (NCER 2015-002). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. http://eric.ed.gov/?id=ED562991

Shanahan T. & Beck, I. (2006). Effective literacy teaching for English-Language Learners. In S. August & T. Shanahan, (Eds), *Developing literacy in second-language learners*: *Report of the National Literacy Panel on language-minority children and youth* (pp. 415-488). Lawrence Erlbaum Associates, Inc.

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, 68, 651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

Silverman, R. D., Martin-Beltran, M., Peercy, M. M., Hartranft, A. M., McNeish, D. M., Artzi, L., & Nunn, S. (2017). Effects of a cross-age peer learning program on the vocabulary and comprehension of English Learners and non-English Learners in elementary school. *Elementary School Journal*, 117(3), 485–512. http://dx.doi.org/10.1086/690210

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English Language Learners. *Review of Educational Research*, 75(2), 247-284. https://doi.org/10.3102/00346543075002247

Snyder, E., Witmer, S. E., & Schmitt, H. (2017a). English Language Learners and reading instruction: A review of the literature. *Preventing School Failure*, 61(2), 136–145. https://doi.org/10.1080/1045988X.2016.1219301

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. E. (2007). *Teaching word meanings*. Routledge.

*Stevens, M. (2018). *Technology enhanced learning for English Language Learners* [Doctoral dissertation, George Mason University]. ProQuest Dissertations and Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/2203805608?accountid=14586

Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, *4*(2), 127-141. https://doi.org/10.1177/1536867X0400400204

Sutton, A.J. (2009). Publication bias. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis (*2nd ed, pp. 435-452). Russell Sage Foundation.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261–285. https://doi.org/10.3102/00346543069003261

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tarlow, K. R. (2016). Baseline Corrected Tau calculator. http://www.ktarlow.com/stats/tau

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline Corrected Tau. *Behavior Modification*, 41(4), 427–467. https://doi.org/10.1177/0145445516676750

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, August, 1–9. https://doi.org/10.1113/jphysiol.2004.078915

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375. https://doi.org/10.1037/met0000011

*Tong, F., Irby, B. J., Lara-Alecio, R., & Koch, J. (2014). Integrating literacy and science for English language learners: From learning-to-read to reading-to-learn. *The*

*Journal of Educational Research,* 107(5), 410-426. http://dx.doi.org/10.1080/00220671.2013.833072

*Ulanoff, S. H., & Pucci, S. L. (1999). Learning words from books: The effects of read-aloud on second language vocabulary acquisition. *Bilingual Research Journal*, 23(4), 409-422. https://doi.org/10.1080/15235882.1999.10162743

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017). What Works Clearinghouse: Standards Handbook (Version 4.0). *Washington, DC: Institute of Education Sciences, US Department of Education*. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks

Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15(6), 471–497. https://doi.org/10.1080/10888438.2010.501091

Valentine, J.C. (2009). Judging the quality of primary research. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 129-146). Russell Sage Foundation.

*Vang, M. (2004). *The effect of sustained silent reading on the reading vocabulary of second grade English Learners* (Master's thesis, California State University-Fresno). ProQuest Dissertations and Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/305035174?accountid=14586

Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single case research: Web based calculators for SCR analysis. (Version 2.0) [Web-based application].

College Station, TX: Texas A&M University. Retrieved from singlecaseresearch.org

*Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., ... & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, 43(3), 449-487. https://doi.org/10.3102/00028312043003449

Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities*, 45(3), 244–256. https://doi.org/10.1177/0022219412442157

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English Language Learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, 2(4), 297-324. https://doi.org/10.1080/19345740903167018

Viechtbauer W (2010). metafor: Meta-analysis package for R. R package version 1.4-0, Retrieved from http://CRAN.R-project.org/package=metafor.

*Wanzek, J., Petscher, Y., Otaiba, S. A., Rivas, B. K., Jones, F. G., Kent, S. C., Schatschneider, C., & Mehta, P. (2017). Effects of a year long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology, 109*(8), 1103–1119. https://doi.org/10.1037/edu0000184

*Weitz, W. E. (2003). *Sustained silent reading with non-native speakers of English: Its impact on reading comprehension, reading attitude, and language acquisition* [Doctoral dissertation, University of Southern California]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/305322233?accountid=14586

Wilson, D.B. (2009). Systematic coding. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (2nd ed., pp.159-176). Russell Sage Foundation.

Wilson, S. J., Dickinson, D. K., & Rowe, D. W. (2012). Impact of an early reading first program on the language and literacy achievement of children from diverse backgrounds. *Early Childhood Research Quarterly,* 28(3), 578–592. https://doi.org/10.1016/j.ecresq.2013.03.006

Wilson, S., Tanner-Smith, E. E., Lipsey, M. W., Steinka-Fry, K., & Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *The Campbell Collaboration*. https://doi.org/10.4073/csr.2011.8

Wood, W. & Eagly, A.H. (2009). Advantages of certainty and uncertainty. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 455-472). Russell Sage Foundation.

Wright, T., & Cervetti, G.N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly, 52*, 203-226. https://doi.org/10.1002/rrq.163

Wright, T., & Neuman, S. (2013). Vocabulary instruction in commonly used kindergarten core reading curricula. *The Elementary School Journal, 113*(3), 386-408. https://doi.org/10.1086/668766

Xiong, E.X. (2018). A literature review on the effects of vocabulary instruction for English Learners [Unpublished manuscript]. Educational Psychology Department, University of Minnesota.

*Yang, P.-L. (2015). *The impact of story reading and retelling on the oral development of English Language Learners* [Doctoral dissertation, Texas A&M University]. ProQuest Dissertations & Theses. http://login.ezproxy.lib.umn.edu/login?url=https://search.proquest.com/docview/1732349882?accountid=14586

Young-Suk Grace Kim (2017) Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER), scientific studies of reading, 21:4, 310-333. https://doi.org/10.1080/10888438.2017.1291643

Zelinsky, N. A. M., & Shadish, W. (2018). A demonstration of how to do a meta-analysis that combines single-case designs with between-groups experiments: The effects of choice making on challenging behaviors performed by people with disabilities. *Developmental Neurorehabilitation*, 21(4), 266–278. https://doi.org/10.3109/17518423.2015.1100690

Zimmerman, K. N., Pustejovsky, J. E., Ledford, J. R., Barton, E. E., Severini, K. E., & Lloyd, B. P. (2018). Single-case synthesis tools II: Comparing quantitative

outcome measures. *Research in Developmental Disabilities*, 79, 65–76. https://doi.org/10.1016/j.ridd.2018.02.001

Zotero (Version 5.0.64) [Computer software]. Retrieved from https://www.zotero.org/

## Appendix A: Search Strategy Results

| Database | Records |
|---|---|
| ASP | 22 |
| Education Source | 341 |
| ERIC | 103 |
| OSF | 61 |
| ProQuest Digital dissertations | 585 |
| PsycInfo | 34 |
| Total | 1146 |

**Retrieval Source: Academic Search Premier *(EBSCO)***

S1 TI ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR TI ( "Limited English Proficiency" or "LEP" ) OR TI ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR TI ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR TI ( Bilingual or Multilingual or "Linguistically diverse" ) OR TI ( "Dual language" or "Dual Language Learner" or "DLL" ) OR TI ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S2 SU ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR SU ( "Limited English Proficiency" or "LEP" ) OR SU ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR SU ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR SU ( Bilingual or Multilingual or "Linguistically diverse" ) OR SU ( "Dual language" or "Dual Language Learner" or "DLL" ) OR SU ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S3 AB ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR AB ( "Limited English Proficiency" or "LEP" ) OR AB ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR AB ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR AB ( Bilingual or Multilingual or "Linguistically diverse" ) OR AB ( "Dual language" or "Dual Language Learner" or "DLL" ) OR AB ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S4 S1 OR S2 OR S3

S5 TI ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR TI ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR TI ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S6 SU ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR SU ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR SU ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S7 AB ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR AB ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR AB ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S8 S5 OR S6 OR S7

S9 TI ( elementary or "high school" or "middle school" or school age ) OR SU ( elementary or "high school" or "middle school" or school age ) OR AB ( elementary or "high school" or "middle school" or school age )

S10 SU ( experiment* or quasi-experiment* ) OR AB ( experiment* or quasi-experiment* )

S11 SU ( "single-case design" or "single case design" or multiple baseline or alternating treatment or multiple probe or AB ) OR AB ( "single-case design" or "single case design" or multiple baseline or alternating treatment or multiple probe or AB )

S12 S10 OR S11

S13 (S10 OR S11) AND (S4 AND S8 AND S9 AND S12)

Number of records: 22
Timestamp: 04/12/19

**Retrieval Source: ERIC via EBSCO**

S1 TI ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR TI ( "Limited English Proficiency" or "LEP" ) OR TI ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR TI ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR TI ( Bilingual or Multilingual or "Linguistically diverse" ) OR TI ( "Dual language" or "Dual Language Learner" or "DLL" ) OR TI ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S2 AB ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR AB ( "Limited English Proficiency" or "LEP" ) OR AB ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR AB ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR AB ( Bilingual or Multilingual or "Linguistically diverse" ) OR AB ( "Dual language" or "Dual Language Learner" or "DLL" ) OR AB ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S3 SU ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR SU ( "Limited English Proficiency" or "LEP" ) OR SU ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR SU ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR SU ( Bilingual or Multilingual or "Linguistically diverse" ) OR SU ( "Dual language" or "Dual Language Learner" or "DLL" ) OR SU ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S4 TI (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR TI ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR TI ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S5 SU (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR SU ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR SU ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S6 AB (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR AB ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR AB ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S7 (S4 OR S5 OR S6)

S8 (S1 OR S2 OR S3)

S9 TI (elementary or "high school" or "middle school" or school age ) OR AB ( elementary or "high school" or "middle school" or school age ) OR SU ( elementary or "high school" or "middle school" or school age )

S10 S7 AND S8 AND S9

S11 AB (experiment* or quasi-experiment* ) OR SU ( experiment* or quasi-experiment* )

S12 AB ( "single-case design" or "single case design" or multiple baseline or alternating treatment or multiple probe or AB ) OR SU ( "single-case design" or

"single case design" or multiple baseline or alternating treatment or multiple probe or AB)
S13 (S11 OR S12)
S14 S10 AND S13

Number of records: 103
Timestamp: 04/12/19

**Retrieval Source: ProQuest Digital Dissertations**

1. SU “English Language Learner” or “ELL” or “English Learner” or “EL” or “English as a second language” or “ESL”
2. SU “Limited English Proficiency” or “LEP”
3. SU “Language minority” or “LM” or “Emergent bilingual” or “EB”
4. SU “Second Language Learner” or “Second language education” or “Second language acquisition”
5. SU Bilingual or Multilingual or “Linguistically diverse”
6. SU “Dual language” or “Dual Language Learner” or “DLL”
7. SU “Non-English speakers” or “English for speakers of other language” or “ESOL”
8. 1 or 2 or 3 or 4 or 5…

9. TI “English Language Learner” or “ELL” or “English Learner” or “EL” or “English as a second language” or “ESL”
10. TI “Limited English Proficiency” or “LEP”
11. TI “Language minority” or “LM” or “Emergent bilingual” or “EB”
12. TI “Second Language Learner” or “Second language education” or “Second language acquisition”
13. TI Bilingual or Multilingual or “Linguistically diverse”
14. TI “Dual language” or “Dual Language Learner” or “DLL”
15. TI “Non-English speakers” or “English for speakers of other language” or “ESOL”
16. 9 or 10 or 11 or 12…

17. AB “English Language Learner” or “ELL” or “English Learner” or “EL” or “English as a second language” or “ESL”
18. AB “Limited English Proficiency” or “LEP”
19. AB “Language minority” or “LM” or “Emergent bilingual” or “EB”
20. AB “Second Language Learner” or “Second language education” or “Second language acquisition”
21. AB Bilingual or Multilingual or “Linguistically diverse”
22. AB “Dual language” or “Dual Language Learner” or “DLL”
23. AB “Non-English speakers” or “English for speakers of other language” or “ESOL”
24. 17 or 18 or 19 or 20…

25. SU (vocabulary or “vocabulary development” or “vocabulary acquisition” or “vocabulary intervention” or “vocabulary instruction” or “vocabulary education” or “vocabulary teaching”)
26. SU (“vocabulary skills” or “vocabulary strategies” or “vocabulary building”)
27. SU (Morpholog* or “morphological awareness” or “word meaning” or “word knowledge” or “word learning” or “word consciousness” or “Word learning strateg*” or “vocabulary learning strateg*” or vocabulary instruction strateg*”)
28. 25 or 26 or 27

29. TI (vocabulary or “vocabulary development” or “vocabulary acquisition” or “vocabulary intervention” or “vocabulary instruction” or “vocabulary education” or “vocabulary teaching”)
30. TI (“vocabulary skills” or “vocabulary strategies” or “vocabulary building”)
31. TI (Morpholog* or “morphological awareness” or “word meaning” or “word knowledge” or “word learning” or “word consciousness” or “Word learning strateg*” or “vocabulary learning strateg*” or vocabulary instruction strateg*”)
32. 29 or 30 or 31

33. AB (vocabulary or “vocabulary development” or “vocabulary acquisition” or “vocabulary intervention” or “vocabulary instruction” or “vocabulary education” or “vocabulary teaching”)
34. AB (“vocabulary skills” or “vocabulary strategies” or “vocabulary building”)
35. AB (Morpholog* or “morphological awareness” or “word meaning” or “word knowledge” or “word learning” or “word consciousness” or “Word learning strateg*” or “vocabulary learning strateg*” or vocabulary instruction strateg*”)
36. 33 or 34 or 35

37. TI NOT “English as a Foreign Language” or “EFL”
38. SU NOT “English as a Foreign Language” or “EFL”
39. 37 or 38

40. TI NOT adults or “college students” or “post-secondary”
41. SU NOT adults or “college students” or “post-secondary”
42. 40 or 41

43. SU SU experiment* or quasi-experiment*
44. SU “single-case design” or “single case design”
45. 43 or 44

46. 8 AND 16 AND 24 AND 28 AND 32 AND 36 AND 39 AND 42 AND 45

***ProQuest Digital Dissertations Search Syntax***

((su("English Language Learner" OR "ELL" OR "English Learner" OR "EL" OR "English as a second language" OR "ESL") OR su("Limited English Proficiency" OR "LEP") OR su("Language minority" OR "LM" OR "Emergent bilingual" OR "EB") OR

su("Second Language Learner" OR "Second language education" OR "Second language acquisition") OR su(Bilingual OR Multilingual OR "Linguistically diverse") OR su("Dual language" OR "Dual Language Learner" OR "DLL") OR su("Non-English speakers" OR "English for speakers of other language" OR "ESOL")) OR (ti("English Language Learner" OR "ELL" OR "English Learner" OR "EL" OR "English as a second language" OR "ESL") OR ti("Limited English Proficiency" OR "LEP") OR

ti("Language minority" OR "LM" OR "Emergent bilingual" OR "EB") OR ti("Second Language Learner" OR "Second language education" OR "Second language acquisition") OR ti(Bilingual OR Multilingual OR "Linguistically diverse") OR ti("Dual language" OR "Dual Language Learner" OR "DLL") OR ti("Non-English speakers" OR "English for speakers of other language" OR "ESOL")) OR

(ab("English Language Learner" OR "ELL" OR "English Learner" OR "EL" OR "English as a second language" OR "ESL") OR ab("Limited English Proficiency" OR "LEP") OR ab("Language minority" OR "LM" OR "Emergent bilingual" OR "EB") OR ab("Second Language Learner" OR "Second language education" OR "Second language acquisition") OR ab(Bilingual OR Multilingual OR "Linguistically diverse") OR ab("Dual language" OR "Dual Language Learner" OR "DLL") OR ab("Non-English speakers" OR "English for speakers of other language" OR "ESOL"))) AND

((su(vocabulary OR "vocabulary development" OR "vocabulary acquisition" OR "vocabulary intervention" OR "vocabulary instruction" OR "vocabulary education" OR "vocabulary teaching") OR su("vocabulary skills" OR "vocabulary strategies" OR "vocabulary building") OR su(Morpholog* OR "morphological awareness" OR "word meaning" OR "word knowledge" OR "word learning" OR "word consciousness" OR "Word learning strateg*" OR "vocabulary learning strateg*" OR "vocabulary instruction strateg*")) OR (ti(vocabulary OR "vocabulary development" OR "vocabulary acquisition" OR "vocabulary intervention" OR "vocabulary instruction" OR "vocabulary education" OR "vocabulary teaching") OR ti("vocabulary skills" OR "vocabulary strategies" OR "vocabulary building") OR ti(Morpholog* OR "morphological awareness" OR "word meaning" OR "word knowledge" OR "word learning" OR "word consciousness" OR "Word learning strateg*" OR "vocabulary learning strateg*" OR "vocabulary instruction strateg*")) OR (ab(vocabulary OR "vocabulary development" OR "vocabulary acquisition" OR "vocabulary intervention" OR "vocabulary instruction" OR "vocabulary education" OR "vocabulary teaching") OR ab("vocabulary skills" OR "vocabulary strategies" OR "vocabulary building") OR ab(Morpholog* OR "morphological awareness" OR "word meaning" OR "word knowledge" OR "word learning" OR "word consciousness" OR "Word learning strateg*" OR "vocabulary learning strateg*" OR "vocabulary instruction strateg*"))) AND

(ab(elementary OR "high school" OR "middle school" OR school age) OR ti(elementary OR "high school" OR "middle school" OR school age) OR su(elementary OR "high school" OR "middle school" OR school age))

Number of records: 585
Timestamp: 04/11/19

**Retrieval Source: PsycInfo *(OVID)***

1 ("English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL").ab,sh,ti.

2 ("Limited English Proficiency" or "LEP").ab,sh,ti.

3 ("Language minority" or "LM" or "Emergent bilingual" or "EB").ab,sh,ti.

4 ("Second Language Learner" or "Second language education" or "Second language acquisition").ab,sh,ti.

5 (Bilingual or Multilingual or "Linguistically diverse").ab,sh,ti.

6 ("Dual language" or "Dual Language Learner" or "DLL").ab,sh,ti.

7 ("Non-English speakers" or "English for speakers of other language" or "ESOL").ab,sh,ti.

8 1 or 2 or 3 or 4 or 5 or 6 or 7

9 (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching").ab,sh,ti.

10 ("vocabulary skills" or "vocabulary strategies" or "vocabulary building").ab,sh,ti.

11 (Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*").ab,sh,ti. 33566

12 9 or 10 or 11

13 (elementary or "high school" or "middle school" or school age).ab,sh,ti.

14 (experiment* or quasi-experiment*).ab,sh.

15 ("single-case design" or "single case design" or multiple baseline or alternating treatment or multiple probe or AB).ab,sh.

16 14 or 15

17 8 and 12 and 13 and 16

Number of records: 34
Timestamp: 4/12/19

**Retrieval Source: Education Source**

S1 AB ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR AB ( "Limited English Proficiency" or "LEP" ) OR AB ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR AB ( "Second Language Learner" or "Second language

education" or "Second language acquisition" ) OR AB ( Bilingual or Multilingual or "Linguistically diverse" ) OR AB ( "Dual language" or "Dual Language Learner" or "DLL" ) OR AB ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S2 SU ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR SU ( "Limited English Proficiency" or "LEP" ) OR SU ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR SU ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR SU ( Bilingual or Multilingual or "Linguistically diverse" ) OR SU ( "Dual language" or "Dual Language Learner" or "DLL" ) OR SU ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S3 TI ( "English Language Learner" or "ELL" or "English Learner" or "EL" or "English as a second language" or "ESL" ) OR TI ( "Limited English Proficiency" or "LEP" ) OR TI ( "Language minority" or "LM" or "Emergent bilingual" or "EB" ) OR TI ( "Second Language Learner" or "Second language education" or "Second language acquisition" ) OR TI ( Bilingual or Multilingual or "Linguistically diverse" ) OR TI ( "Dual language" or "Dual Language Learner" or "DLL" ) OR TI ( "Non-English speakers" or "English for speakers of other language" or "ESOL" )

S4 S1 OR S2 OR S3

S5 AB ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR AB ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR AB ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S6 SU ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR SU ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR SU ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S7 TI ( vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" ) OR TI ( "vocabulary skills" or "vocabulary strategies" or "vocabulary building" ) OR TI ( Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strateg*" or "vocabulary learning strateg*" or "vocabulary instruction strateg*" )

S8 S5 OR S6 OR S7

S9 AB ( elementary or "high school" or "middle school" or school age ) OR SU ( elementary or "high school" or "middle school" or school age ) OR TI ( elementary or "high school" or "middle school" or school age )

S10 S4 AND S8 AND S9

Number of records: 341
Timestamp: 04/12/19

**Retrieval Source: OSF (Open Science Framework) Preprints**

***Search #1***

("English Language Learner" or "English Learner" or "English as a second language" or "Limited English Proficiency" or "Language minority" or "Emergent bilingual" or "Second Language Learner" or "Second language education" or "Second language acquisition" or Bilingual or Multilingual or "Linguistically diverse" or "Dual language" or "Dual Language Learner" or "Non-English speakers" or "English for speakers of other language") AND (vocabulary or "vocabulary development" or "vocabulary acquisition" or "vocabulary intervention" or "vocabulary instruction" or "vocabulary education" or "vocabulary teaching" or "vocabulary skills" or "vocabulary strategies" or "vocabulary building")

Number of records: 34
Timestamp: 04/11/19

***Search #2***

("English Language Learner" or "English Learner" or "English as a second language" or "Limited English Proficiency" or "Language minority" or "Emergent bilingual" or "Second Language Learner" or "Second language education" or "Second language acquisition" or Bilingual or Multilingual or "Linguistically diverse" or "Dual language" or "Dual Language Learner" or "Non-English speakers" or "English for speakers of other language") AND (Morpholog* or "morphological awareness" or "word meaning" or "word knowledge" or "word learning" or "word consciousness" or "Word learning strategy" or "vocabulary learning strategy" or "vocabulary instruction strategy")

Number of records: 27
Timestamp: 4/11/19

TOTAL OSF records: 61

## Appendix B: Abstract Screening Checklist

**Study ID _____** **Screener Date: _____**

**Full citation**:

**What is the source of the study?**

1. Digital research database: Select this option if the study was identified from ASP, ProQuest Digital Dissertations, PsycInfo, Education Resources Information Center (ERIC), and Education Source.
2. Internet search: Select this option if the study was identified from Google Scholar…
3. Reference list: Select this option if the study was identified from reference lists by August & Shanahan (2006), Baker et al. (2014), Snyder et al. (2017) and identified as part of Phase 4.
4. Researcher: Select this option if the study was submitted by a researcher via email.

**Criteria 1. Does this study include English Learner participants?**

**Yes**: Select *yes* if the study mentions English Learner, Limited English Proficiency, language minority, English Language Learner, English as a second language, Second Language Learners, bilingual, multi-lingual, and dual language learner

**No**: Select *no* if the study does not mention the above terms, and/or mentions English as a Foreign Language (EFL)

**Undermined**: Select *undetermined* if participant characteristics are not discussed in the abstract, or if it is unclear that participants are ELs.

**Criteria 2. Does this study include participants in grades kindergarten to 12$^{th}$ grade?**

**Yes**: Select *yes* if study mentions participants who are in grades K-12$^{th}$.

**No**: Select *no* if study mentions pre-kindergarten, adults, or college participants only.

**Undermined**: Select undetermined if a grade level is not mentioned or difficult to determine the grade level.

**Criteria 3. Was the study conducted in the U.S.?**

**Yes**: Select *yes* if study was conducted in the United States

**No**: Select *no* if study was conducted outside of the United States.

**Undetermined**: Select *undetermined* if information is insufficient to determine place of origin.

**Criteria 4. Does this study address vocabulary instruction?**

**Yes**: Select *yes* if the study mentions vocabulary, word learning, word knowledge, word meaning, morphology, and/or syntax.

**No**: Select *no* if the study focuses only on oral language development, phonemic awareness, phonics, fluency, and comprehension. Select no if the study focuses on test/assessment construction and validation.

**Undermined**: Select *undetermined* if information is insufficient to be certain vocabulary is part of the intervention.

*Studies that evaluate assessment tools will be excluded (Goodwin et al., 2012).*

*When **NO** is marked on any of the above criteria, the study is excluded from advancing to the next phases and excluded from the meta-analysis.*

## Appendix C: Methods Screening Checklist

**Study ID: _____** **Screener Date: _____**

**Coder initials: _____**

**Criteria 1a. Does this study include English Learner participants?**

**Yes**: Select *yes* if the study identifies participants as English Learner, Limited English Proficiency, language minority, English Language Learner, English as a second language, Second Language Learners, bilingual, multilingual, and dual language learner.

**No**: Select *no* if the study mentions English as a Foreign Language (EFL), or does not identify students as ELs in any of its variations.

**Criteria 1b. Does this study include students in grades kindergarten to 12th grade?**

**Yes**: Select *yes* if participants are in grades K-12th.

**No**: Select *no* if participants are only pre-kindergarten, adults, or college students.

**Criteria 1c. If the study includes BOTH pre-k and kindergarten participants, does it allow data to be disaggregated only for kindergarten scores?**

**Yes**: Select *yes* if study provides means, sd, or outcome statistics for kindergarten students.

**No**: Select *no* if study aggregates all data.

**Criteria 1d. If the study includes both high school and adult participants, does the study allow data to be disaggregated only for high school participants?**

**Yes**: Select *yes* if study provides means, sd, or outcome statistics for high school participants.

**No**: Select *no* if study aggregates all data.

**Criteria 2a. Was the study conducted in the U.S.?**

**Yes**: Select *yes* if study was conducted in the United States

**No**: Select *no* if study was conducted outside of the United States (e.g., Canada, China)

**Criteria 2b. Was the study reported in English?**

**Yes**: Select *yes* if study was reported in English.

**No**: Select *no* if study was reported in a language other than English (e.g., Spanish)

**Criteria 3a. Is vocabulary part of the intervention or instructional program?**

**Yes**: Select *yes* if the study mentions instruction on vocabulary, word learning, word awareness, word knowledge, word meaning, morphology, and syntax

**No**: Select *no* if the study focuses only on oral vocabulary, phonemic awareness, phonics, fluency, and comprehension. Select no if the study focuses on test/assessment construction.

*Studies focusing on teacher professional development or teacher instructional performance without incorporating student outcomes will be excluded (Rance-Roney, 2010).*

**Criteria 3b. Was vocabulary instruction manipulated or part of the independent variable?**

**Yes**: Select *yes* if vocabulary instruction is part of the independent variable that was manipulated or changed.

**No**: Select *no* if vocabulary instruction is not part of the independent variable, and/or observed.

**Criteria 4a. Is the study an experimental or quasi-experimental study?**

**Yes**: Select *yes* if the study employs single-case design (e.g., multiple baseline design, multiple probe design), randomized assignment, pre-/post-test design, and/or post-test only design.

**No**: Select *no* if the study is a case study, correlation design study, or qualitative study (e.g., ethnographic study).

**Criteria 4b. Are there at least two groups of participants? (Group Design Studies)**

**Yes**: Select *yes* if the study includes an intervention group and at least one control/comparison group.

**No**: Select *no* if the study includes only one group.

**Criteria 4c. Are there at least two conditions, subsequent comparison legs, or a control measure? (Single-case Design Studies)**

**Yes**: Select *yes* if the study includes a baseline phase, subsequent intervention legs, a reversal, and/or repeated measures of a control outcome measure.

**No**: Select *no* if the study includes only the intervention phase.

**Criteria 5a. Is there at least one vocabulary outcome measure?**

**Yes**: Select *yes* if at least one outcome assessment measures word learning, word knowledge, word meaning, word awareness, word mapping, word analysis, and word identification.

**No**: Select *no* if no outcome measure exists that assesses word learning, word knowledge, word meaning, morphology, and syntax. Select no if the outcome measure assesses oral fluency, oral language development or English language proficiency.

**Criteria 5b. If the outcome measure is a global reading measure, does the study provide disaggregated data to isolate vocabulary learning?**

**Yes**: Select *yes* if data are available to analyze the effects of vocabulary learning in insolation.

**No**: Select *no* if data are aggregated.

**Criteria 6a. Does the study provide disaggregated descriptive data and sufficient vocabulary outcome data to calculate effect sizes for ELs in the intervention condition?**

**Yes**: Select *yes* if study provides means, sd, or outcome statistics for EL participants.

**No**: Select *no* if study aggregates all data for the study sample.

**Criteria 6b. Does the study provide disaggregated descriptive data and sufficient vocabulary outcome data to calculate effect sizes for ELs in the control/comparison condition? (Not applicable for SCDs)**

**Yes**: Select *yes* if study provides means, sd, or outcome statistics for EL participants.

**No**: Select *no* if study aggregates all data for the study sample.

**Not applicable:** Select *not applicable* when the study employs a single case experimental design.

*When **NO** is marked on any of the above criteria for its corresponding research design (group and single-case), the study is excluded from advancing to the next phase or the meta-analysis.*

# Appendix D: Coding Manual

## Coding notes for each tab on Coding Spreadsheet

| Spreadsheet Tabs | Notes |
|---|---|
| **Study Information** | This tab is used to record general study information<br>If a manuscript includes more than 1 study, be sure to add a new row by identifying the study ID with a .1 or .2 to indicate the 1st study and 2nd study<br>Example: ASP3**.1**<br>ASP3**.2** |
| **HLang** | This tab is used to record information on languages other than Spanish or English spoken by the student sample.<br><br>Insert a new row for each unique language reported by the author<br>Example: ASP 1 Hmong<br>ASP 1 Tagalog<br>ASP 1 Russian |
| **Intervention Information** | This tab is used to record general intervention information. If a manuscript includes more than one study, be sure to add a new row by identifying the study ID with a .1 or .2 to indicate the 1st study and 2nd study |
| **Vocab Strategies** | This tab is used to record all unique vocab strategies reported by authors in the introduction or methods section, or previous publication.<br><br>Insert a new row for each vocabulary/reading strategy reported<br>When creating a new row, be sure to include the ID (study ID) for each vocab strategy<br>Example: ASP1 read aloud<br>ASP1 teaching of definitions<br>ASP1 modeling oral use of vocab words in a sentence |
| **DVvocab/DVcomp** | This tab is used to record information specific to dependent measures.<br><br>Insert a new row for each unique dependent variable; be sure to include the Study ID for each variable that is added<br><br>**Start all Vocab DVs** for each study with **V01** (label subsequent vocab DVs as **V02, V03**…)<br><br>**Start all Comprehension DVs** for each study with **C01** (label subsequent comprehension DVs as **C02, C03**…) |

| | |
|---|---|
| **VocabES/CompES** | This tab is used to record information pertaining to calculating effect sizes that relate to vocab and reading comprehension.<br><br>Insert a new row for each unique vocab or comprehension ES |
| **MethodC** | This tab is used to record information regarding quality indicators that each manuscript exhibits. QIs are adopted from both CEC and WWC.<br><br>Use coding handout to support with final WWC rating. |

**Coding Variables**

| **Study Information** | | |
|---|---|---|
| **Variable Name** | **Assigned Codes** | **Variable Label and Descriptions** (when applicable) |
| Study_ID | | **Study ID**<br><br>A unique identification number given to all studies. All Study IDs will begin with the corresponding search source followed by a numeric value (e.g., ASP01 [Academic Search Premier], ER28 [ERIC], GS35 [Google Scholar]). |
| Cite | | **Full citation** |
| AuthX | | **First Author (Last name, initials)**<br><br>Ex: Xiong, E.Z. |
| PubYear | | **Year of Publication** |
| Jname | NA = not applicable<br>Journal/Publication Name | **Journal/Publication Name**<br><br>NA is used for unpublished manuscripts |
| ManType | 1. Journal Article<br>2. Dissertation<br>3. Government/Technical Report<br>4. Unpublished manuscript | **Manuscript Type**<br><br>1. Journal Article- Manuscript published in a journal<br>2. Dissertation |

| | | |
|---|---|---|
| | 9. Cannot determine/Unknown | 3. Government/Technical Report- Indicated as a government or technical report (e.g., WWC, IES)<br>4. Unpublished manuscript- manuscripts provided by authors that are not dissertations<br>9. Cannot determine/Unknown |
| PubStat | 1. Peer-reviewed publication<br>2. Nonpeer-reviewed publication<br>3. Unpublished | **Publication Status**<br><br>1. Peer-reviewed publication: Study was published in a peer-reviewed or refereed journal<br>Examples:<br>- Journal of Applied School Psychology<br>- Journal of Early Childhood Literacy<br>- Learning Disability Quarterly<br>- Reading Research Quarterly<br>- Journal of Research on Educational Effectiveness<br>- The Reading Teacher<br>- TESOL Quarterly<br>2. Nonpeer-reviewed publication: Study was published in an outlet that is not peer-reviewed<br>Examples:<br>- Teaching Reading<br>- Language Contact and Bilingualism<br><br>3. Unpublished: Study has not been published in an outlet<br>Examples:<br>- Dissertation<br>- Technical report<br><br>Note: If you are not sure whether a journal is peer-reviewed or not, use UrlichsWeb. Sign into your UMN library account, and search for Urlichs web. Non-peer reviewed journals will appear without a referee icon, or if you click on the journal, the descriptor "refereed" will display as "No" |

| | | |
|---|---|---|
| | | http://ulrichsweb.serialssolution.com/ |
| CodeDate | | **Date of coding**<br><br>Date that coding was completed for the study<br>(Month/date/year) |
| CodeX | | **Coder initials**<br><br>The initials of the individual that completed the coding for the study. |
| Region | 9. Not reported | **Region**<br><br>States/regional areas are **coded based on the explicit reporting** from original authors.<br><br>9. Not reported: original authors do not identify a state or region in which the study took place.<br><br>*Consider converting (after coding) to NOAA mapping for consistency in regional identification.* |
| GeoArea | 1. Rural<br>2. Suburban<br>3. Urban/metropolitan<br>4. Mixed<br>9. Not reported | **Geographic area**<br><br>Geographic compositions are coded based on the explicit reporting from original authors.<br><br>9. Not reported: original authors do not identify a geographic area in which the study took place. |
| RAge | 99. Not reported | **Sample age range**<br><br>The age range of student sample as reported by author. |
| MAge | | **Sample mean age**<br><br>The mean age reported by author. |

| | | |
|---|---|---|
| | | Note: Ignore this variable if the author reports age range. |
| Grade | | **Student grade**<br><br>All grade levels represented in the sample (e.g., K,4 or 2-3, or 5, 7, 9) |
| GradeClass | 1. Elementary<br>2. Middle School<br>3. High School<br>4. Mix | **Grade classification**<br><br>1. Elementary: Students in kindergarten to 5th grade<br>2. Middle School: Students in 6th grade to 8th grade<br>3. High School: Students in 9th grade to 12th grade<br>4. Mix: Students are in grades that span more than one grade classification |
| HLang | 99. Not reported/insufficient information<br>1. Spanish<br>2. Other | **Languages other than English spoken by student sample**<br><br>1. Spanish: Students are reported to speak Spanish per original authors.<br>2. Other: When original authors report a language other than Spanish and English<br>99. Not reported/insufficient: Authors do not provide information about the sample's home language<br><br>Note: When studies are based on students enrolled in bilingual or native language transition programs, if the authors do not report home language, code as 99 (NR/insufficient information).<br><br>If Spanish AND another language is reported, code 2, and in the Olang tab list all of the languages reported including Spanish. |
| Disability | 9. Not Reported<br>0. No/excluded<br>1. Yes | **Included students with disabilities** |

| | | |
|---|---|---|
| | | Original authors reported including students with disabilities or provided demographic information on students with disabilities. |
| DisType | 99. Not reported<br>1. Autism Spectrum Disorder<br>2. Deaf/Hard of Hearing<br>3. Intellectual Disability<br>4. Learning Disability<br>5. Other<br>6. Other Health Impairment<br>7. Speech-language | **Disability Type**<br><br>The disability type that is reported by original authors. |
| NTot | | **Number of participants in total study sample**<br><br>The total number of participants at the beginning of the study per original author(s) reporting |
| nEL | NR-not reported | **Number of EL participants in study sample**<br><br>The total number of EL participants at the beginning of the study per original author(s) |
| nEO | NR-not reported | **Number of non-EL or English-only participants in the study sample**<br><br>nEO is calculated by this author by taking NTot – nEL. |
| PctEL | | **Percent of EL participants in study**<br><br>Calculated by dividing the number of EL participants (nEL) by the total number of study participants (NTot). This value will be calculated by this author. |
| RaceTot | NR-not reported | **Racial composition of overall sample**<br><br>% of African American/Black |

| | | |
|---|---|---|
| | | % of Asian American/Asian<br>% of Hispanic/Latinx<br>% of White<br>% of Native American/Alaska Native<br><br>NR = original author did not report racial composition |
| RaceEL | NR-not reported | **Racial composition of EL students**<br><br>% of African American/Black<br>% of Asian American/Asian<br>% of Hispanic/Latinx<br>% of White<br>% of Native American/Alaska Native<br><br>NR = original author did not report racial composition |
| BoysTot<br><br>GirlsTot | NR-not reported | **Gender representation of overall study sample**<br><br>BoysTot: Total percentage of boys in overall sample<br><br>GirlsTot: Total percentage of girls in overall sample<br><br>NR = original author did not report gender composition |
| FRL | 99. Not reported | **Percent of free/reduced lunch of overall sample**<br><br>Percent of students in the overall sample reported as receiving free/reduced lunch. (This variable is coded for the sample, not the population.)<br><br>Note: If multiple schools are included in a study, report the range of FRL. |
| ELidProc | 9. Not reported<br>1. Limited English proficiency test | **EL identification process**<br><br>The process original authors used to identify students as ELs |

| | | |
|---|---|---|
| | 2. Identified by school site<br>3. Other | 1. Limited English proficiency test: when original study used a test to identify students as ELs (e.g., language proficiency test)<br>2. Identified by school site: original authors report that students were identified as ELs by school/teacher or those enrolled in ELL/ESL classes<br>3. Other: original authors report a process other than a test or school/teacher identification to identify students as ELs |
| ELidtest | NR-Not Reported | **Name of the language proficiency test**<br><br>Code the name of the language proficiency test used to identify ELs as ELs (e.g., WIDA, OWL) |
| ELidCut<br><br>ELidCat | NR-Not Reported | **EL identification cut-point**<br><br>The cut-score values or limited proficiency categories original authors used to identify students as ELs.<br><br>ELidCut: the value of the cut-score original authors used to identify students as ELs<br><br>ELidCat: the proficiency category that original authors used to identify students as ELs |
| OLang | | **Languages other than English and Spanish spoken by student sample**<br><br>Name other languages author identified (e.g., Russian, Arabic, Hmong) |
| RdznCat | 1. Group<br>2. Single-case | **Research design category** |

| | | |
|---|---|---|
| | | 1. Group: When a study employs a research design with a large sample of participants to understand the average performance of groups.<br>2. Single-case: When a study employs a research design to 1 or a small sample of participants to understand individual performance. |
| Rdzn | 9. Not reported<br>1. Pre-test post-test group comparison<br>2. Post-test only group comparison<br>3. Other | **Research design (Group Design)**<br><br>1. Pre-test post-test group comparison: Research design conducts both pre-test and post-test assessments.<br>2. Post-test only group comparison: Research design conducts only post-test assessment.<br>3. Other: Indicate in the notes the experimental design used |
| Rdzn | 1. AB<br>2. Alternating<br>3. Changing criterion<br>4. Multiple baseline<br>5. Multiple probe<br>6. Parallel<br>Other | **Research design (Single-case Design)**<br><br>1. AB: Study that employs a baseline and intervention phase only.<br>2. Alternating treatments: Study employs rapid alternation of two or more interventions<br>3. Changing criterion: Study employs a baseline phase and repeatedly implements the intervention in a stepwise fashion using a pre-established criterion<br>4. Multiple baseline: Study that staggers the introduction of the intervention across a series of legs.<br>5. Multiple-probe: Study that staggers the introduction of the intervention across a series of legs and probes intermittently.<br>6. Parallel treatments: Study that rapidly alternates interventions with a time-lagged introduction of |

| | | |
|---|---|---|
| | | the intervention across a series of legs.<br>7. Other: All other SCDs (specify design in notes) |
| Asn | 9. Not reported<br>1. Random<br>2. Systematic (non-random)<br>3. Self-selection<br>4. Undetermined | **Mechanism of assignment to conditions (GS only)**<br><br>The process original authors used to assign participants to conditions.<br><br>1. Random: Assignment to the intervention or control condition employs a random process (regardless of the unit of assignment)<br>2. Systematic (non-random): Assignment to the intervention or control condition employs a predictable pattern or convenient process (regardless of the unit of assignment) decided upon by researchers or another authority (e.g., school administrators).<br>3. Self-selection: Assignment to intervention or control condition is based on volunteer selection by the participant.<br>4. Undetermined: There is insufficient information reported to determine the specific mechanism for assigning conditions. |
| UAsn | 9. Not reported<br>1. Student level<br>2. Teacher/classroom level<br>3. School site level | **Unit of assignment to condition (GS only)**<br><br>1. Student level: Assignment to conditions was conducted at the student level<br>2. Teacher/classroom level: Assignment to conditions was conducted at the teacher or classroom level |

| | | |
|---|---|---|
| | | 3. School site level: Assignment to conditions was conducted at the school level |

| **HLang** | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| **OLang** | | **List native languages other than Spanish of study sample**<br><br>Example: Hmong, Russian, Tagalog |

| **Intervention Information** | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| WordSel | 1. Basic/Functional<br>2. General academic<br>3. Content-specific<br>4. Mixed method<br>5. Insufficient information<br>99. Not applicable | **The category of vocabulary words used for instruction**<br><br>The category of vocabulary words indicated by original authors that can be determined from the introduction or methods section.<br><br>1. Basic/Functional: Words that are learned with little explicit instruction, communicate basic needs, or label everyday items/things<br>(e.g., happy [emotions], clock [common items], up [direction], dog, who)<br>2. General academic: Words drawn from the curriculum and/or used across multiple disciplines, appear frequently, are needed to access more complex topics, and typically have abstract meanings<br>(e.g., parallel, analyze, noun, fortunate).<br>3. Content-specific: Words that are specific to a particular academic |

| | | |
|---|---|---|
| | | discipline (e.g., science, math, language arts) tend to have technical definitions, and rarely found outside of the specific content area.<br>(e.g., condensation, hemoglobin, lava)<br>4. Mixed method: Uses a combination of any of the above descriptions *(indicate the combinations in notes area)*<br>5. Insufficient information to determine category<br>99. Not applicable: when studies exclusively use implicit strategies such as independent reading, shared book reading, or provide only teacher manual suggestions without identifying a specific target word list for students |
| TxProv | 9. Not reported<br>1. Author/research team<br>2. Classroom teacher<br>3. Licensed staff<br>4. Paraprofessionals<br>5. Self-administered<br>6. Mix<br>7. Other | **Intervention provider**<br><br>1. Author/research team: Original authors or graduate students implemented the intervention. (This includes researcher-hired staff who may be licensed teachers.)<br>2. Classroom teacher: Teachers of participating students implemented the intervention.<br>3. Licensed staff: A licensed staff other than the classroom teacher implemented the intervention.<br>4. Paraprofessionals: School aides (unlicensed) implemented the intervention.<br>5. Self-administered: The participant accessed the intervention and self-guided their progress without an intervention provider. This may include computer-delivered interventions or tape-recorded interventions. |

| | | |
|---|---|---|
| | | 6. Mix: A mix of providers defined in 1-5.<br>7. Other: A person other than individuals 2-4 implemented the intervention. |
| PrfDev | 0. Not reported/No<br>1. Yes | **Professional development provided by the research team for purposes of the study.**<br><br>1. Yes: Original authors report providing professional development or instructional guidance (at any point) throughout the study. |
| PDhrs | NR- not reported | **Professional development hours provided**<br><br>The number of hours authors provided on professional development for purposes of the intervention. |
| TxSet | NR = Not reported | **Instructional setting**<br><br>The group size that instruction was delivered. Record the group size or range based on the original author's reporting.<br><br>Code 1 for 1:1 settings<br><br>If authors provide descriptors of group sizes such as **small group or whole class without reporting a specific value**, use the descriptor that the authors used. |
| CType | 9. Not reported/insufficient information<br>1. No instruction/nothing<br>2. Business as usual | **Type of control/comparison condition**<br><br>1. No instruction/nothing: Baseline condition includes no instruction (typically in SCDs) |

| | | |
|---|---|---|
| | 3. Control outcome measure<br><br>99. Not applicable | 2. Business as usual: Control condition consisted of instruction or programming students would have received otherwise.<br>3. Control outcome measure: A control outcome measure is used (typically in SCDs).<br>Ex: When an intervention targets science vocabulary and a social studies vocabulary list is used as the control measure as part of an adaptive alternating treatments design<br>99. Not applicable: Experimental designs that do not require a baseline or control<br>Ex: Alternating treatments design don't require a baseline phase to be valid |
| InstrType | 9. Not reported<br>1. Explicit Instruction<br>2. Implicit Instruction<br>3. Combination of Explicit and Implicit | **Type of vocabulary instruction**<br><br>1. Explicit instruction: Instruction is provided on word meanings or strategies, and/or rules and external cues are provided to acquire word knowledge such that clear models and demonstrations are used<br>2. Implicit instruction: Students acquire word knowledge through exposure and reading and are expected to infer or derive meanings/concepts of words from the text without direct instruction<br>Ex: sustained silent reading, use of discussions, independent reading/wide reading<br>3. Combination of explicit and implicit<br>Ex: Instruction that includes teaching definitions and shared book reading. |
| PrType | 1. Depth | **Type of vocabulary program** |

| | | |
|---|---|---|
| | 2. Breadth<br>3. Combination | 1. Depth: Intervention program consisted of learning a limited/controlled set of words, and learning about words in various contexts. Instruction is often focused on learning about the different qualities of word knowledge, which may include learning about semantic relationships, collocations or syntactic patterning of words, their meaning, and related words<br><br>2. Breadth: Intervention program consisted of learning a large number of words and their meanings. These programs are focused on increasing vocabulary size such that students are learning many words. Instruction is usually focused on teaching only definitions without placing words in context, or having students read independently and/or read a wide selection of text.<br><br>*Code studies that use shared book reading, read aloud, independent reading, wide reading or sustained reading as a single instructional strategy or accompanied with teaching definitions as* ***breadth****.*<br><br>3. Combination: Intervention program consisted of a combination of depth and breadth. |
| Nwords | NR = not reported | **Total number of words taught**<br><br>The total number of words taught throughout the intervention as reported by original authors. |

| | | |
|---|---|---|
| DoseDay | | **Dosage (days)**<br><br>DoseDay: The number of days per week that the intervention was implemented |
| DoseMinSess | NR = not reported | **Dosage (minutes**)<br><br>DoseMin: The length of time in minutes that the intervention was implemented per session.<br><br>**Note:** Report either DoseMinSess or DoseMinT. Whichever that the author reports |
| DoseMinT | NR = not reported | **Dosage (total minutes)**<br><br>DoseMinT: the total number of minutes that participants received instruction throughout the study<br><br>**Note:** Report either DoseMinSess or DoseMinT. Whichever that the author reports |
| DoseWk | NR = not reported | **Dosage (duration)**<br><br>DoseWeek: The number of weeks the intervention was implemented |
| DoseTot | NR = not reported | **Dosage (total sessions)**<br><br>DoseTot: The total number of sessions the intervention was implemented<br><br>When the original author does not report the total number of sessions, this number will be calculated by this author given that DoseDay and DoseWeek were reported.<br><br>DoseTot = DoseDay * DoseWeek |
| AvgInteg | NR = not reported | **Mean percent of treatment integrity** |

| | | |
|---|---|---|
| | | **Report the range if a mean is not available**<br>**In the notes section, indicate when it is not reported in percentage (e.g., likert scale was used)** |
| FidelityObs | | **The proportion of sessions that treatment integrity/fidelity was collected**<br><br>Record the proportion of sessions (i.e., 0.40 for 40%) that authors collected treatment integrity/fidelity data. |
| TxLang | 1. English<br>2. Spanish<br>3. Combination of English and Spanish<br>4. Other | **Language of intervention instruction** |
| PctAttrEL | NR= Not reported | **Percent of overall participant attrition from baseline for ELs (GS only)**<br><br>Data is recorded based on the original authors' reporting of attrition rates for ELs from the beginning of to the end of the study. |
| PctAttrAll | NR= Not reported/insufficient information | **Percent of overall participant attrition from baseline for overall study sample (GS only)**<br><br>Data is recorded based on the original authors' reporting of attrition rates for overall study sample from the beginning to the end of the study.<br><br>Note: If a study does not mention attrition and quantitative and narrative data is insufficient to understand if attrition occurred, code NR. |

| Vocab Strategies | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| Vstrat | Example strategies<br><br>• Activates prior knowledge related to target words<br>• Connecting target words to familiar student experiences<br>• Embedded teacher-guide suggested adaptations<br>• Exposure to target words with various contexts<br>• Expressive use of the word between peer-to-peer<br>• Modeling pronunciation of target words<br>• Oral repetition of target word<br>• Picture pairing of target word<br>• Presented definitions of target words<br>• Providing corrective feedback<br>• Providing multiple exposure/repeated exposure of target words<br>• Providing Spanish translation<br>• Story book reading aloud by teacher/interventionist<br>• Student comprehension verification of meaning or understanding by teacher/interventionist<br>• Student-teacher co-charting word meanings<br>• Student-teacher co-construction of definitions<br>• Students orally use target word in a sentence<br>• Students write target word in a sentence | **Vocabulary strategy**<br><br>All unique vocabulary strategies explicitly reported by original authors are recorded (e.g., modeling, teaching definition, constructing sentences with target words). |

| | | |
|---|---|---|
| | • Teacher orally uses the word in a sentence<br>• Teaching morpheme analysis<br>• Teaching semantic analysis<br>• Use dictionary to teach target words<br>• Uses examples and nonexamples<br>• Uses games to teach target words<br>• Uses graphic organizer to teach target words<br>• Uses music/songs to teach target words<br>• Uses of cognates to teach target words<br>• Uses peer-tutoring to teach target words<br>• Uses polysemy to teach target words<br>• Uses student discussion focused on target words<br>• Uses synonyms/antonyms to teach target words<br>• Uses videos to teach target words | |

| **DV vocab/comp** | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| DVID | V--<br>Example: V01, V02<br><br>C--<br>Example: C01, C02 | **Dependent variable identifier**<br><br>A unique alpha-numeric code used to identify the dependent measure.<br><br>All identifiers for vocabulary measures will begin with the letter V.<br><br>All identifiers for comprehension measures will begin with the letter C. |

| | | |
|---|---|---|
| | | Note: Make sure to communicate with IOA partner to use similar labeling<br><br>Flag SCDs that use only pre and post vocab DVs. Will need to reconsider if these studies should be included as part of the meta. |
| pp | | **List the page number(s) that information is found regarding dependent variables** |
| DVname | | **Name of the outcome measure** |
| PartID | Example: 1, 2, 3 | **Participant identification (SCD only)**<br><br>All participants are identified with a numerically starting with the first participant as 1. |
| PartName | | **Name of the participant**<br><br>Use this variable to support with identification. Record the name of the participant if reported by the author, otherwise ignore. |
| NPhase | | **The total number of phase contrasts for each participant (SCD-only)**<br><br>The number of AB contrasts calculated for each time-series graph.<br><br>Example: For a multiple baseline design that consists of 3 participants, each participant has 1 AB contrast (a baseline phase [A] and intervention phase [B]) |
| MxProd | 9. Not reported<br>1. Commercial/standardized<br>2. Author created | **The process in which the measure was produced** |

| | | |
|---|---|---|
| | 3. Other | 1. Commercial/standardized: Copyrighted assessments published assessments.<br>2. Author created: Assessments created by the authors<br>3. Other: All other assessments |
| MxBroad | 0. No<br>1. Yes<br>99. Insufficient information | **Broad measure**<br><br>Yes: Vocabulary or comprehension outcome measures that typically include more than one construct/area, and are aimed at comprehensive performance (e.g., Gates Reading Achievement). |
| MxProxi | 0. No<br>1. Yes | **Proximal measure**<br><br>Vocabulary and comprehension measures will be classified as being a proximal measure (near-transfer) or not.<br><br>No: The measure is a distal measure that assesses generalized skills or skills and constructs that were not instructed on in the intervention.<br><br>Yes: The measure assesses specific skills or constructs taught in the intervention. |
| MxVscale | 1. Productive<br>2. Receptive<br>3. Mix<br>9. Insufficient information | **The type of vocabulary scale**<br><br>Vocabulary outcome measures will be classified into the following scale categories.<br><br>1. Productive/expressive: Measures that require students to use/produce/generate their knowledge of words, demonstrate their skills of words, describe attributes of a word (e.g., root |

| | | |
|---|---|---|
| | | word, parts of speech). Tasks often require students to speak or write what they know about words/concepts.<br>Ex: constructing sentences using the target word, writing out definitions in their own words<br>2. Receptive: Measures that require students to show their understanding by recognizing words, definitions, vocabulary skills or attributes of a word (e.g., root word, parts of speech). Tasks often require students to students to identify correct responses via reading (without the student having to independently produce the definition/meaning/concept)<br>Ex: matching a word to its definition, naming a picture<br>Sorting and matching words to their parts of speech categories<br>3. Mix: Measure required the student to produce their knowledge and to recognize concepts<br>9. Insufficient information: There is insufficient information provided by authors to determine vocab scale |
| MxVArea | 1. Context learning (WL)<br>2. Word knowledge (WK)<br>3. Word analysis (WL)<br>4. Word awareness (WL)<br>5. Word identification (WK)<br>6. Word mapping (WL)<br>7. Sentence construction (WL)<br>8. Strategy application (WL)<br>9. Word consciousness (WL)<br>10. Combination<br>99. Insufficient information | **Vocabulary subarea of outcome measure**<br><br>1. Context learning: Measures that prompt students to use context clues (i.e., known words around the unfamiliar word or information around the unfamiliar word) or utilize polysemy (i.e., words with many meanings).<br>2. Word knowledge: Measures that prompt students to produce a definition for a word or produce a word for a definition (orally or in written form). |

| | | |
|---|---|---|
| | | 3. Word analysis: Measures that prompt students to analyze whole words and/or parts of words (e.g., morphological derivation, identifying root words).<br>4. Word awareness: Measures that prompt students to use cognates (i.e., words in different languages that have a common origin or similar meanings)<br>5. Word identification: Measures that prompt students to recognize the meaning of words or match words to their definitions (e.g., multiple-choice). This also includes labeling (written or speech) a picture, or matching a word to a picture.<br>6. Word mapping: Measures that prompt students to identity words/concepts that share attributes with the target words (e.g., word association tasks, analogies).<br>7. Sentence construction: Measures that prompt students to use target words in a sentence.<br>8. Strategy application: Measures that prompt students to describe orally or in writing the steps needed/strategies to learn a word or acquire the meaning of a word<br>9. Word consciousness: Measures that inquire about attitudes toward words and learning about words<br>10. Combination: The measure integrated a combination of vocabulary subareas. Indicate the combination in the notes.<br>99. Insufficient information: The author does not provide enough information and public information is not available to determine which vocabulary subarea is targeted. Public |

| | | |
|---|---|---|
| | | information can include assessment websites (e.g., PPVT, Stanford Achievement) |
| ORelT | | **Reliability type of technical adequacy reporting**<br><br>Reliability may include but are not limited to coefficient stability, coefficient equivalence (alternate form), internal consistency, or criterion reliability. Indicate the specific reliability metric used. |
| OMxRel | | **Mean of previously established reliability properties of the outcome measure reported by original authors** |
| OMxIOA | NR= Not reported | **Mean of reliability IOA score from current study of the outcome measure reported by original authors** |
| UnitMx | 1. Percent<br>2. Points<br>3. # of words read correct<br>4. Standardized score<br>5. Grade Equivalence<br>6. Other | **Unit of measurement**<br><br>Indicate the unit of measurement for each score/dependent measure.<br><br>1. Percent<br>2. Points<br>3. # of words<br>4. Standardized score (e.g., z-score, standard score)<br>5. Grade Equivalence<br>6. Other (indicate in notes the unit of measurement) |
| DVTime | 1. Pre-test Only<br>2. Post-test Only<br>3. Pre and post-test<br>4. Follow-up post intervention only<br>5. Other | **Dependent variable time of administration**<br><br>Record the timeframe in which each dependent measure was administered. |

| | | |
|---|---|---|
| | 99. Insufficient information | 1. Pre-test Only: prior to implementing the intervention<br>2. Post-test Only: at the conclusion of the intervention<br>3. Pre and post-test: administered prior to implementing the intervention and at the conclusion of the intervention<br>4. Follow-up post intervention only: a delayed administration after the intervention concluded<br>5. Other: any timing not listed above<br>99. Insufficient information to determine<br><br>Note: If a measurement is not a true pre-test, such that it was administered after the start of an intervention, make sure to indicate that in the notes. |

| **Vocab/Comp ES** | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| DVID | V--<br>Example: V01, V02<br><br>C--<br>Example: C01, C02 | **Dependent variable identifier**<br><br>A unique alpha-numeric code used to identify the dependent measure.<br><br>All identifiers for vocabulary measures will begin with the letter V.<br><br>All identifiers for comprehension measures will begin with the letter C.<br><br>*This should be the same DV ID (as above section) used to describe the dependent variables.* |
| ESID | Example: 01, 02, 03, 04 | **Effect size identification** |

| | | |
|---|---|---|
| | | All relevant effect sizes will be labeled numerically beginning with 01 as the sequence |
| pp | | **List page number(s) of statistics pertaining to ES** |
| nTxPre | | **Intervention group size (n) at the start of study of EL subsample** |
| nTxpost | | **Intervention group size (n) at the end of study of EL subsample** |
| nCPre | | **Control group size (n) at the start of study of EL subsample** |
| nCPost | | **Control group size (n) at the end of study of EL subsample** |
| MonPost | NR= Not reported | **Timeframe posttests were measured at the conclusion of the intervention (record in months)**<br><br>MonPost: (e.g., 3 months); enter 0 if post-test is administered immediately after the intervention |
| PreTxM | | **Intervention group mean at pre-test of EL subsample** |
| PostTxM | | **Intervention group mean at post-test of EL subsample** |
| PreCM | | **Control group mean at pre-test of EL subsample** |
| PostCM | | **Control group mean at post-test of EL subsample** |
| PreTxSD | | **Intervention group standard deviation at pre-test of EL subsample** |

| | | |
|---|---|---|
| PostTxSD | | **Intervention group standard deviation at post-test of EL subsample** |
| PreCSD | | **Control group standard deviation at pre-test of EL subsample** |
| PostCSD | | **Control group standard deviation at post-test of EL subsample** |
| AdjM | 0. No<br>1. Yes | **Means were adjusted of EL subsample**<br><br>The original authors report that the means of EL subsample were adjusted |
| ESdir | 1. Positive<br>2. Negative<br>3. Inconclusive | **Direction of effect of EL subsample**<br><br>The direction in which intervention effects (calculated by this author) favored the treatment group.<br><br>1. Positive: When results favored the intervention group<br>2. Negative: When results favored the control group<br>3. Inconclusive: When no difference in intervention effects were revealed, or when effect size = 0<br><br>*Ignore during coding* |
| OEqual | 0. No<br>1. Yes<br>99. Undetermined | **Groups tested for equivalence of EL subsample**<br><br>Groups were tested for equivalence as reported by the original study |
| OEStype | 1. ES<br>2. Cohen's d<br>3. Hedge's g<br>4. Eta Square | **Effect size statistic used by original author for EL subsample analysis** |

| | | |
|---|---|---|
| | 9. Not reported/did not calculate ES for ELs<br>10. PND<br>11. Other | |
| OES | NR = Not reported/did not calculate ES for ELs | **Effect size reported by original authors**<br><br>Numeric value of the effect size statistic<br><br>Note: Report this variable only when mean and SD are not reported |
| Ot | NA= not applicable | **t statistic from a t-test (original reporting) of EL subsample**<br><br>Note: Report this variable only when mean and SD are not reported |
| Odf | NA= not applicable | **Degrees of freedom value used**<br><br>Note: Report this variable only when mean and SD are not reported |
| Of | NA= not applicable | **F-value (original reporting) of EL subsample**<br><br>Note: Report this variable only when mean and SD are not reported |
| TyScore | 1. Pre-post gain score<br>2. Post-test group comparison score<br>3. Other | **Type of Score**<br><br>1. Pre-post gain score:<br>2. Post-test group comparison score<br>3. Other<br><br>Note: Report this variable only when mean and SD are not reported |
| FuncT | 0. No<br>1. Yes | **Trend (SCD-only)**<br><br>The data path follows in the desired or expected direction for the target behavior (e.g., increasing number of words acquired) |

| | | |
|---|---|---|
| FuncL | 0. No<br>1. Yes | **Level (SCD-only)**<br><br>There is a change in level in the desired direction from baseline to intervention.<br><br>Yes: The mean of the first three intervention data points is larger than the mean of the last three baseline data points. |
| FuncR | 0. No<br>1. Yes | **Functional relation is demonstrated (SCD-only)**<br><br>A change in the outcome measure resulted in the introduction of the intervention and is demonstrated by showing a positive trend and a change in level.<br><br>1. No: Effects are consistently and/or sporadically observed without the introduction of the intervention<br>Yes: both FuncT and FuncL are coded as yes. |
| FuncRep | 0. No<br>1. Yes | **There are three or more replication of effects across participants, or behaviors. (SCD-only)** |
| FuncRo | 0. No replication<br>1. Participants<br>2. Behaviors | **Replication of effects was observed across: (SCD-only)**<br><br>Participants: Independent participants within the same study<br>Behaviors: different types of words, different content area, different strategy use |

| **Datapoints** | | |
|---|---|---|
| DVID | | *Same coding used in the previous section* |

| | | |
|---|---|---|
| PartID | | **Participant identification (SCD only)**<br><br>All participants are identified with a numerically starting with the first participant as 1. |
| PartName | | **Name of the participant**<br><br>Use this variable to support with identification. Record the name of the participant if reported by the author, otherwise ignore. |
| Session | | **Session**<br><br>Indicate the session that the datapoint was collected. Label values 0 – $n^{th}$ value.<br><br>Example:<br>A study collected baseline data during sessions 1, 3, and 6. Enter 1, 3, and 6 into their own row for this variable. |
| Trt | 0. Baseline<br>1. Treatment/intervention | **Treatment coding of phases**<br><br>Indicate whether the datapoint was collected during baseline or intervention phase. |
| Outcome | | **The outcome value of the dependent variable.**<br><br>These are the y-values extracted from WebplotDigitizer. |
| TrtSess | | **Treatment Session**<br><br>Assign treatment sessions that correspond to each y-value. Label values from 0 – $n^{th}$ value.<br><br>All baseline datapoints are labeled with 0. First treatment session is |

| | | |
|---|---|---|
| | | coded as 0, second treatment session as 1, and etc. |

| **Methodological Characteristics** | | |
|---|---|---|
| **Variable Label** | **Assigned Codes** | **Descriptions** (when applicable) |
| QIsett_geo<br>QIsett_ses<br>QIsett_T | 0. No<br>1. Yes<br>NR= not reported | **Methodological reporting on setting**<br><br>The study provided sufficient information to determine:<br>QIset_geo: the geographic location (*both region and geographic area must be reported to be coded as yes*)<br>QIsett_ses: the socio-economic status of the school environment<br><br>QIsett:_T Sum of all items in this area |
| QIpart_age<br>QIpart_sex<br>QIpart_ethnic<br>QIpart_frl<br>QIpart_elid<br>QIpart_T | 0. No<br>1. Yes<br>NR= not reported | **Methodological reporting on participants**<br><br>The study provided sufficient information to determine participant:<br>QIpart_age: Ages or age range of the study sample<br>QIpart_sex: Gender composition of the study sample<br>QIethnic: Racial/ethnic composition of the study sample<br>QIpart_frl: socio-economic status of the sample population (free/reduced lunch is used as a proxy)<br>QIpart_elid: methods used to determine EL status<br><br>QIpart:_T Sum of all items in this area |
| QIagent_txprov<br>QIagent_provLic<br>QIagent_trn<br>QIagent_T | 0. No<br>1. Yes<br>NR = not reported/insufficient information | **Methodological reporting on intervention agent**<br><br>The study provided sufficient information to determine: |

| | | |
|---|---|---|
| | | QIagent_txprov: who delivered the intervention (e.g., teacher, researcher, computer program)<br>QIagent_provLic: qualification, educational background or licensure of the intervention provider<br>QIagent_trn: that the training or qualification (e.g., professional credential) required to implement the intervention was provided (e.g., # of hours training provided)<br><br>QIagent_T: Sum of all items in this area |
| QIintv_Vstrat<br>QIintv_dose<br>QIintv_matS<br>QIintv_matT<br>QIintv_T | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on intervention program/curriculum**<br><br>The study provided sufficient information to determine:<br>QIintv_Vstrat: the specific instructional strategies used during the intervention<br>QIintv_dose: the overall dosage of intervention implemented<br>QIintv_matS: the intervention materials (e.g., manipulatives, worksheets) used with students or cited at least one accessible source providing the information<br>QIintv_matT: the intervention materials intervention providers used (e.g., teacher's manual) or cited at least one accessible source providing the information<br><br>QIintv_T: Sum of all items in this area |
| QIfi_ad<br>QIfi_avg<br>QIfi_T | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on fidelity of implementation**<br><br>The study provided sufficient information to determine:<br>QIfi_ad: that implementation fidelity related to adherence was collected using direct observational checklists or methods |

| | | |
|---|---|---|
| | | QIfi_avg: that an overall, mean or range of implementation fidelity was reported<br><br>QIfi_T: Sum of all items in this area |
| QIval_C<br>QIval_CLim<br>QIval_Asn<br>QIval_attr<br>QIval_exp<br>QIval_T | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on internal validity**<br><br>The study provided sufficient information to determine:<br><br>QIval_Cl: characteristics of baseline and control/comparison conditions<br>QIval_CLim: that control/comparison conditions were restricted or had limited access to the treatment intervention<br>QIval_Asn: how assignments to control/comparison and intervention or intervention sequence (ABBAB) were made<br>QIval_attr: that the rate of attrition from baseline is reported or can be determined by information in the manuscript **(GSs-only)**<br>QIval_exp: that showed at least three demonstrations of experimental effects at three different times (**SCD-only)**<br><br>QIval_T: Sum of all items in this area |
| QIoutc_dv<br>QIoutc_results<br>QIoutc_ioa<br>QIoutc_rel<br>QIsoutc_soc<br>QIoutc_T | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on outcome measures**<br><br>The study provided sufficient information to determine:<br>QIoutc_dv: the constructs that outcome measures assessed<br>QIroutc_results: that the results of all outcome measures are reported and not just results with positive findings<br>QIioutc_ioa: that inter-observer reliability was collected and reported |

| | | |
|---|---|---|
| | | QIoutc_rel: that reliability coefficients of all outcome measures were reported **(GS only)**<br>QIsoutc_soc: that social validity data were collected and reported **(SCD only)**<br><br>QIoutc_T: Sum of all items in this area |
| QIdata_viz<br>QIdata_es<br>QIdata_T | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on data analyses**<br><br>The study provided sufficient information to determine:<br>QIdata_viz: that graphs were clear and displayed for all participants to allow reliable visual analysis **(SCD only)**<br>QIdata_es: that an effect size statistic is reported for all outcome variables regardless of its statistical significance<br><br>QIdata_T: Sum of all items in this area |
| QIrq | 0. No<br>1. Yes<br>NA = Not applicable | **Methodological reporting on conceptualization**<br><br>The study provided sufficient information to determine:<br>QIrq: that research questions or hypotheses were explicitly stated |
| QItot | | **QItot**: the total sum of scores across methodological variables (max value = xx [SCD]; xx [group]) |

| **WWC GS** | | |
|---|---|---|
| WWC_Rdzn | 0. No<br>1. Yes | **Group Design Category**<br><br>Is intervention and comparison group membership determined through a random process? |

| | | |
|---|---|---|
| | | If (yes-RCT), **go to attrition**, if (no-Quasi) go to baseline equivalence. |
| WWC_attri | 0. No<br>1. Yes<br>NR = Not reported | **Sample attrition**<br><br>Is the combination of overall and differential attrition low?<br><br>Follow the flow chart of attrition and potential bias.<br><br>*Low attrition* = Green zone<br><br>*High attrition* = red zone; potentially yellow zone if it is determined that attrition rates are related to the intervention |
| WWC_BE | 0. No<br>1. Yes<br>NR = Not Reported/not sufficient info | **Baseline Equivalence**<br><br>Is equivalence established at baseline for the groups in the analytic sample?<br><br>● \|Baseline ES\| > 0.25 **(not equivalent)**<br>● ≤ \|Baseline ES\| ≤ 0.05 **(equivalent).**<br>● 0.05 < \|Baseline ES\| ≤ 0.25 **(equivalent if statistical adjustment is applied).**<br><br>Note: If authors do not calculate BE, use coding handout to calculate BE from the pretest scores of the group. |
| WWC_FD | 0. Does not meet<br>1. Meets without reservations<br>2. Meets with reservations | **Final determination of WWC Design Standards for GSs**<br><br>If both WWC_Rdzn and WWC_attri are yes then 1-Meets without reservations, or else, follow flowchart to determine final designation. |

| **WWC SCDs** |
|---|

| | | |
|---|---|---|
| WWC_man | 0. No<br>1. Yes | **Manipulation of IV**<br><br>The independent variable is systematically manipulated |
| WWC_dem | 0. No<br>1. Yes | **Intervention demonstration**<br><br>At least three demonstrations of an intervention effect are observed across timepoints or phases. |
| WWC_5pt | 0. No<br>1. Yes | **Minimum data points collected**<br><br>There are at least 5 points within each phase/condition for each participant and outcome measure.<br><br>Note:<br>Make sure to apply this definition to both baseline and intervention phases. Hence, a 1 is only awarded if at least 5 datapoints are collected in baseline and at least 5 datapoints are collected during intervention.<br><br>Flag SCDs that use only pre and post vocab DVs. Will need to reconsider if these studies should be included as part of the meta.<br>An alternating treatment design needs *five repetitions* of the alternating sequence to *Meet Standards.* Designs such as ABABBABAABBA, BCBCBCBCBC, and AABBAABBAABB would qualify, even though randomization or brief functional assessment may lead to one or two data points in a phase. A design with four repetitions would *Meet Standards with Reservations*, and a design with fewer than four repetitions *Does Not Meet Standards (WWC SCD standards)*<br>To *Meet Standards* a multiple baseline design must have a minimum of six |

| | | |
|---|---|---|
| | | phases with at least 5 data points per phase. To *Meet Standards with Reservations* a multiple baseline design must have a minimum of six phases with at least 3 data points per phase. Any phases based on fewer than three data points *cannot* b*e used to demonstrat*e existence or lack of an effect. |
| WWC_IOA20 | 0. No<br>1. Yes | **Inter-assessor agreement collected across phases**<br><br>At least 20% of data points are collected for IOA in each condition/phase. |
| WWC_TH | 0. No<br>1. Yes | **Inter-assessor agreement threshold**<br><br>IOA meets the appropriate threshold. >0.80-0.90 for percentage agreement >0.60 for Cohen's kappa |
| WWC_FD | 0. Does not meet<br>1. Meets without reservations<br>2. Meets with reservations | **Meets WWC Design Standards for SCDs**<br><br>Final determination for WWC standards. If responded yes to all WWC items, then Meets without Reservations, or else follow handout flowchart for appropriate designation. |

# Appendix E: dmetar Outlier Analysis Scenario Results

## *Group Study Outlier Analysis Results and Scenario (Hedge's g)*

**Studies identified as potential outliers**

```
"Kieffer et al. 2012"  "Proctor et al. 2011"  "Proctor et al. 2011"
"Proctor et al. 2011"  "Denton et al. 2008"   "Vaughn et al. 2006"
"McBroom 2009"   "Mieure 2014"                "Yang 2015"
"Burns 2001"     "Crevecoeur et al. 2014" "Cervetti et al. 2015"
```

**Setting study weights to 0**

| | SMD | 95%-CI | %W(random) | exclude |
|---|---|---|---|---|
| Kieffer et al. 2012 | 0.2221 | [ 0.0065; 0.4376] | 3.6 | |
| Kieffer et al. 2012 | 0.0597 | [-0.1553; 0.2747] | 0.0 | * |
| Nelson et al. 2011 | 0.5967 | [ 0.3033; 0.8901] | 3.0 | |
| Nelson et al. 2011 | 0.0289 | [-0.2581; 0.3159] | 3.0 | |
| Proctor et al. 2011 | 0.0073 | [-0.3512; 0.3658] | 2.6 | |
| Proctor et al. 2011 | 0.6198 | [ 0.2527; 0.9869] | 2.5 | |
| Proctor et al. 2011 | 1.2586 | [ 0.8658; 1.6514] | 0.0 | * |
| Proctor et al. 2011 | 1.2799 | [ 0.8860; 1.6739] | 0.0 | * |
| Proctor et al. 2011 | 1.3233 | [ 0.9271; 1.7196] | 0.0 | * |
| Denton et al. 2008 | -0.6463 | [-1.2940; 0.0015] | 0.0 | * |
| Vaughn et al. 2006 | 0.0106 | [-0.4454; 0.4666] | 2.0 | |
| Vaughn et al. 2006 | -0.0079 | [-0.4454; 0.4295] | 2.1 | |
| Vaughn et al. 2006 | -0.0297 | [-0.4715; 0.4121] | 2.1 | |
| Vaughn et al. 2006 | 0.1334 | [-0.3081; 0.5749] | 2.1 | |
| Vaughn et al. 2006 | 0.1526 | [-0.2631; 0.5684] | 2.2 | |
| Vaughn et al. 2006 | -0.2045 | [-0.6157; 0.2066] | 0.0 | * |
| Vaughn et al. 2006 | 0.1122 | [-0.2962; 0.5206] | 2.3 | |
| Vaughn et al. 2006 | 0.2468 | [-0.1629; 0.6564] | 2.3 | |
| Vaughn et al. 2009 | 0.5651 | [ 0.1516; 0.9787] | 2.2 | |
| Vaughn et al. 2009 | 0.0303 | [-0.3616; 0.4222] | 2.4 | |
| Crum 2017 | 0.2396 | [-0.1527; 0.6320] | 2.4 | |
| Cruz-Cruz 2005 | 0.2963 | [-0.4271; 1.0197] | 1.1 | |
| Dack 1996 | 0.7168 | [ 0.1576; 1.2760] | 1.6 | |
| Dack 1996 | 0.9224 | [ 0.3524; 1.4924] | 1.5 | |
| Dack 1996 | 0.8244 | [ 0.2598; 1.3889] | 1.6 | |
| Frasco 2008 | 0.5552 | [-0.1143; 1.2247] | 1.2 | |
| McBroom 2009 | 0.2837 | [-0.7027; 1.2700] | 0.7 | |
| McBroom 2009 | 0.0448 | [-0.9361; 1.0257] | 0.7 | |
| McBroom 2009 | 1.0172 | [-0.0334; 2.0679] | 0.6 | |
| McBroom 2009 | 1.7159 | [ 0.5472; 2.8846] | 0.0 | * |
| Mieure 2014 | 1.2881 | [ 0.7701; 1.8060] | 0.0 | * |
| Mieure 2014 | 0.9173 | [ 0.4205; 1.4140] | 1.8 | |
| Benoit 2017 | 0.4694 | [-0.0374; 0.9761] | 1.8 | |
| Stevens 2018 | 0.7893 | [ 0.3654; 1.2132] | 2.2 | |
| Vang 2004 | -0.1654 | [-0.8434; 0.5126] | 1.2 | |
| Weitz 2003 | 0.1370 | [-0.2848; 0.5589] | 2.2 | |
| Yang 2015 | 1.7543 | [ 1.1828; 2.3258] | 0.0 | * |
| Yang 2015 | 0.1861 | [-0.2990; 0.6712] | 1.9 | |
| Burns 2001 | -0.1326 | [-0.5725; 0.3073] | 0.0 | * |

| | | | | |
|---|---|---|---|---|
| Burns 2001 | 0.1958 | [-0.2448; 0.6363] | 2.1 | |
| Wanzek et al. 2017 | 0.1802 | [-0.3441; 0.7045] | 1.7 | |
| Wanzek et al. 2017 | 0.0971 | [-0.4304; 0.6246] | 1.7 | |
| Avila & Sadoski 1996 | 0.6947 | [ 0.0360; 1.3533] | 1.3 | |
| Avila & Sadoski 1996 | 1.0650 | [ 0.3804; 1.7495] | 1.2 | |
| Neuman & Kaefer 2018 | 0.1783 | [-0.2494; 0.6059] | 2.2 | |
| Neuman & Kaefer 2018 | 0.6305 | [ 0.1932; 1.0679] | 2.1 | |
| Neuman & Kaefer 2018 | -0.0310 | [-0.4579; 0.3958] | 2.2 | |
| Neuman & Kaefer 2018 | 0.9314 | [ 0.4820; 1.3809] | 2.1 | |
| Crevecoeur et al. 2014 | 0.3109 | [-0.2810; 0.9027] | 1.5 | |
| Crevecoeur et al. 2014 | 1.1817 | [ 0.5537; 1.8097] | 0.0 | * |
| Graves et al. 2011 | -0.1085 | [-0.6229; 0.4060] | 1.8 | |
| Graves et al. 2011 | -0.1392 | [-0.7181; 0.4397] | 1.5 | |
| August et al. 2009 | 0.2856 | [ 0.1194; 0.4518] | 3.9 | |
| Bravo & Cervetti 2014 | 0.6564 | [ 0.2832; 1.0296] | 2.5 | |
| Cena et al. 2013 | 0.7911 | [ 0.2234; 1.3587] | 1.6 | |
| Cena et al. 2013 | 0.8514 | [ 0.2804; 1.4224] | 1.5 | |
| Cena et al. 2013 | 0.8142 | [ 0.2453; 1.3832] | 1.5 | |
| Cena et al. 2013 | 0.1337 | [-0.4130; 0.6804] | 1.6 | |
| Cervetti et al. 2015 | -0.1334 | [-0.4687; 0.2019] | 0.0 | * |
| Lawrence et al. 2012 | 0.0319 | [-0.3822; 0.4460] | 2.2 | |
| Ulanoff & Pucci 1999 | 0.5937 | [-0.0451; 1.2325] | 1.3 | |
| Tong et al. 2014 | 0.7962 | [ 0.2578; 1.3346] | 1.7 | |
| Tong et al. 2015 | 0.7839 | [ 0.2474; 1.3204] | 1.7 | |
| Tong et al. 2015 | 0.5749 | [ 0.0475; 1.1022] | 1.7 | |
| Kittley-Koshenina 2009 | 0.1468 | [-0.8932; 1.1868] | 0.6 | |

Number of studies combined: k = 53

| | SMD | 95%-CI | t | p-value |
|---|---|---|---|---|
| **Random effects model** | 0.3592 | [ 0.2674; 0.4510] | 7.85 | < 0.0001 |
| Prediction interval | | [-0.0707; 0.7891] | | |

**Quantifying heterogeneity:**
tau^2 = 0.0438 [0.0172; 0.1109]; tau = 0.2092 [0.1310; 0.3330];
I^2 = 46.6% [26.3%; 61.4%]; H = 1.37 [1.16; 1.61]

**Test of heterogeneity:**

| Q | d.f. | p-value |
|---|---|---|
| 97.41 | 52 | 0.0001 |

***Single Case Design Study Outlier Analysis Results and Scenario (BCTau)***

**Studies identified as potential outliers**

"Helman 2015" "Helman 2015" "Helman 2015"
"Hinrichs 2008" "Kim & Linan-Thompson 2013" "Kim & Linan-Thompson 2013"
"Kim & Linan-Thompson 2013" "Alison et al. 2017"

**Setting study weights to 0**

| | SMD | 95%-CI | %W(random) | exclude |
|---|---|---|---|---|
| Anderson 2014 | 0.7620 | [ 0.6528; 0.8712] | 2.2 | |
| Anderson 2014 | 0.7860 | [ 0.6859; 0.8861] | 2.6 | |
| Anderson 2014 | 0.7940 | [ 0.7092; 0.8788] | 3.7 | |
| Anderson 2014 | 0.7800 | [ 0.6894; 0.8706] | 3.2 | |
| Anderson 2014 | 0.7810 | [ 0.6962; 0.8658] | 3.7 | |
| Anderson 2014 | 0.7700 | [ 0.6811; 0.8589] | 3.3 | |
| Anderson 2014 | 0.7450 | [ 0.5277; 0.9623] | 0.6 | |
| Anderson 2014 | 0.7450 | [ 0.5277; 0.9623] | 0.6 | |
| Anderson 2014 | 0.7070 | [ 0.4270; 0.9870] | 0.3 | |
| Anderson 2014 | 0.7250 | [ 0.4596; 0.9904] | 0.4 | |
| Anderson 2014 | 0.5980 | [ 0.1778; 1.0182] | 0.1 | |
| Anderson 2014 | 0.6200 | [ 0.2178; 1.0222] | 0.2 | |
| Anderson 2014 | 0.5770 | [ 0.1422; 1.0118] | 0.1 | |
| Anderson 2014 | 0.5770 | [ 0.1422; 1.0118] | 0.1 | |
| Anderson 2014 | 0.7070 | [ 0.4270; 0.9870] | 0.3 | |
| Anderson 2014 | 0.6900 | [ 0.3965; 0.9835] | 0.3 | |
| Anderson 2014 | 0.7450 | [ 0.5277; 0.9623] | 0.6 | |
| Anderson 2014 | 0.7450 | [ 0.5277; 0.9623] | 0.6 | |
| Anderson 2014 | 0.5770 | [ 0.1422; 1.0118] | 0.1 | |
| Anderson 2014 | 0.5770 | [ 0.1422; 1.0118] | 0.1 | |
| Anderson 2014 | 0.7070 | [ 0.4270; 0.9870] | 0.3 | |
| Anderson 2014 | 0.7070 | [ 0.4270; 0.9870] | 0.3 | |
| Anderson 2014 | 0.7750 | [ 0.5793; 0.9707] | 0.7 | |
| Anderson 2014 | 0.7750 | [ 0.5793; 0.9707] | 0.7 | |
| Helman 2015 | 0.8060 | [-0.4673; 2.0793] | 0.0 | |
| Helman 2015 | 0.8670 | [ 0.8246; 0.9094] | 14.7 | |
| Helman 2015 | 0.6250 | [ 0.4658; 0.7842] | 0.0 | |
| * | | | | |
| Helman 2015 | 0.8450 | [ 0.7916; 0.8984] | 9.2 | |
| Helman 2015 | 0.8590 | [ 0.8101; 0.9079] | 11.0 | |
| Helman 2015 | 0.6000 | [ 0.4329; 0.7671] | 0.0 | |
| * | | | | |
| Helman 2015 | 0.9400 | [ 0.9204; 0.9596] | 0.0 | |
| * | | | | |
| Helman 2015 | 0.8570 | [ 0.7990; 0.9150] | 7.8 | |
| Helman 2015 | 0.7070 | [ 0.2170; 1.1970] | 0.1 | |
| Helman 2015 | 0.5160 | [-0.2038; 1.2358] | 0.1 | |
| Helman 2015 | 0.3330 | [-0.5390; 1.2050] | 0.0 | |
| Helman 2015 | 0.7750 | [ 0.3834; 1.1666] | 0.2 | |
| Helman 2015 | 0.7750 | [ 0.3834; 1.1666] | 0.2 | |
| Helman 2015 | 0.7750 | [ 0.3834; 1.1666] | 0.2 | |
| Helman 2015 | 0.3330 | [-0.5390; 1.2050] | 0.0 | |

```
Helman 2015                0.5770 [-0.0755; 1.2295]        0.1
Hinrichs 2008              0.7113 [ 0.5246; 0.8980]        0.8
Hinrichs 2008              0.6433 [ 0.4271; 0.8596]        0.6
Hinrichs 2008              0.6797 [ 0.4823; 0.8770]        0.7
Hinrichs 2008              0.7000 [ 0.5074; 0.8926]        0.7
Hinrichs 2008              0.5423 [ 0.2890; 0.7956]        0.0
*
Lia 2010                   0.7818 [ 0.5319; 1.0318]        0.4
Lia 2010                   0.7887 [ 0.5547; 1.0227]        0.5
Lia 2010                   0.6165 [ 0.2763; 0.9567]        0.2
Lia 2010                   0.8092 [ 0.5958; 1.0226]        0.6
Green et al. 2015          0.7300 [ 0.4257; 1.0343]        0.3
Green et al. 2015          0.7560 [ 0.4760; 1.0360]        0.3
Guardino et al. 2014       0.8013 [ 0.6204; 0.9823]        0.8
Guardino et al. 2014       0.8243 [ 0.6212; 1.0275]        0.6
Guardino et al. 2014       0.9050 [ 0.8062; 1.0038]        2.7
Helman et al. 2015         0.7170 [ 0.5049; 0.9291]        0.6
Helman et al. 2015         0.7910 [ 0.6075; 0.9745]        0.8
Helman et al. 2015         0.7980 [ 0.6685; 0.9275]        1.6
Helman et al. 2015         0.7280 [ 0.5235; 0.9325]        0.6
Helman et al. 2015         0.8260 [ 0.6701; 0.9819]        1.1
Helman et al. 2015         0.8070 [ 0.6825; 0.9315]        1.7
Kim & Linan-Thompson 2013  0.6250 [ 0.3596; 0.8904]        0.4
Kim & Linan-Thompson 2013  0.7350 [ 0.5963; 0.8737]        1.4
Kim & Linan-Thompson 2013  0.7170 [ 0.5709; 0.8631]        1.2
Kim & Linan-Thompson 2013  0.6050 [ 0.4491; 0.7609]        0.0
*
Kim & Linan-Thompson 2013  0.6960 [ 0.4434; 0.9486]        0.4
Kim & Linan-Thompson 2013  0.6230 [ 0.4383; 0.8077]        0.0
*
Kim & Linan-Thompson 2013  0.7170 [ 0.5709; 0.8631]        1.2
Kim & Linan-Thompson 2013  0.6880 [ 0.5585; 0.8175]        0.0
*
Alison et al. 2017         0.5350 [ 0.4015; 0.6685]        0.0
*
Alison et al. 2017         0.7620 [ 0.6836; 0.8404]        4.3
Alison et al. 2017         0.7670 [ 0.6902; 0.8438]        4.5
Cannon et al. 2010         0.6603 [ 0.4564; 0.8642]        0.6
Cannon et al. 2010         0.7097 [ 0.5312; 0.8882]        0.8
Cannon et al. 2010         0.6053 [ 0.3722; 0.8384]        0.5
Cannon et al. 2010         0.6113 [ 0.3754; 0.8473]        0.5

Number of studies combined: k = 67

                         SMD            95%-CI     t  p-value
Random effects model 0.8029 [0.7868; 0.8191] 99.11 < 0.0001
Prediction interval         [0.7867; 0.8191]

Quantifying heterogeneity:
 tau^2 = 0 [0.0000; 0.0008]; tau = 0 [0.0000; 0.0287];
 I^2 = 0.0% [0.0%; 26.0%]; H = 1.00 [1.00; 1.16]
```

**Test of heterogeneity:**

| Q | d.f. | p-value |
|---|---|---|
| 63.24 | 66 | 0.573 |

# Appendix F: Supplemental Materials

## Moderator Analyses for Only Group Study Effects

*Relation Between Grade Level Groupings and Vocabulary Effects*

| | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $I^2$ | $T^2$ |
|---|---|---|---|---|---|---|---|---|
| Elementary[a] | 50 | 0.45 | 0.08 | <.001 | 0.28 | 0.62 | 70.11 | 0.11 |
| Middle School | 14 | -0.31 | 0.08 | 0.00 | -0.48 | -0.15 | | |
| High School | 1 | -0.23 | 0.14 | 0.12 | -0.52 | 0.06 | | |

*Relation Between Study Characteristics and Vocabulary Effects*

| Study Characteristics | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $I^2$ | $T^2$ |
|---|---|---|---|---|---|---|---|---|
| Instructional Programming | | | | | | | | |
| Combination[a] | 21 | 0.47 | 0.12 | 0.005 | 0.19 | 0.75 | 71.82 | 0.13 |
| Explicit Instruction | 24 | -0.13 | 0.17 | 0.47 | -0.49 | 0.23 | | |
| Incidental Instruction | 2 | -0.46 | 0.19 | 0.19 | -1.72 | 0.80 | | |
| NR | 18 | -0.11 | 0.16 | 0.49 | -0.45 | 0.23 | | |
| Intervention Provider | | | | | | | | |
| Research Team[a] | 6 | 0.04 | 0.09 | 0.64 | -0.20 | 0.29 | 67.07 | 0.10 |
| Classroom Teacher | 35 | 0.27 | 0.12 | 0.07 | -0.03 | 0.56 | | |
| Combination | 14 | 0.51 | 0.21 | 0.04 | 0.02 | 1.00 | | |
| Paraprofessional | 4 | 0.54 | 0.33 | 0.23 | -0.73 | 1.81 | | |
| Self-administered | 6 | 0.80 | 0.10 | 0.01 | 0.42 | 1.19 | | |
| Intervention Dosage | | | | | | | | |
| Frequency of Training | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 40 or fewer sessions[a] | 12 | 0.26 | 0.15 | 0.12 | -0.09 | 0.60 | 69.22 | 0.10 |
| more than 40 sessions | 40 | 0.07 | 0.17 | 0.69 | -0.30 | 0.44 | | |
| Duration | | | | | | | | |
| 30 hr or less[a] | 25 | 0.31 | 0.10 | 0.01 | 0.09 | 0.54 | 67.98 | 0.10 |
| More than 30 hr | 24 | 0.00 | 0.15 | 0.99 | -0.31 | 0.30 | | |
| Intensity | | | | | | | | |
| 20 min or less[a] | 15 | 0.24 | 0.09 | 0.03 | 0.03 | 0.45 | | |
| more than 20 min | 38 | 0.12 | 0.13 | 0.35 | -0.16 | 0.40 | | |
| Target Word Domain | | | | | | | | |
| Content-specific[a] | 7 | 0.35 | 0.15 | 0.06 | -0.03 | 0.72 | 71.19 | 0.13 |
| General academic | 23 | 0.19 | 0.20 | 0.35 | -0.24 | 0.63 | | |
| NR | 21 | -0.16 | 0.20 | 0.41 | -0.59 | 0.26 | | |
| Combination | 14 | 0.10 | 0.18 | 0.59 | -0.30 | 0.50 | | |

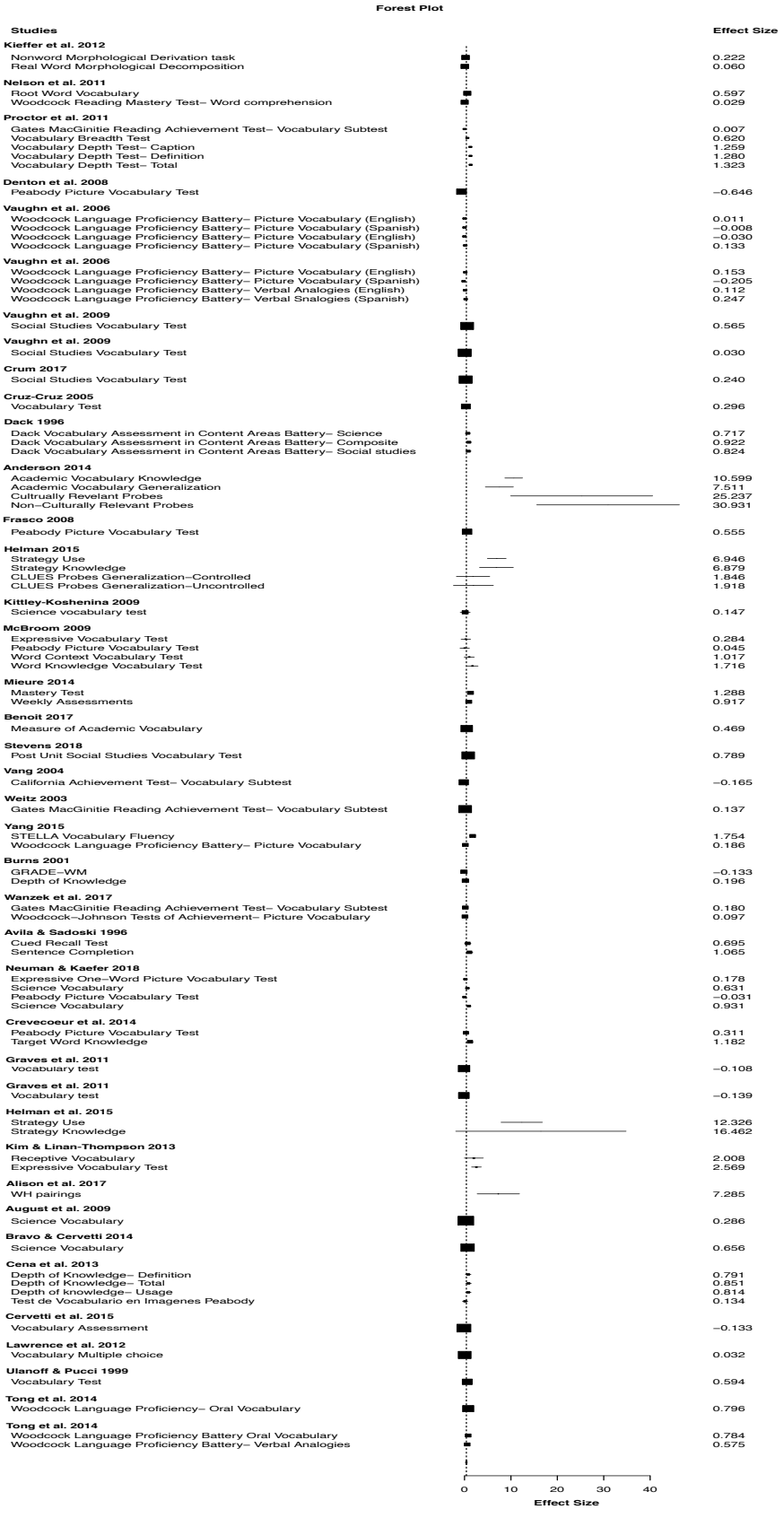
*Relation Between Outcome Characteristics and Vocabulary Effects*

| | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $I^2$ | $T^2$ |
|---|---|---|---|---|---|---|---|---|
| Measurement Production | | | | | | | | |
| Author-Created[a] | 49 | 0.46 | 0.09 | <.001 | 0.27 | 0.65 | 69.35 | 0.11 |
| Standardized | 29 | -0.25 | 0.13 | 0.06 | -0.51 | 0.01 | | |
| Vocabulary Scale | | | | | | | | |
| Combination[a] | 2 | 0.45 | 0.18 | 0.25 | -1.89 | 2.80 | 70.13 | 0.12 |
| Expressive | 24 | 0.12 | 0.22 | 0.66 | -1.12 | 1.35 | | |
| Receptive | 39 | -0.19 | 0.20 | 0.49 | -1.81 | 1.42 | | |
| Taxonomy Scale | | | | | | | | |
| NR[a] | 2 | 0.17 | 0.13 | 0.40 | -1.43 | 1.77 | 71.12 | 0.13 |
| Word Knowledge | 43 | 0.15 | 0.15 | 0.49 | -1.10 | 1.39 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Word Learning | 20 | 0.35 | 0.18 | 0.21 | -0.61 | 1.32 | | |

*Relation Between Methodological Rigor and Vocabulary Effects*

| | | *k* | *g* | *SE* | *p* | $CI_{95}$ L | $CI_{95}$U | $T^2$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|---|
| WWC | | | | | | | | | |
| | DNM[a] | 21 | 0.37 | 0.13 | 0.02 | 0.08 | 0.66 | 0.12 | 71.53 |
| | Met | 28 | -0.05 | 0.16 | 0.77 | -0.38 | 0.29 | | |
| | MWR | 16 | 0.09 | 0.18 | 0.63 | -0.30 | 0.48 | | |
| CEC | | | | | | | | | |
| | DNM80[a] | 63 | 0.37 | 0.07 | <.001 | 0.24 | 0.51 | 0.12 | 70.07 |
| | Met80 | 2 | 0.23 | 0.07 | 0.002 | -0.37 | -0.09 | | |


Forest Plot
Studies
Effect Size
Kieffer et al. 2012
Nonword Morphological Derivation task 0.222
Real Word Morphological Decomposition 0.060
Nelson et al. 2011
Root Word Vocabulary 0.597
Woodcock Reading Mastery Test– Word comprehension 0.029
Proctor et al. 2011
Gates MacGinitie Reading Achievement Test– Vocabulary Subtest 0.007
Vocabulary Breadth Test 0.620
Vocabulary Depth Test– Caption 1.259
Vocabulary Depth Test– Definition 1.280
Vocabulary Depth Test– Total 1.323
Denton et al. 2008
Peabody Picture Vocabulary Test −0.646
Vaughn et al. 2006
Woodcock Language Proficiency Battery– Picture Vocabulary (English) 0.011
Woodcock Language Proficiency Battery– Picture Vocabulary (Spanish) −0.008
Woodcock Language Proficiency Battery– Picture Vocabulary (English) −0.030
Woodcock Language Proficiency Battery– Picture Vocabulary (Spanish) 0.133
Vaughn et al. 2006
Woodcock Language Proficiency Battery– Picture Vocabulary (English) 0.153
Woodcock Language Proficiency Battery– Picture Vocabulary (Spanish) −0.205
Woodcock Language Proficiency Battery– Verbal Analogies (English) 0.112
Woodcock Language Proficiency Battery– Verbal Snalogies (Spanish) 0.247
Vaughn et al. 2009
Social Studies Vocabulary Test 0.565
Vaughn et al. 2009
Social Studies Vocabulary Test 0.030
Crum 2017
Social Studies Vocabulary Test 0.240
Cruz-Cruz 2005
Vocabulary Test 0.296
Dack 1996
Dack Vocabulary Assessment in Content Areas Battery– Science 0.717
Dack Vocabulary Assessment in Content Areas Battery– Composite 0.922
Dack Vocabulary Assessment in Content Areas Battery– Social studies 0.824
Anderson 2014
Academic Vocabulary Knowledge 10.599
Academic Vocabulary Generalization 7.511
Cultrually Revelant Probes 25.237
Non–Culturally Relevant Probes 30.931
Frasco 2008
Peabody Picture Vocabulary Test 0.555
Helman 2015
Strategy Use 6.946
Strategy Knowledge 6.879
CLUES Probes Generalization–Controlled 1.846
CLUES Probes Generalization–Uncontrolled 1.918
Kittley-Koshenina 2009
Science vocabulary test 0.147
McBroom 2009
Expressive Vocabulary Test 0.284
Peabody Picture Vocabulary Test 0.045
Word Context Vocabulary Test 1.017
Word Knowledge Vocabulary Test 1.716
Mieure 2014
Mastery Test 1.288
Weekly Assessments 0.917
Benoit 2017
Measure of Academic Vocabulary 0.469
Stevens 2018
Post Unit Social Studies Vocabulary Test 0.789
Vang 2004
California Achievement Test– Vocabulary Subtest −0.165
Weitz 2003
Gates MacGinitie Reading Achievement Test– Vocabulary Subtest 0.137
Yang 2015
STELLA Vocabulary Fluency 1.754
Woodcock Language Proficiency Battery– Picture Vocabulary 0.186
Burns 2001
GRADE–WM −0.133
Depth of Knowledge 0.196
Wanzek et al. 2017
Gates MacGinitie Reading Achievement Test– Vocabulary Subtest 0.180
Woodcock–Johnson Tests of Achievement– Picture Vocabulary 0.097
Avila & Sadoski 1996
Cued Recall Test 0.695
Sentence Completion 1.065
Neuman & Kaefer 2018
Expressive One–Word Picture Vocabulary Test 0.178
Science Vocabulary 0.631
Peabody Picture Vocabulary Test −0.031
Science Vocabulary 0.931
Crevecoeur et al. 2014
Peabody Picture Vocabulary Test 0.311
Target Word Knowledge 1.182
Graves et al. 2011
vocabulary test −0.108
Graves et al. 2011
Vocabulary test −0.139
Helman et al. 2015
Strategy Use 12.326
Strategy Knowledge 16.462
Kim & Linan-Thompson 2013
Receptive Vocabulary 2.008
Expressive Vocabulary Test 2.569
Alison et al. 2017
WH pairings 7.285
August et al. 2009
Science Vocabulary 0.286
Bravo & Cervetti 2014
Science Vocabulary 0.656
Cena et al. 2013
Depth of Knowledge– Definition 0.791
Depth of Knowledge– Total 0.851
Depth of knowledge– Usage 0.814
Test de Vocabulario en Imagenes Peabody 0.134
Cervetti et al. 2015
Vocabulary Assessment −0.133
Lawrence et al. 2012
Vocabulary Multiple choice 0.032
Ulanoff & Pucci 1999
Vocabulary Test 0.594
Tong et al. 2014
Woodcock Language Proficiency– Oral Vocabulary 0.796
Tong et al. 2014
Woodcock Language Proficiency Battery Oral Vocabulary 0.784
Woodcock Language Proficiency Battery– Verbal Analogies 0.575
0
10
20
30
40
Effect Size