The Effects of Vocabulary Instruction on Bilingual Students' Lexico-Semantic Acquisition in English-Medium Contexts: A Systematic Analysis

Alain Bengochea

University of Miami, alainbengochea@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/1322
THE EFFECTS OF VOCABULARY INSTRUCTION ON BILINGUAL STUDENTS’ LEXICO-SEMANTIC ACQUISITION IN ENGLISH-MEDIUM CONTEXTS: A SYSTEMATIC ANALYSIS

By

Alain Bengochea

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Coral Gables, Florida

December 2014
THE EFFECTS OF VOCABULARY INSTRUCTION ON BILINGUAL STUDENTS’ LEXICO-SEMANTIC ACQUISITION IN ENGLISH-MEDIUM CONTEXTS: A SYSTEMATIC ANALYSIS

Alain Bengochea

Approved:

Maria S. Carlo, Ph.D.
Associate Professor of
Teaching and Learning

Batya Elbaum, Ph.D.
Professor of
Teaching and Learning

Mary Avalos, Ph.D.
Assistant Research Professor
of Teaching and Learning

M. Brian Blake, Ph.D.
dean of the Graduate School

Soyeon Ahn, Ph.D.
Associate Professor
of Educational and
Psychological Studies
This systematic review provides a comprehensive analysis of the empirical literature on the effectiveness of vocabulary instructional methods for English language learners (ELLs) in prekindergarten through 6th grade settings. Thus, a central goal of this study is to appraise the available research on vocabulary learning and assess intervention studies’ levels of evidence for effectiveness in order to make a critical assessment of the variables that are currently represented in the field. Specifically, the research questions guiding this systematic review are: (1) How do intervention studies developing L2 word-meaning acquisition with ELLs in pre-kindergarten through 6th grade settings vary in their methodological characteristics (i.e., research design, learners, instructional design and outcome)? (a) What are the studies’ levels of evidence for effectiveness based on their design and outcome characteristics? and (2) What is the impact of studies that meet requisite evidence standards on learners’ L2 word-meaning acquisition?

A total of 18 studies meeting the inclusion criteria were identified. Independent coding of 20% of the articles \( n=4 \) was completed, the results were cross-checked against the original codes, and further discrepancies in coding were resolved through discussion between the two coders. The final criteria were subsequently reapplied by the author to all studies in the review. Coding categories for each study were organized by
four main characteristics: (1) design characteristics (i.e., random assignment, baseline equivalence, comparison group type); (2) learner characteristics (i.e., L1/L2 proficiency, SES, age and grade level); (3) treatment characteristics (i.e., focus of instruction, level of L1 support, word tier targeted, total instructional time, training, explicitness of instruction, contextual/definitional support, group size, and level of word exposure) and (4) outcome characteristics (i.e., outcome objective, type of measure, type of ability measured, modality, word tier assessed, and type of knowledge). Based on the studies’ design and outcome characteristics, their level of evidence for effectiveness was also examined. Subsequent analyses revealed that a large proportion of studies underreported information and did not meet evidence standards for effectiveness. Additionally, studies with available effect sizes for measures of breadth and depth of vocabulary knowledge revealed inconclusive evidence about the best instructional practices for ELLs due to the large variations among a limited number of studies in their instructional approach, learners, and outcome characteristics.

An analysis of the merits and limitations of the reviewed studies reveals a number of methodological problems that threaten the validity and credibility of vocabulary research. These findings highlight the need for increased research in this field of study, particularly in revisiting existing practices to add to current understandings relating to vocabulary practices with ELLs. Additionally, this study points to specific issues in current research and provides recommendations for researchers seeking to address the significant knowledge gap in research that is centered on ELL vocabulary learning.
DEDICATION

I dedicate this dissertation to some very special people.

A mis padres, Norka y Armando Bengochea, aunque es posible que no reconozcan totalmente el impacto que han tenido en mi carrera, he tomado sus consejos en cuenta y he basado muchas de mis decisiones en esas lecciones de vida. Gracias a ustedes siempre he reconocido la necesidad de preservar mi identidad multicultural, de encontrar constantemente maneras de comunicarme en español tanto como en inglés y—lo más importante—hacer bien en el mundo. Gracias por todo el apoyo que me han dado y por sus sacrificios a lo largo, los que han ayudado en encaminarme hasta este punto.

To my siblings, Diana, Junior, Juan, Mandy, Vannessa and Jessica, although my work may had inadvertently taken away from quality time, I always knew I had your support throughout this process. I look forward to seeing us grow old and closer no matter where our paths take us.

To my closest friends, Damian Alberti, Stephen Bamonte, and Jorge Vazquez, we have passed through many life stages together in the last two decades, and I owe much of my drive to you all. Perhaps those long hours we used to put in laying out yearbook pages in high school really solidified our work ethic then. Thanks for always remaining to be my biggest supporters and for challenging me on many levels.

À ma petite, Juliet Menendez, tu as été là dès le début, en me voyant passer d’instituteur à chercheur. Merci pour toujours partager tes sages paroles pendant ce parcours.
To my dearest Alex Tejero, you were a firsthand witness to so much of this experience. Thank you for always being there as my sounding board before presenting my ideas to a wider audience and for providing a valuable perspective.
ACKNOWLEDGMENT

I would like to thank several individuals that contributed to my development as a researcher.

My advisor, Dr. Maria Carlo, thank you for your words of inspiration and for reminding me to stay positive about my work. You believed in me from the beginning to which I will always remain grateful.

My committee members, Dr. Mary Avalos, Dr. Batya Elbaum, Dr. Soyeon Ahn, thank you for your wonderful feedback throughout the various stages of my dissertation. Your insights helped greatly to advance this work.

One of my mentors, Dr. Miledis Gort, you took me under your wing and have also become a great friend. Since the beginning, you have been challenging me to never accept ideas at face value in the world of educational research simply because they are widely accepted. Thank you for always pushing me to be retrospective, reflective and forward-thinking.

My friends and colleagues, Sabrina Sembiante and Ryan Pontier, I had the honor of collaborating with you two throughout the PhD program and have become a better researcher as a result. I look forward to seeing us spread our passion for all things bilingual in the future.
**Table of Contents**

List of Figures ................................................................................................................................................. x

List of Tables ................................................................................................................................................ xi

List of Appendices ......................................................................................................................................... xii

Chapter 1 Introduction ......................................................................................................................... 1

  Background and Statement of the Problem ...................................................................................... 1

  Opportunities for vocabulary learning ......................................................................................... 5

  Opportunities to draw on linguistic and cultural resources ...................................................... 7

  Purpose .................................................................................................................................................. 8

  Significance of the Study ............................................................................................................... 8

  General Research Questions ...................................................................................................... 9

Chapter 2: Literature Review ............................................................................................................. 10

  How is Vocabulary Learned? ........................................................................................................ 10

  Broader Scholarly Literature: Enhancing Students’ Lexico-Semantic Knowledge .... 12

  Type of Learning Tasks ................................................................................................................ 13

  Rich, Varied Experiences with Words .......................................................................................... 14

    Providing contextual and definitional information about words ..................................... 15

    Word exposure, refinement and reinforcement ................................................................... 23

    Bilingual lexicon as an incremental, integrated system .................................................... 24

    Opportunities for receptive and expressive vocabulary instruction ................................ 27

  Review of Relevant Syntheses .................................................................................................... 29

Chapter 3: Methodology .................................................................................................................... 32

  Inclusion and Exclusion Criteria .................................................................................................. 34

  Design characteristics .................................................................................................................. 34
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 6</td>
<td>Synthesis and Discussion of the Findings</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Contextual approaches and breadth of curriculum-based vocabulary</td>
</tr>
<tr>
<td>1.5.2</td>
<td>Definitional approaches and breadth of general vocabulary</td>
</tr>
<tr>
<td>1.5.3</td>
<td>Breadth in Knowledge of Curriculum-Based Vocabulary</td>
</tr>
<tr>
<td>1.5.4</td>
<td>Balanced approaches and breadth of curriculum-based vocabulary</td>
</tr>
<tr>
<td>1.5.5</td>
<td>Contextual approaches and breadth of curriculum-based vocabulary</td>
</tr>
<tr>
<td>1.5.6</td>
<td>Definitional approaches and breadth of curriculum-based vocabulary</td>
</tr>
<tr>
<td>1.5.7</td>
<td>Depth in Knowledge of Curriculum-Based Vocabulary</td>
</tr>
<tr>
<td>1.5.8</td>
<td>Definitional approaches and depth of curriculum-based vocabulary</td>
</tr>
<tr>
<td>Fixed Variables Relating to Learner and Setting Characteristics</td>
<td>1.5.9</td>
</tr>
<tr>
<td>1.5.10</td>
<td>Demographic and contextual factors</td>
</tr>
<tr>
<td>1.5.11</td>
<td>Linguistic and cultural factors</td>
</tr>
<tr>
<td>Manipulated Variables Relating to Vocabulary Instruction</td>
<td>1.5.12</td>
</tr>
<tr>
<td>1.5.13</td>
<td>Vocabulary teaching as an instructional factor</td>
</tr>
<tr>
<td>1.5.14</td>
<td>Factors relating to instruction in treatment-comparison group contrasts</td>
</tr>
<tr>
<td>1.5.15</td>
<td>Contextual and definitional support as a factor in word learning</td>
</tr>
<tr>
<td>Response Characteristics</td>
<td>1.5.16</td>
</tr>
<tr>
<td>1.5.17</td>
<td>Baseline equivalence</td>
</tr>
<tr>
<td>1.5.18</td>
<td>Study outcomes</td>
</tr>
<tr>
<td>Implications</td>
<td>1.5.19</td>
</tr>
<tr>
<td>1.5.20</td>
<td>Limitations of Study</td>
</tr>
<tr>
<td>1.5.21</td>
<td>Conclusion</td>
</tr>
<tr>
<td>References</td>
<td>1.5.22</td>
</tr>
</tbody>
</table>

ix
List of Figures

Figure 1. Flowchart summarizing study searches............................................................ 40

Figure 2. Chart displaying overall and differential attrition levels............................ 43

Figure 3. Framework for assessing studies’ levels of evidence standards.................. 51

Figure 4. Effect sizes and confidence intervals for studies assessing breadth of general vocabulary ................................................................................................................................. 92

Figure 5. Effect sizes and confidence intervals for studies using curriculum-based measures to assess breadth of word knowledge................................................................. 95

Figure 6. Effect sizes and confidence intervals for studies using curriculum-based measures to assess depth of word knowledge.......................................................... 100
List of Tables

Table 1. Design and outcome characteristics used to evaluate levels of evidence for effectiveness........................................................................................................................................................................... 85
List of Appendices

Appendix A: Outcome Characteristics for Included Studies ........................................ 137
Appendix B: Figures Displaying the Proportion of Studies by Outcome Characteristics ... .................................................................................................................. 141
Appendix C: Curriculum-Based Measures ................................................................. 142
Appendix D: Learner Characteristics for Included Values ............................................ 144
Appendix F: Treatment Characteristics for Included Studies at the Study Level ........ 151
Appendix H: Proportion of Studies Meeting Evidence Standards by Level................. 158
Appendix I: Effect Sizes for Studies Meeting Evidence Standards Using General Vocabulary Measures to Assess Definitional Knowledge.................................................. 159
Appendix J: Effect Sizes for Studies Meeting Evidence Standards Using Curriculum-Based Measures to Assess Breadth of Word Knowledge..................................................... 162
Chapter 1 Introduction

This dissertation presents the results of a systematic review on the effectiveness of vocabulary instruction for bilingual learners who are frequently referred to as English language learners (ELLs) and arrive at school with varying linguistic and cultural knowledge in their two languages. Chapter 1 provides the background, statement of the problem and the significance of the study. Chapter 2 reviews the relevant first and second language literatures for vocabulary teaching and learning. Chapter 3 details the methodology, and Chapter 4 presents the results of the various analyses. Lastly, Chapter 5 discusses how the results address the research questions as well as the limitations and future research suggested by this systematic review.

Background and Statement of the Problem

Emergent bilinguals encompass those learners who develop two or more languages through their experiences in their homes, communities and schools (Garcia, Kleifgen, & Falchi, 2008). Due to the emphasis of English as the medium of instruction in schools across the United States, emergent bilinguals are often identified as English language learners (ELLs) but represent a highly diverse group of students that know English as well as their first language to varying degrees when entering schools and come from an environment where a language other than English is dominant. The ELL\(^1\) category has certain limitations as its use devalues other languages and places English in a greater position of legitimacy (Garcia, Kleifgen, & Falchi, 2008). Although nearly one in ten students are classified as ELLs in the U.S.—4.6 million ELLs of an approximate

\(^1\) The label “English language learner” (ELL) will be used throughout this dissertation as it is the commonly accepted, popular term in discussions of certain emergent bilingual populations, their designation, and the policies surrounding them (Garcia, Kleifgen & Falchi, 2008).
total of 55.5 million students (U.S. Department of Education, 2014)—this category fails to fully represent the dynamic nature of these students’ linguistic and cultural knowledge as they progress through their schooling. Based on data gathered by the National Assessment of Educational Progress (2013), ELLs continue to fall well below reading achievement levels on English reading assessments in relation to their monolingual counterparts who have steadily shown a 40- to 50-point advantage from 1998 to 2013. However, unlike other subgroup classifications used in national assessment data, the ELL category demonstrates less stability over time: higher-performing ELLs may transition out of the ELL category upon meeting certain achievement criteria and thus reclassified while lower-performing ELLs persist in this category. Additionally, those students who may be formerly designated as ELLs may not potentially fare much better in the long-term despite having previously met achievement criteria that would enable them to exit the ELL category (National Center for Education Services, 2009). As a result of the influx of newly classified ELLs and efflux of ELLs meeting achievement criteria, the persistent disparity remains and the ways in which to best address the learning needs for these diverse learners who vary in levels of proficiency continues to be a challenge in today’s classrooms.

Vocabulary development in the first years of formal instruction in particular is a key factor that contributes to the achievement gap between those who do well and those who do not. As evidenced by Cunningham and Stanovich (1998), vocabulary knowledge is crucial to later development as it significantly predicts reading comprehension in middle and high school grades. For ELLs whose exposure to their second language (L2) vocabulary (i.e., English) may have been heavily rooted in their schooling experiences
unlike their native English peers, they may correspondingly face potential challenges upon encountering the greater academic and language demands as they progress through the upper grades. Because language and culture are inextricably bound (Nieto, 2004), ELLs may also face fewer opportunities to express vocabulary knowledge developed through their everyday experiences when solely receiving instruction in English and having limited opportunities to bridge their two languages. Without the appropriate supports for these diverse and unique learners who may have fewer opportunities for exposure to English words than their monolingual counterparts in the earliest years, there may be longstanding effects on later learning as they must predominantly rely on English to communicate for academic purposes.

**Centrality of vocabulary knowledge in academic performance.** ELLs, whose development pathways are unique and commonly not addressed in curriculum planning, are often faced with the enduring inability to access information from oral and written language at the same rates as their native English-speaking peers (Laufer and Yano, 2001). For many of these students, starting school with less developed English vocabularies will have longstanding effects on their later academic success as well as meeting standard levels of English proficiency (White et al., 1990). ELLs like their native English-speaking counterparts must develop the same five major areas contributing to reading success: phonemic awareness, phonics, fluency, vocabulary, and comprehension (National Reading Panel, 2006). Although some of these are developmental skills that transfer across language and can be fully mastered, vocabulary knowledge remains as the most common source of difficulty faced by ELLs as they encounter increasing reading demands and unfamiliar, complex words in their second
language (Droop & Verhoeven, 2003), persisting as the primary cause of academic failure in grades 3 through 12 (Bauman and Kameenui, 1991). Furthermore, even though a deficit in low-order skills (e.g., decoding, reading fluency, word recognition) during adolescence may disrupt comprehension of grade-level texts, researchers contend that the greatest and most widespread challenge among this group is accessing vocabulary and semantic knowledge (Kamil, 2003; Snow et al., 2007). As demonstrated by report by the National Literacy Panel on Language-Minority Children and Youth, through instruction, word-level components (e.g., decoding, spelling) are or can be developed to the levels of their monolingual peers (August & Shanahan, 2006). Though, this is not the case for text-level skills, like vocabulary and reading comprehension, which require rich and extensive exposure over time.

For ELLs, a delay in precursor literacy skills such as oral language proficiency may also constrain future language and literacy development, thereby limiting access to many of the words already acquired by native English speakers. The crux of the matter lies in the great disproportion in depth of academic vocabulary between ELLs and native English speakers (August et al., 2005). The ensuing disproportionate growth between native English speakers and ELLs’ knowledge of words increases with time as their access to word-learning opportunities becomes hampered by increasing vocabulary demands (Verhoeven, 1990). Native English speakers enter their schooling with approximately 5,000 to 7,000 words before receiving formal reading instruction (Biemiller & Slonim, 2001) and acquire approximately 3,000 words on average every school year (Nagy & Anderson, 1984). Although there is a dearth of information on the average size of their vocabularies, previous research has revealed that there is a large
range in the number of known words among them and that the vocabulary gap between them and their native English counterparts endures (August, Carlo, Dressler & Snow, 2005; Proctor, Carlo, August & Snow, 2005; Swanson, Saez & Gerber, 2006). Assessing individuals’ vocabulary knowledge poses a challenge especially for ELLs due to a greater number of variables influencing their lexico-semantic development, or word-meaning acquisition. Investigating the multifaceted ways vocabulary instruction impacts ELLs’ vocabulary development will elucidate how varying approaches and the interplay of variables influence students’ word-meaning acquisition.

Although most research in the area of vocabulary instruction is predominantly conducted with English-only students, there has been strong support for ensuring that all students’ oral vocabularies are expanded as it will consequently contribute to their reading vocabulary (i.e., words they understand when reading a text) when they grow older (Hiebert, Pearson, Taylor, Richardson, & Paris, 1998; Snow, Burns & Griffin, 1998; Sticht & James, 1984). Although learners’ listening and speaking competence develops in advance of their reading and writing competence, researchers have more recently emphasized teaching vocabulary and developing early reading skills concurrently as a way to introduce young learners to unfamiliar vocabulary before being able to comprehend them in texts on their own.

**Opportunities for vocabulary learning.** Vocabulary instruction is an integral part of students’ knowledge acquisition throughout their entire schooling, and word-learning activities in the years prior to formal reading instruction in particular can provide opportunities to enrich oral language—the basis for understanding and producing written language (i.e., reading and writing)—in ways that are sensitive to ELLs’ linguistic
abilities and background knowledge. More emphasis made on acquiring word meanings in primary grades, can avoid deferring vocabulary learning at a pivotal point in young learners’ education and give preference to teaching words that meet instructional goals while also enhancing early reading skills (i.e., decoding) (Biemiller, 2005). This is possible by broadening instruction to encompass phonics- and meaning-based instruction when necessary. Instruction that emphasizes making print accessible and highlighting oral vocabulary students may already know to the neglect of learning more difficult, unknown words, may not contribute to the their long-term vocabulary growth, which may hinder also their knowledge acquisition upon encountering increasingly demanding tasks throughout the upper grades. In the relatively few studies examining the contributions of oral language on later reading comprehension, studies of ELLs in elementary school have found consistently that oral language proficiency in English and English reading comprehension are positively correlated (Beech & Keys, 1997; Carlisle, Beeman, Davis and Spharim, 1999; Peregoy & Boyle, 1991). Consequently, as ELLs take on learning tasks with less teacher-mediated support, they may also benefit from learning about word meanings that will enable them to engage in increasingly student-centered literacy activities, such as independent reading and writing activities, and to negotiate word meanings more efficiently through receptive and productive means (Davis, Carlisle & Beenan, 1999; Dufva & Voeten, 1999). Interventions with multiple components that provide an integrative approach to vocabulary teaching at any grade level can therefore highlight word meanings and their applications using various modalities (e.g., listening, speaking, reading, writing) throughout students’ learning. Due to the dearth in vocabulary research, however, more narrowly focused interventions can also prove
beneficial in identifying specific elements that may impact ELLs’ word learning with these diverse learners.

Given the short instructional time devoted to vocabulary instruction in most classrooms (Biemiller, 2005), as well as its influence on children’s word learning and literacy acquisition (National Reading Panel, 2000), students are faced with learning tasks that may require increasing contextual support during oral (Short & Echevarria, 2005) and written communication (Schwanenflugel, Stahl & McFalls, 1997) in order to acquire a considerable proportion of vocabulary. As such, the study aims to investigate the effectiveness of varying word-teaching approaches for improving receptive and expressive vocabulary outcomes for ELLs, intending to reveal through a systematic approach the multilayered ways that instruction can impact ELLs’ lexico-semantic development in an effective and efficient manner.

**Opportunities to draw on linguistic and cultural resources.** ELLs receiving formal instruction in a second language bring with them existing linguistic and cultural resources that support L2 word learning. In particular, ELLs are able to draw on their existing knowledge in their L1 to recognize the cross-linguistic relationships between their two languages as they assimilate information within a conceptual framework (August, Carlo, Dressler & Snow, 2005). Vocabulary instruction that has integrated L1 supports have shown to benefit ELLs as it permits them to access underlying conceptual knowledge (Avila & Sadoski, 1996; Carlo et al., 2004; Li, 2013; Proctor et al., 2005; Ulanoff & Pucci, 1999). To this end, the ability to access information relating to their existing vocabulary knowledge may be contingent on the instructional supports available to them. Previous reviews examining the effects of instruction on word learning have
primarily focused on monolingual populations (Marulis & Neuman, 2010; Stahl & Fairbanks, 1996). This study examines the effects of instruction on ELLs’ vocabulary learning, as well as contributes to the literature on the opportunities afforded to students to draw on native languages to expand their vocabulary repertoire in the targeted language.

**Purpose**

This systematic analysis explores the effects of multiple vocabulary instructional approaches geared toward enhancing ELLs’ word-meaning acquisition. The results of this study will help to inform the field on the most responsive learning approaches to the needs of ELLs who exhibit varying language and literacy needs. In addition, this analysis will also point to issues in need of further investigation after examining studies’ level of evidence for effectiveness by presenting the many factors may impact students’ word learning. The findings will present the effectiveness of instruction on ELLs’ receptive and expressive vocabulary knowledge through oral and written communication.

**Significance of the Study**

The results of the systematic review will offer a synthesis of multiple studies and is intended to contribute to an expanding literature on specific instructional approaches deemed responsive for bilingual learners identified as ELLs in US schools. Previous studies have begun to explore how various vocabulary-related instructional approaches enhance ELLs’ development of receptive and expressive vocabulary knowledge. However, research syntheses have not been used to appraise the existing literature and translate the results of intervention studies into well-defined ways to promote word learning among ELLs. Standardized mean effect size estimates for vocabulary outcomes
will provide systematic evidence of the effectiveness of vocabulary instruction with this population across studies meeting sufficient levels of evidence standards. Additionally, this study enables an analysis of factors that are important to consider when interpreting current and future research and when considering instructional decisions that might arise during the implementation of vocabulary teaching in actual classroom contexts.

**General Research Questions**

The systematic review in this dissertation aims to respond to two primary research questions, which will help guide the search and analysis of the findings. These specific questions are revisited in Chapter 3 after variables of interest have been discussed in the literature review in Chapter 2.

1. How do intervention studies developing L2 word-meaning acquisition with ELLs in pre-kindergarten through 6\textsuperscript{th} grade settings vary in their methodological characteristics (i.e., research design, learners, instructional design and outcome)?
   - What are the studies’ levels of evidence for effectiveness based on their design and outcome characteristics?

2. What is the impact of studies that meet requisite evidence standards on learners’ L2 word-meaning acquisition?
Chapter 2: Literature Review

As indicated, the purpose of this systematic review is to synthesize the empirical literature on the effectiveness of vocabulary instructional methods for English language learners in preK-6th grade settings. The most relevant literature to review consists of previous meta-analyses and quantitative studies of vocabulary instruction; however, important studies in this area, particularly cited reviews and syntheses, are included to ensure that relevant instructional, theoretical and empirical variables are also investigated. This literature review explores the current research relating to the vocabulary-learning task and the factors leading to word-meaning acquisition, which will inform the coding framework discussed in the subsequent chapter.

How is Vocabulary Learned?

Vocabulary knowledge refers broadly to an intricate system of concepts that are acquired in unique ways based on their varying degrees of complexity and use. Therefore, when thinking of students’ lexical development, increasing students’ level of word understanding is rendered more difficult by the many dimensions that characterize the vocabulary learning process. To gain a thorough understanding of a word, learners must develop depth of knowledge that includes learning how a word’s phonology relates to its orthographic representation; the syntactic structures into which the word form enters; its potential for collocations, derivational processes and polysemy; its morphological structure; a word’s semantic richness; its connotations; and knowledge of pragmatics (Paribakht & Wesche, 1993). Verhallen & Schoonen (1993) have also described the multifaceted nature of vocabulary learning as it entails understanding a word’s literal
meaning; various connotations; semantic associates; morphological options it offers; multiple meanings; and its syntactic rules.

To increase the depth of one’s word knowledge, learners must be exposed to and use words in multifaceted ways and across multiple contexts. Nagy and Scott (2000) highlight the many dimensions and complexities that may impact vocabulary acquisition: the role of word types (i.e., function words versus content words); the ability to use words in multidimensional ways (e.g., recognizing a word versus using a word expressively); opportunities for incremental exposure to refine a word’s meaning; words’ polysemous nature whereby meanings may show subtle and vast disparities; the extent to which words are interrelated to other words; and opportunities to apply word-learning strategies.

By providing more than a single encounter, context and source of lexical information learners are able to engage in deep processing of words (Beck, McKeown & Kucan, 2002). Similarly, Graves (2007) also indicates multiple components that are characteristic of robust vocabulary teaching with monolingual English learners: rich language and word experiences to enhance incidental word learning, direct word teaching, instruction of word-learning strategies, and word-consciousness cultivation. With this in mind, learners are able to focus on diverse and rare words throughout literacy-based interactions (e.g., shared readings, text-based discussions) guided by their instructors to develop the semantic richness of novel, sophisticated vocabulary and text structures, which are infrequent in students’ everyday linguistic environment. However, for ELLs, vocabulary acquisition becomes even more complex as multiple factors, such as the
breadth and depth of their vocabulary knowledge, the levels of proficiency in students’
L1 and L2 and the learning context, may impact the ability to acquire word meanings
(Kaivanpanah & Alavi, 2008).

**Broader Scholarly Literature: Enhancing Students’ Lexico-Semantic Knowledge**

Researchers in area of vocabulary instruction have applied a variety of
instructional methods to achieve L2 vocabulary acquisition among ELLs. Attention to
exclusively focusing on teaching word-learning strategies that would serve as an impetus
for recalling and retaining vocabulary knowledge has been narrowly investigated (Avila
& Sadoski, 1996; Zhang & Schumm, 2000). Several researchers emphasize teaching
words in unconventional ways through television viewing (Uchikoshi, 2006) as well as
multimedia texts with embedded instruction (Proctor et al., 2005) and dictionary-based
language supports (Li, 2013). Storybook-based interventions are also commonly used to
test teacher-centered instructional practices on word learning (Collins, 2004; Lugo-Neris,
Jackson, & Goldstein, 2010; Ulanoff & Pucci, 1999; Silverman, 2007). Dialogic
vocabulary-focused interventions that include reinforcement activities may also provide
rich, varied exposures to words across multiple learning contexts (Carlo et al, 2004;
Cohen & Byrnes, 2007; Cohen, Kramer-Vida, & Frye, 2012; Dockrell, Stuart, & King,
2010; Lesaux, Kieffer, Faller, & Kelley, 2010; Roberts & Neal, 2004; Taboada &
Rutherford, 2011; Townsend & Collins, 2009; Vaughn et al., 2009). Instruction
promoting ELLs’ ability to not only discuss and refine word meanings through oral
interaction but also to enable students to engage in various word-focused literacy
practices varies from study to study.
Type of Learning Tasks

The type of learning tasks may affect the amount of vocabulary knowledge that is gained (Griffin & Harley, 1996; Stoddard, 1929). ELLs who are taught through adult-directed practices typically engage in vocabulary learning through listening activities that are also constrained by the types of vocabulary used. For instance, simply engaging in everyday conversations exposes students to a limited pool of words they will learn to understand (i.e., receptive vocabulary) and eventually produce (i.e., expressive vocabulary) (Justice, Meier, and Walpole, 2005). Interventions conducted with native English students have demonstrated that by simply listening to stories they are able to acquire unfamiliar words (Robbins & Ehri, 1994) and even improve their expressive vocabulary (Senechal and Cornell, 1993). On the other hand, for ELLs, accessing meaning from their L2 requires them to use grammatical and semantic cues that may be surrounded by a greater number of unfamiliar words. For ELLs who exhibit limited expressive vocabulary knowledge, exposing them to print-based interactions can introduce them to sophisticated and complex language not typically used in their regular interactions (Crain-Thoreson, Dahlin & Powell, 2001). However, hearing these once inaccessible, unknown words used in context may enhance knowledge of these through mere exposure (Biemiller & Boote, 2006). Although rich and varied communicative experiences tend to expand students’ receptive and productive vocabularies (August & Shanahan, 2006), the magnitude of the effects of how adult-directed ways and more student-centered activities impact ELLs’ word learning is still uncertain. Interventions that apply dialogic approaches that afford opportunities to engage in receptive and
productive ways afford ELLs opportunities to negotiate and gain more complete and precise word meanings (Carlo et al., 2004; Cohen et al., 2012; Taboada & Rutherford, 2011).

In addition, many intervention studies emphasize exposing young ELLs to reading vocabulary well in advance of their oral vocabularies as a means of prematurely building semantic representations—prior to developing decoding skills—for typically unknown and inaccessible words (Dockrell et al., 2010; Silverman, 2007; Silverman & Hines, 2009; Valdasy et al., 2013). However, evidence has also shown that greater vocabulary knowledge from the outset may provide more proficient learners with a word-learning advantage as they are able to infer meanings more successfully with less scaffolding (Nicholson & Whyte, 1992).

**Rich, Varied Experiences with Words**

Some research contends that enriching one’s word knowledge in the earlier stages of development requires heavier attention to communication through oral experiences as a way of developing word knowledge prior to developing literacy skills. Many point to early literacy skills, such as phonemic awareness as strong predictors of a child’s decoding ability (Ball & Blachman, 1991; Perfetti, Beck, Bell & Hughes, 1987)—a discrete skill particularly crucial to vocabulary growth and early reading in the first years of schooling. Conversely, other studies report how vocabulary knowledge (e.g., knowing words and their meanings) can also contribute to low-order skills (i.e., phonemic awareness and word recognition skills; Nagy, Berninger, Abbot, Vaughan, & Vermeulen, 2003) et al., 2003; Nation & Snowling, 2004), which are also necessary for reading comprehension. Limited vocabulary knowledge has also shown to affect oral
comprehension and impede communication with others, which in turn constrains further vocabulary development (Gass & Selinker, 2008). The interdependent relationship between lower-order skills and vocabulary knowledge demonstrates the promising benefits of providing vocabulary instruction that integrates top-down, or comprehension-oriented approaches to vocabulary learning, typically neglected for lower-order decoding skills especially for young ELLs (Roberts & Neal, 2003).

**Providing contextual and definitional information about words.** Research has shown that the need for contextual support to facilitate comprehension is based on the level of concreteness of unknown words as it also affects students’ word-learning ability (Nagy et al., 1987). As such, contexts that present too many unfamiliar, complex lexical items can hinder students’ word learning. Alternatively, contexts that are deemed too easy may also fail to expand students’ lexicon (Carlisle & Katz, 2005). Therefore, the interplay between students’ L2 proficiency and the conceptual difficulty of words found in learning activities may hinder their comprehension, thereby limiting the growth of their English lexicon. At any given time, there are precise words of varying conceptual difficulty, or concreteness, that all students need to know (i.e., receptive vocabulary) and use (i.e., expressive vocabulary) in order to comprehend and participate in school subjects with little interruption (Anderson & Nagy, 1992). Exposing ELLs to information-rich literacy environments lays the foundation for a wider range of learning and teaching about words. Through the exposure of vocabulary-rich contexts that are familiar to them, students are introduced to novel vocabulary in ways that are accessible while also enriching conceptual knowledge of known words. In this vein, without explicit attention to the meanings of words, ELLs are able to infer meanings of surrounding
words when a large proportion of contextual information is comprehensible (Neuman & Koskinen, 1992; Uchikoshi, 2006). When this is not the case, instructional practices that target specific words can facilitate their lexical and knowledge acquisition. Consequently, interventions that supply appropriate context-rich mediums can promote indirect word learning that enable learners to derive meanings and add new words to their repertoires (Biemiller & Boote, 2004; Robbins & Ehri, 1994; Senechal, Thomas, & Monker, 1995), but the probability of acquiring word meanings varies based on the degree of contextual support afforded to the learner. Context-embedded instruction in which definitional information is provided can therefore provide additional support when contextual information alone is insufficient to promote understanding.

As evidenced by the National Reading Panel’s findings (2000), research on contextual approaches to vocabulary instruction resulted in insightful ways that have guided teaching about words over the past decade. Similar to the results found in Stahl and Fairbank’s meta-analytic review (1986), the uses of both definitional and contextual vocabulary instruction have emerged as an effective practice rather than solely using one of these methods. These guiding principles are based on the patterns emerging from vocabulary intervention studies relating to word learning among native English speakers. Determining the degree to which these areas of instruction should be fostered is still in question as it relates to ELLs who represent a diverse group with varying levels of L1 and L2 vocabulary and concomitant conceptual knowledge.

Currently, the degree to which contextual and definitional emphasis can impact ELLs’ vocabulary acquisition has not been thoroughly researched. In a comparison of incidental word learning and explicit teaching conditions, Johnson and Yeates (2006)
indicate that most information learned about words occurs through explicit instruction. Through the provision of definitional information, interventions are able to promote word learning by supplying additional semantic information (e.g., examples, non-examples, synonyms/antonyms, morphological variations, visual information) not gleaned from an immediate or individual context, which may increase the possibility for vocabulary growth (Biemiller & Boote, 2006; Carlo et al., 2004; Collins, 2004; Silverman, 2007, Proctor et al., 2007; Ulanoff & Pucci, 1996). Through intentional teaching (i.e., targeting words in explicit and implicit ways), ELLs are able to gain a richer semantic representation of a word not afforded by receptive routes. Stahl & Fairbanks (1986) also points out that explicitly teaching vocabulary in heavily context-dependent ways (e.g., through book-based interactions, thematic discussions) may reveal greater effects on curriculum-based measures during vocabulary instruction more so than any other method (e.g., decontextualized vocabulary instruction; incidental word learning). These findings are based on a few studies relating to native English participants, which may differentially impact ELLs’ vocabulary development as a result. Studies conducted with ELLs, however, reveal mixed results for ELLs as both decontextualized instruction (Avila & Sadoski, 1996; Rodriguez & Sadoski, 2000; Zhang & Schumm, 2000) and contextualized word learning (Neuman & Koskinen, 1992; Uchikoshi, 2006) yielding favorable results for ELLs.

*Incidental vocabulary learning.* ELLs’ exposure to language-rich environments provide opportunities for incidental word learning as they encounter words in varied oral or written contexts. Multicomponential studies that permit learners to engage with targeted words in multiple contexts may increase the possibilities of incidental word
learning and facilitate target word learning through discussions with peers and teachers (Dockrell et al., 2010; Taboada & Rutherford, 2011; Vaughn et al, 2009); by presenting learners with videos and illustrations based on a text (Roberts & Neal, 2004); supplying Spanish audio and written versions of an English text (Carlo et al., 2004); and providing independent readings with embedded target vocabulary (Roberts & Neal, 2004). Through word-leaning opportunities that afford incidental exposure, ELLs’ learning may require minimal support by teachers in order to promote students’ lexico-semantic development and to introduce novel words. Instruction can also expand students’ conceptual knowledge about words through oral readings, interactive discussions or by engaging in independent reading. While listening to oral reading, teachers may provide implicit instruction to support the development of conceptual structures to aid word vocabulary acquisition without overtly specifying learning goals (i.e., developing word knowledge) (Cohen et al., 2012). Similarly, Taboada and Rutherford (2011) emphasize the importance of using text-based questioning to promote discussions about previously encountered words without providing definitional information to students.

**Intentional vocabulary instruction.** Researchers and teachers delivering instruction may implement a variety of ways to target specific words throughout interventions. Reflective of the many dimensions of vocabulary knowledge, much of ELLs’ instruction may entail developing their phonological, syntactic, morphological and semantic knowledge of words. In addition, variations in the amount of contextual support may range from relying wholly on the context-dependent use of words to teaching a particular aspect of a word with no contextual support throughout instruction.
Decontextualized vocabulary instruction. Interventions that include decontextualized ways of learning words through definitional support aim to teach discrete aspects of words, independent of an immediate communicative context. Decontextualized techniques may include provide information about word’s phonological and orthographic aspects as well as teaching definitional material, with little assumption of shared background knowledge between the learner and instructor. To increase ELLs’ semantic knowledge, students can be supplied with formal definitions of target words (Cohen & Byrnes, 2007; Proctor et al., 2011; Zhang & Schumm, 2000); Spanish translations of novel L2 words (Avila & Sadoski, 1996; Proctor et al., 2011; Zhang & Schumm, 2000), as well as visual representations (i.e., drawings) of target words (Avila & Sadoski, 1996; Proctor et al., 2011; Zhang & Schumm, 2000). To improve students’ phonological awareness and to reinforce decoding skills, intervention protocols can be designed to present ELLs with a word’s pronunciation (Proctor et al., 2011; Silverman, 2007) or to include a blending and spelling activity as an integral component of an intervention to stimulate later word learning and text comprehension (Nelson, Vadasy, & Sanders, 2011).

Contextualized vocabulary instruction. Intervention studies may also provide varying degrees of contextual support to ELL’s word learning. ELLs provided with contextual support during literacy-rich experiences in which vocabulary instruction is embedded are given support for word-meaning acquisition as they are able to glean information about oral and written language. Context-bound vocabulary instruction, therefore, entails using information from a text to introduce and shape understandings of unfamiliar words with less reliance on individuals’ prior experience to access meanings.
However, to build students’ lexicon, techniques that may afford less contextual support could also be implemented outside familiar activities (e.g., listening to predictable story events) through direct instruction to enhance students’ depth of word knowledge, given that they may benefit from instruction that advances their knowledge of the multifaceted nature of words beyond definitional information found in their immediate context.

Interventions in which vocabulary instruction is purely context-dependent may focus on teachers’ explanations of words that expand on word meanings, strictly providing definitional information relevant to a story, and highlight words’ different semantic features as they appear in the same storybook (Lugo-Neris et al., 2010; Collins, 2004). Due to the recurring nature of targeted vocabulary in storybook contexts, teachers are able to elaborate on and reinforce initial word meanings. However, vocabulary instruction may be limited to teachers’ own explanations of words. In contrast, words may be explicitly taught and systematically presented throughout multiple contexts, enabling learners to develop their lexicon in diverse ways upon encountering words in both contextualized and decontextualized instruction (Carlo et al., 2004; Cohen & Byrnes, 2007; Dockrell et al., 2010; Proctor et al., 2011; Silverman, 2007; Taboada & Rutherford, 2011; Ulanoff & Pucci, 1999).

Teaching about words in authentic contexts can provide additional support to ELLs’ active meaning-making, such as when presenting words during read alouds, while independently reading multimedia texts, or in conversations (Silverman, 2007; Taboada & Rutherford, 2011). Word teaching may also include integrating both word-level and text-level skills due to words’ recurring possibilities in connected text and speech. For instance, words’ phonological representations may be highlighted as they are presented
during a teacher-led read alouds (Lugo-Neris et al., 2010; Silverman, 2007) and during multimedia texts that may provide immediate access and support a novel word’s pronunciation as it is encountered during independent reading (Li, 2013). Teaching about the orthographic representations of words may also be the focus of instruction when presenting students with a target word’s written form as it encountered during a shared reading (Cohen & Byrnes, 2007).

Instructional strategies can be employed to support semantic depth throughout rich content-related interactions. Extralinguistic strategies may be utilized as a way to develop learners’ visual concepts of words through gestures to scaffold ELLs’ understandings of target words as they appear in a story (Collins, 2004; Silverman, 2007) or providing visual supports, such as pointing students’ attention toward illustrations or using realia, to label vocabulary items found in a story (Cohen et al., 2012; Collins, 2004; Silverman, 2007). Oral vocabulary teaching strategies provide a means of explicitly teaching about definitional information relating to target words as they are discussed in an interactive, scaffolded conversation (Nelson et al., 2011; Kieffer and Lesaux, 2012) and as a way to build on story-based meanings of words (Cohen & Byrnes, 2007; Collins, 2004). Semantically-related words (i.e., synonyms, antonyms) can also be introduced in context-dependent ways to expand on the meanings of words found in a text (Collins, 2004; Silverman, 2007). Opportunities to learn words in different contexts can also enrich semantic representations of words while concurrently developing syntactic awareness. These efforts may be achieved by relating words to personal experiences that are relevant to students (Collins, 2004; Silverman, 2007) and using the target words in example sentences (Proctor et al., 2011; Silverman, 2007).
Generative vocabulary strategies. To increase learner autonomy and enhance independent word learning, intervention studies can reinforce the use of generative vocabulary strategies. By adopting these methods in vocabulary-centered interventions, students are able to engage in certain practices that could potentially lead to incidental vocabulary acquisition and less reliance on teacher support. One common method is providing ELLs with opportunities to utilize dictionaries as a resource to access multiple meanings of words and learn about the importance of selecting appropriate definitions to fit the learning context (Carlo et al., 2004; Li, 2013; Proctor et al., 2011).

Instruction can focus on strengthening students’ word analysis skills through morphemic, cognate and contextual analysis (Carlo et al., 2004; Kieffer & Lesaux, 2012). By developing students’ word analysis skills, ELLs can learn of the ways to derive an unfamiliar word’s meaning by examining its meaningful parts (e.g., prefixes, root, suffixes), all in support of their morphological awareness—to recognize the stems of morphologically complex words and their semantic relationships to other morphological forms as well as to develop an awareness of their syntactic possibilities. Cognate awareness is enhanced through instruction that directs ELLs’ attention to etymologically-related words across languages to instruct on how words may relate in meaning, spelling and phonology (Carlo et al., 2004; Proctor et al., 2011) to bolster understanding of an L2 word when conceptual information is available in their L1. Contextual analysis may be developed when students are afforded opportunities to infer the meanings of unfamiliar words by scrutinizing surrounding text (Carlo et al., 2004; Lesaux et al., 2010). Alternatively, ELLs can be also taught the keyword method, a memory technique implemented to teach students how to associate a pair of lexical items (i.e., one known
and one unknown word) to corresponding visual images and using the known word—the keyword bearing a phonetic similarity to the novel word—as a means of recalling words and facilitate retention.

**Word exposure, refinement and reinforcement.** Rich, extended vocabulary instruction has shown to expose ELLs to target words frequently within and across lessons and allows them to manipulate words in varied and meaningful ways (Carlo et al., 2004; Proctor et al., 2005; Roberts & Neal, 2003; Valdasy et al., 2013; Vaughn, 2009). Instruction with limited opportunities for exposure and contextual use may expectedly restrict students’ vocabulary knowledge to the ways they have learned about words. For instance, repeated exposures in literacy-based interactions that tend to reinforce students’ understandings of words in activities like shared readings (Biemiller & Boote, 2006) may only show gains for familiar concepts without exhibiting significant increases in their knowledge of unfamiliar concepts (Leung, 1992). In these cases, however, students can rely on the contextual supports afforded to them during oral (Echevarria, J., Vogt, M.E., & Short, D., 2004) and written communication (Schwanenflugel, Stahl, & McFalls, 1997) in order to make meaning when encountering unfamiliar words. The conceptual difficulty of words can play a critical role even when presented in context; when a greater number of words are presented, a learner’s incidental word learning can be stifled (Nagy et al., 1987), which can expectedly be challenging for ELLs whose cognitive resources become overwhelmed as they attend to lower-order skills upon encountering more unknown words than their native-English counterparts. In contrast, vocabulary learned in multiple contexts can assist learners in discovering their logical relationships with other words and how they function in different contexts (Stahl and Kapinus, 2001). More
recent research in the form of multi-componential interventions has emphasized the use of varying levels of contextual support in order to provide contextualized and decontextualized word-learning opportunities. These studies provide additional information about words beyond a single context to introduce and/or refine word meanings for students. On the other hand, studies that assess the effectiveness of less comprehensive yet more focused approaches (i.e., testing a single strategy) help to reveal task-specific influences on learners’ vocabulary learning. However, research remains scant on the impact of multicomponential interventions as compared with more discrete approaches to vocabulary instruction in developing ELLs’ receptive and expressive vocabulary.

**Bilingual lexicon as an incremental, integrated system.** When providing systematic ways of teaching and/or exposing students to words, one’s instruction should be responsive to the students’ linguistic and cultural needs (Biemiller & Boote, 2006). ELLs are more likely to possess a less integrated system of word meanings in their L2, which deter them from focusing on knowledge acquisition and word relationships as they mainly learn individual word meanings in discrete ways (Soderman, 1993). Due to varying bilingual proficiencies among ELLs, it is important to consider their level of exposure and use with words. Incidentally learning a word for a new concept has been described as less likely for students who are in the process of learning a label for an unknown concept rather than learning a label for a known concept (Nagy, 1997; Nagy et al., 1987). As such, ELLs with more developed L1 proficiencies who have developed conceptual knowledge in their L1 may differ in the type of support needed to further enrich semantic networks. In addition, the degree to which a novel word may be
conceptually difficult could hamper its acquisition (Nagy et al., 1987). Ensuring that ELLs are given multiple opportunities to learn words in a variety of rich contexts through which they are able to manipulate their various forms and enrich their semantic understandings of them will help them accumulate an understanding of the range in usage of a given word (Nation, 2001). In the same vein, Graves (1987) explains that words that are semantically related (e.g., *ingest/consume*) to a known concept (e.g., *eat*) are much easier to learn than words representing an entirely new concept.

For ELLs who enter their schooling with varying levels of native and English proficiency and diverse background knowledge, finding effective ways to expand their vocabulary may depend on the best way to structure learning tasks, given their varied exposures to words. ELLs’ lexical systems have been described as separate, which gradually map onto shared semantic representations through experience and use with words (Kroll & Curley, 1988; Kroll & Sholl, 1992). Since ELLs tend to show less rich semantic representations of words in their L2 than monolinguals due to limited exposure in that language (Verhallen & Schoonen, 1998), vocabulary instruction geared toward building on existing conceptual knowledge (i.e., in their L1 as well as their L2) may hasten their learning as these learners are able to gain information about the same or similar concepts through two language routes. With a common underlying conceptual base, ELLs can learn to restructure existing knowledge belonging to a particular label in their L1 under the influence of new knowledge gained through L2 (Cummins, 2000). As such, ELLs receiving formal instruction in a second language bring with them existing linguistic resources that also support L2 word learning. In particular, ELLs are able to draw on their existing knowledge in their L1 to recognize the cross-linguistic
relationships between their two languages. Through explicit instruction, ELLs can be taught to recognize cognate relationships as a means of inferring the meanings of unknown words as they become familiar with the phonological, morphological, and semantic overlap between words across their two languages (Carlo et al., 2004; Proctor et al., 2005). ELLs who speak non-alphabetic languages can also benefit from instruction that provides a more direct route to semantic information already developed in their L1 (e.g., through dictionary-based language supports; Li, 2013). Consequently, ELLs are able to enhance their literacy experiences in support of word learning through the use of their native language.

**L1 bridging strategies.** Instruction can be designed to purposely draw from ELLs’ existing knowledge to bridge students’ understandings in their L1 to bolster L2 word learning. To support this cross-linguistic transfer, L1 bridging strategies can include using explicit instruction (e.g., through translation techniques); by implicit teaching techniques (e.g., through bilingual interactions without direct word instruction); and other ways to promote incidental word learning across both languages. Text-level comprehension can be improved by permitting students to preview Spanish versions of an English read aloud (Carlo et al., 2004), through concurrent translations of an English read aloud to facilitate understanding as a story progresses (Ulanoff & Pucci, 1999), or by providing ELLs with a multimedia bilingual dictionary through which one could search for an unknown word’s translation and definition in both L1 and L2 as it was encountered in an English text (Li, 2013). In contrast, definitional approaches with less contextual support could also be used to also draw attention to English words and their translations and Spanish outside their contextual use (Carlo et al., 2004; Proctor et al., 2011; Zhang &
To ensure ELLs develop the conceptual knowledge needed to learn L2 target words, instruction can also focus on building ELLs’ conceptual understanding of a word in their L1 that will serve as an impetus for further English word-meaning acquisition (Avila & Sadoski, 1996; Zhang & Schumm, 2010).

**Opportunities for receptive and expressive vocabulary instruction.** Instruction can exhibit variations in the degrees to which ELLs’ word-level processing is enhanced while engaging in various receptive and expressive modalities. These variations may be explained by the types of learning activities ELLs engage in as they construct their vocabulary knowledge. Due to predetermined learning objectives found in teacher-driven activities, this instruction may limit opportunities for comprehensible input and output from ELLs (Avila & Sadoski, 1996; Collins, 2010; Uchikoshi, 2006; Ulanoff & Pucci, 1999; Zhang & Schumm, 2000). However, instruction can be designed to increase the possibilities for ELLs to monitor their learning and consider the incremental nature of their word-meaning acquisition. Multi-componential interventions through their subcomponents can include learner-centered and interactive activities that support students’ receptive (i.e., through listening or reading activities) as well expressive ways of learning words (i.e., through oral interactions or writing) (Carlo et al., 2004; Cohen & Byrnes, 2007; Cohen et al., 2012; Dockrell, 2010; Lugo-Neris et al., 2010; Dockrell et al., 2010; Lesaux et al., 2010; Li; Nelson et al., 2011; Proctor et al., 2011; Roberts & Neal, 2004; Silverman, 2007; Taboada & Rutherford, 2011; Townsend & Collins, 2009; Vaughn et al., 2009).
Dialogic opportunities for learning words can promote both receptive and productive knowledge of words. During these interactions, ELLs are able to orally communicate and listen to other’s ideas about word meanings prior to encountering their contextual use in ensuing activities (Carlo et al., 2004; Cohen & Byrnes, 2007; Cohen et al., 2012; Dockrell et al., 2010; Lesaux et al., 2010; Nelson et al, 2011). Developing ELLs’ depth of word knowledge can be an objective throughout students’ collaborative efforts as they share background knowledge while brainstorming word meanings (Lesaux et al., 2010) or by acting out scenarios centered on a particular word (Dockrell et al., 2010; Silverman et al., 2007). Interactive discussions in can also aimed to develop ELLs’ ability to infer meaning from written words as they discuss words’ morphological forms and how inflections influence word meaning (Lesaux et al., 2010).

Interactive discussions can also take place to reinforce students’ word knowledge following their explicit instruction as a way to stimulate productive vocabulary knowledge. Implementing word reinforcement activities may support students’ recall and retention of words meanings previously encountered during instruction. To reinforce students’ understanding of the meaning of target words learned in a previous instructional context, students can revisit words and demonstrate their understandings through scaffolded oral interactions with others (Carlo et al., 2004; Nelson et al., 2011; Silverman, 2007).

Interventions that integrate written tasks may also increase opportunities for ELLs’ receptive and productive vocabulary learning. In their writing, students are able to explore meanings of words, representing their visual concepts through drawings (Cohen & Byrnes, 2007) and writing down words’ definitions (Carlo et al., 2004; Cohen &
Byrnes, 2007). ELLs are able to demonstrate productive knowledge of words’ grammatical function, syntax and meaning through more complex activities as they provided example sentences using a target word (Cohen & Byrnes, 2007), responding to story comprehension questions in which target words are highlighted (Proctor et al., 2011; Nelson et al., 2011) and write statements in which they describe their personal connections to words (Proctor et al., 2011).

**Review of Relevant Syntheses**

In one of the earliest reviews relating to vocabulary instruction with monolingual English students, Petty, Herold and Stoll (1968) investigated the role of vocabulary but did not have enough data available to indicate the best single method of vocabulary instruction; however, they concluded that using several different techniques seemed more effective than those involving a single technique. This has been a recurring trend in ensuing decades, even upon the emergence of meta-analyses, which have been increasingly used to address the effects of vocabulary instruction on literacy outcomes, particularly in the area of reading comprehension and with native English speakers. Mezynski (1983) found factors relating to the successful implementation of vocabulary teaching methods in order to improve students’ reading comprehension: the amount of practice and active processing of words as well as the breadth of instruction in using targeted words. Similarly, by examining the contributing nature of word teaching, Stahl and Fairbanks (1986) examined the effects of vocabulary pre-instruction as a way to improve reading comprehension. In this study, vocabulary had strong effects on the comprehension of passages containing taught words, but also had slightly significant
effects on comprehension of passages without taught words, which indicates that instruction of both non-targeted and targeted vocabulary may generally improve reading comprehension.

In more recent years, meta-analyses have been conducted to understand the effectiveness of intervention on vocabulary outcomes with monolingual English students. Emphasizing the contextualizing support found in storybook reading interventions, Mol, Bus, and deJong (2009) explored how this approach impacted students’ oral vocabulary, reporting a moderate effect size for expressive vocabulary and their results also revealed large effects when executed by researchers rather than teachers. To date, however, only one meta-analytic study has investigated the effects of vocabulary instruction on word learning, particularly examining oral language outcomes among native English speakers (Marulis & Neuman, 2010). Through a moderator analysis, this meta-analysis revealed there were greater effects for trained adults providing the treatment, the combination of both explicit and implicit instruction rather than one method of instruction, and author-created measures compared to standardized measures. In addition, children from middle- and upper-income homes were more likely to benefit from vocabulary intervention than those from low socioeconomic backgrounds. This study also showed that all group configurations benefited equally and substantially from vocabulary instruction.

In the only meta-analysis examining the effectiveness of vocabulary instruction with English language learners, Won (2008) investigated how certain instructional supports, such as the degree of contextual support and the use of L1 bridging strategies, may improve achievement outcomes. In this study, varying degrees of contextual support provided medium to large effects for ELLs in ESL or EFL learning contexts. The level
of language proficiency also showed greater effect sizes for more advanced learners of English than those described as intermediate learners. Although this study shed light on important features in vocabulary instruction relating to bilingual populations, it did not explain the outcome types associated with these gains. No indication was made of whether these outcomes addressed vocabulary growth, only referring to the measures in a broader sense as achievement outcomes.
Chapter 3: Methodology

The study provides a systematic review to examine the effectiveness of vocabulary instruction on ELLs’ word-meaning acquisition and apply evidence standards to identify effective and promising approaches to vocabulary instruction. Although this study primarily presents a narrative review of relevant research relating to vocabulary instruction, this study also aims to provide a cohesive picture of the phenomenon by statistically integrating effect sizes and quantifying study findings when possible. These findings are compared by various study characteristics.

Results from preliminary analyses warranted the procedures followed throughout this study. Although efforts were taken to apply meta-analytic methods to combine and contrast results of the intervention effects on vocabulary outcomes that would enable finding sources of heterogeneity among those results through moderator analyses, an a priori statistical power analysis revealed insufficient power to detect a small amount of effect size heterogeneity. Expected statistical power for finding the significance of the overall mean effect under the random-effects model was computed for subgroups of effect sizes, representing the various dimensions of the outcomes under analysis in this study. The statistical power for finding a significant but small intervention effects, when a small degree of between-studies heterogeneity is assumed, ranged from 0.05 ($k = 2$ for depth of receptive, definitional knowledge assessed through curriculum-based measures) to 0.42 ($k = 11$ for breadth of receptive definitional knowledge assessed through standardized measures).

---

2 Subgroups of effect sizes were formed by their outcome objective (i.e. breadth versus depth of knowledge); type of knowledge (i.e., definitional versus contextual knowledge); and type of ability (i.e., receptive versus expressive knowledge).
Since no statistical significant amount of variation could therefore be detected among the subgroups of effect sizes, which in turn would exclude the possibility of examining whether characteristics of the studies are associated with any heterogeneity, a systematic review that was descriptive in nature was instead used to appraise available intervention studies given the small sample of studies that examined word-meaning acquisition among ELLs. Additionally, included studies will be evaluated based on their level of evidence for effectiveness. The impact of those studies that meet evidence standards will also be analyzed, and effect sizes will be investigated by the vocabulary instructional approaches utilized in these studies.

To ensure a rigorous and systematic approach to cumulating the data was taken and an unbiased description of the evidence was given, Cooper (1982)’s standard procedures for conducting a research review were followed, which consists of five stages: Problem formulation, Data collection, Data evaluation, Data analysis, and Report of results. First, tasks for the problem formulation (stage 1) stage, wherein the variables of interest are conceptually and operationally defined, and their relationships are specified, were described in Chapter 1. The literature review in Chapter 2 supports the data collection process (stage 2) that is be described in this chapter, wherein the relevant studies are located and information related to instructional, methodological and theoretical variables is gathered from the included studies.

In this chapter, details regarding the data collection stage are explained, presenting inclusion and exclusion criteria, the search procedures, study coding categories, coding procedures, and relevant statistical analyses. Specifically, the methods for locating the relevant studies are articulated, expressing retrieval techniques as well as
stating any problems encountered throughout the search process. By well delineating the search strategies and selecting inclusion and exclusion criteria, researchers are better able to minimize differences between the results of retrieved studies and those that may not be found. These criteria were consistently applied to all studies regardless of whether the results support or refute the hypotheses under investigation (Cooper, 2010). In addition, a number of indicators for data evaluation (stage 3) are discussed in this chapter. Finally, reporting of results (stage 5) followed by data analysis (stages 4) will be further discussed in the subsequent chapters for studies meeting evidence standards based on certain quality indicators derived from the What Works Clearinghouse (2014) Procedures and Standards Handbook.

**Inclusion and Exclusion Criteria**

To identify the relevant studies that address the research questions posited in this systematic review and determine which of the gathered evidence is usable, a set of inclusion/exclusion criterion were established in terms of study design, study setting, outcome measure, sample, and statistics.

**Design characteristics.** The focus of the studies selected addressed the multifaceted nature of vocabulary instruction, thereby teaching about meaning-based skills integral to vocabulary learning, which include instruction of morphological structures, definitional and contextual use of words, semantically related words such as paradigmatic (i.e., hyponymy, synonymy, antonymy) and syntagmatic relationships (i.e. collocations), and cognates. Therefore, studies implemented a variety of instructional activities, but some form of meaning-based instructional activity was required.
To cast a wide net due to the limited number of studies addressing lexico-semantic acquisition among ELLs, both experimental and quasi-experimental studies formed part of this study. In particular, a pretest-posttest control-group design was required for every study in order to determine how two or more groups of participants are compared for each time point of the desired outcome. Studies were able to vary in their instructional approach, implementing narrow as well as more comprehensive instructional methods to develop students’ word-meaning acquisition. Consequently, studies that included more than one vocabulary-focused intervention group were included. When the primary study author(s) did not identify a comparison group, a comparison group, providing a lesser degree of vocabulary-focused instruction (e.g., a diluted form of the treatment) or no vocabulary teaching, was therefore selected.

**Setting characteristics.** Although it would have been favorable to review studies that primarily serve the needs of English learners, it was necessary to reflect the realities of the various instructional program models in which English learners find themselves. For that reason, studies that provided vocabulary-related interventions in the varying contexts in which ELLs participate were examined, which include but are not limited to mainstream and bilingual classrooms that had heterogeneous as well as homogeneous samples. Studies that included ELLs with speech and language delays were excluded due to these learners’ unique developmental trajectories. Multiple grouping practices could have been used to target ELLs (i.e., whole-class, small-group, dyads, individual and mixed groupings).
**Outcome characteristics.** Lexico-semantic acquisition, which is the primary outcome of interest in this systematic review, was assessed using measures of receptive and/or expressive knowledge. Studies were included when testing for growth in number of single word meanings acquired for individual words (i.e., breadth) and in the number of multiple meanings, or semantic features, learned for particular words (i.e., depth). The instruments for the outcome measures assessed through standardized measures of general vocabulary (i.e., These words were not necessarily targeted throughout the intervention.) or more proximal measures (i.e., curriculum-based measures). Additionally, the types of knowledge assessed included definitional (e.g., through definition production measures, multiple-choice tests with a range of definitions as possible answers) and/or contextual knowledge (i.e., measured knowledge of a word’s use in context, such as through sentence cloze tests).

**Sample characteristics.** Participants included students identified as English language learners (ELLs), even though identification methods in each study may have varied. ELLs encompass a vast group of individuals that possess cultural and linguistic knowledge in their native language (L1) and who are also in the process of acquiring ways of speaking, writing and thinking in English. For the purposes of this study, ELLs who are portrayed as not meeting standard levels of English proficiency will form part of this study; this will be determined by the descriptions of their classroom setting (e.g., ESL/ESOL programs) and/or home language (i.e., primarily speaks a language other than English at home). Those students whose primary home language is not English but participate in mainstream classrooms were classified as ELLs due to the fact that there is evidence that illustrates an enduring vocabulary gap between former students in ESL
programs and native English speakers even when having reached standard levels of vocabulary proficiency (Laufer and Yano, 2001; National Center for Education Services, 2009). Additional labels describing these learners were also used throughout the search: bilinguals, dual language learners, limited English proficient (LEP), language minority students, and culturally and diverse learners (CLD).

All available studies conducted with ELLs were included regardless of the year of publication. The target language used in the vocabulary intervention must have been in English; therefore, ELLs from any language background can be included in the study as well as interventions taking place outside of the United States in English-medium contexts were also included. Student participants were selected based on their grade levels, ranging from prekindergarten to the end of primary grades (i.e., pre-kindergarten through 6th grade) in order to inform of ways to narrow vocabulary and concomitant conceptual gaps that exist throughout ELLs’ earlier years of formal instruction.

Summary. The following inclusion/exclusion criteria were used to locate the relevant studies included in this systematic review:

- The implemented intervention was designed to expose learners to contextual or definitional information as a means of improving word-meaning acquisition.
- A pretest-posttest control-group design was utilized in both experimental and quasi-experimental studies.
- A treatment group providing instruction through meaning-based activity (e.g., direct vocabulary instruction, opportunities for incidental word learning) in English was included.
• Studies assessed receptive and/or expressive vocabulary outcomes for targeted words and/or general vocabulary to evaluate an implemented vocabulary intervention.

• If not all participants were ELLs in treatment and comparison groups, more than half in each group were composed of ELLs. Studies with heterogeneous language groups in which ELLs were not the majority should have provided disaggregated data for ELLs.

• The participants were preschool-aged or in the elementary grades (i.e., kindergarten through 6th grade).

Search Strategies for Identifying Relevant Studies

To locate the relevant studies, a computerized search was first conducted by the author on the October of 2012, followed by an additional search on February of 2013. Journals that were searched included: *Applied Linguistics; Bilingual Research journal; Early Childhood Education Journal; Early Childhood Research Quarterly; International Journal of Bilingual Education and Bilingualism; Journal of Literacy Research; Journal of Early Childhood Literacy; Language and Education; Reading Research Quarterly; Research in the Teaching of English; and TESOL quarterly.*

Second, a comprehensive search of both published and unpublished studies (i.e., dissertations and technical reports) was performed on *PsycInfo, ERIC, PsycArticles, and Proquest* via EBSCO by the author on the October of 2012, followed by an additional search on the February of 2013. Search terms and their various combinations used in the computerized searches included: word, vocabulary, or lexical; knowledge, development, instruction, learning, teaching, or intervention; ESL, EFL, second language, or dual
language; implicit and explicit; semantic awareness or competence; limited English proficiency or LEP; English language learners or ELLs; bilingual; language minority; culturally and diverse learners; literacy development; and experimental study. Finally, backward and forward searches were also performed on all included studies, by searching a list of references found in vocabulary-related studies and published syntheses to ensure a comprehensive search.

As shown in Figure 1, these searches yielded a total of 7,942 studies from which 7,664 were identified as non-intervention studies and were thus excluded. That is, these studies consisted of literature reviews of related topics, qualitative studies, or quantitative studies that did not compare gain scores between independent groups. From the remaining studies, studies without vocabulary measures \(k = 123\) were identified and removed; although many of these interventions highlighted vocabulary as an instructional component, they did not necessarily assess vocabulary growth (e.g., certain interventions examined students’ reading comprehension as an outcome). Subsequently, studies that did not target English word learning were removed \(k = 16\); however, studies in which English and a second language were targeted were included.

A more in-depth examination of the remaining studies was done to select studies that adhered to the inclusion/exclusion criteria relating to sample and statistical characteristics. Therefore, studies with participants who were not in pre-K through 6\(^{th}\) grade \(k=8\) or identified students with language delays \(k=5\) were excluded. In addition, studies were excluded \(k = 5\) when participants in treatment and comparison group(s)
Figure 1. Flowchart summarizing study searches

were not comparable, either by language status (e.g., compared with native English
speakers) or by level of proficiency (e.g., compared with ELLs with higher or lower
proficiency). As a result, a total of 18 studies meeting all inclusion criteria will be
discussed in the systematic analysis.

Details of Study Coding Categories

Various study characteristics were coded in order to answer the primary research
question relating to how included studies varied by their methodological characteristics
before evaluating levels of evidence for effectiveness and determining their relative
impact word-meaning acquisition. The coded study characteristics are classified into the
following: design characteristics such as use of random assignment and baseline
equivalence; sample characteristics such as age and language background; treatment
characteristics such as duration and frequency of the intervention; and outcome characteristics such as the constructs being used to describe aspects of vocabulary knowledge.

**Design characteristics.** A number of methodological indicators that may differentially affect study outcomes were coded. These design characteristics include: use of random assignment, baseline equivalence, attrition rates and comparison group type. Small- and large-sample studies utilizing experimental and quasi-experimental designs were included, and their use of random assignment was coded as 1) using random assignment or 2) did not use random assignment. Secondly, experimental studies were coded by their overall and differential attrition rates. Overall attrition as well as differential rates of attrition for the treatment and comparison groups were examined using the What Works Clearinghouse’s *Study Review Guide Instructions for Reviewing Randomized Controlled Trials and Quasi-experimental Designs* for evaluating attrition bias (What Works Clearinghouse, 2014). A study’s overall attrition rate is computed as

\[ 1 - \frac{n_{post.t} + n_{post.c}}{n_{pre.t} + n_{pre.c}}, \]

where \( n_{post.t} + n_{post.c} \) is the sum of the analytic sample, and \((n_{post.t} + n_{post.c})\) is the sum of the baseline sample. A study’s differential attrition rate is computed as

\[ \left(1 - \frac{n_{post.t}}{n_{pre.t}}\right) \cdot \left(1 - \frac{n_{post.c}}{n_{pre.c}}\right), \]

where \( \frac{n_{post.t}}{n_{pre.t}} \) is the quotient of the treatment group’s analytic sample and baseline sample, and \( \frac{n_{post.c}}{n_{pre.c}} \) is the quotient of the comparison group’s analytic sample and baseline sample.

Subsequently, the studies will be identified as having low or high attrition based on the What Works Clearinghouse’s (2011) attrition categories in Figure 2, which classifies
levels of attrition according to overall and differential attrition rates. In this study, the conservative boundary was used to ensure the highest level of rigor was applied. The figure presents the maximum difference in the attrition rate for the treatment and comparison group that is acceptable for a given level of overall sample attrition. Maximum differential attrition rates are found below the columns labeled “Conservative boundary” and “Liberal boundary.” Upon locating the overall attrition rate on the figure, if the differential attrition rate in the same row exceeds the number listed, the study was classified as a “high attrition” study as opposed to those that fell below these numbers, which were classified as “low attrition” studies.

Thirdly, quasi-experimental studies as well as experimental studies showing a high attrition level were coded by whether study authors attempted to establish baseline equivalence (i.e., Yes or No) and by the methods used to minimize initial differences to create more similar treatment and comparison groups: 1) matching groups on non-meaning-based outcomes (e.g., through decoding-related assessments); 2) matching groups on school-based classifications that indicate English language proficiency (i.e., ESOL levels, reading levels); 3) matching groups on meaning-based outcomes (e.g., through English language proficiency assessments); 3) using non-meaning-based covariates to adjust for differences at pretest; 4) or using meaning-based covariates to adjust for differences at pretest.
Lastly, studies were coded by different types of comparison groups as (1) diluted versions of the treatment whereby most instruction was implemented in the same manner with modifications to the strategies being used or to the subcomponents of a multi-componential intervention; (2) alternative treatments whereby two distinct instructional approaches were contrasted; (3) “treatment as usual” conditions whereby students’ instruction remained unaltered, following regular instructional practices; and 4) “no instruction” conditions in which no information is provided on the manner in which instructional time was used. Although studies may have simply compared two
instructional conditions, studies comparing the effectiveness of three or more conditions will be coded as paired units. Thus, it was acceptable to have different comparison group types per study.

**Outcome characteristics.** To facilitate the interpretation of the findings, outcomes were coded at six levels: 1) assessment type; 2) type of ability; 3) modality; 4) outcome objective 4) word types; 5) type of knowledge; and 6) reliability of the outcome. To enable descriptive analyses, at the first level, types of assessments were coded as: a) standardized assessments (i.e., norm-referenced measures) or b) curriculum-based assessments (i.e., proximal measures of the intervention). Second, the language abilities being measured were also identified: a) receptive outcomes, b) expressive outcomes, or c) a combination of receptive and expressive vocabulary outcomes yielding a composite score for both constructs. At the third level, modalities used during the assessment period were coded as a) oral language outcomes and b) written language outcomes or c) a combination of oral and written language outcomes. Oral language outcomes were those that focused on speaking and listening, while written language outcomes were those that included reading and written tasks. Fourth, assessments were coded by their overall objective (i.e., the level of processing students engaged in to demonstrate their word knowledge): a) to increase the number of words known and only learning single meanings (i.e., breadth) or b) to assess the number of meanings known for a word (i.e., depth). Fifth, word types were assessed according to the outcome under analysis: (a) global vocabulary measure, in which the words being assessed were not necessarily targeted or (2) specific words targeted during the instructional period. Sixth, the type of knowledge being exhibited during the assessment period was also identified: (a)
definitional knowledge measures, such as definition production measures, measuring the
knowledge of definitions and synonyms for targeted words; (b) contextual knowledge
measures, such as sentence cloze tests, which measure knowledge of a word’s use in
context; or (c) a measure that yields a combined score for definitional knowledge and
contextual knowledge. Lastly, information relating to the reliability for each measure
was recorded, which could have been interpreted using any of the following: a) internal
consistency; b) test-retest reliability; and c) inter-rater reliability.

Sample characteristics. To account for sample characteristics that may affect
students’ overall vocabulary growth (e.g., language exposure), samples were coded for
their level of proficiency in their two languages (L1/L2) as well as their socioeconomic
status (SES). These characteristics based on the authors’ descriptions were coded as (1)
low, (2) medium, (3) high or (4) non-reported information. In addition, students’ age (i.e.,
group mean of age) and/or grade level were coded. Due to the distinct learning demands
and increasing conceptual difficulty of vocabulary encountered as they progress through
their schooling, students were grouped into two age groups: (1) early childhood (i.e., pre-
kindergarten and kindergarten) and (2) elementary (i.e., grades 1 through 6).

Treatment characteristics. The following coding categories represent the
multiple factors that may influence the implementation of instruction. First, in an effort to
quantify the instructional time found in the various interventions, three treatment
characteristics were coded. These include duration, frequency, and intensity of
intervention. The duration of the intervention was described as the total number of weeks;
the frequency detailed the number of days per week that instruction was implemented; and the intensity provides the number of minutes per day. When sufficient information is provided, the total instructional time was summed (i.e., represented in minutes).

Second, intervention groups were classified by the focus of their instruction. When vocabulary instruction was central to the objectives of the intervention, these studies were coded for as (1) vocabulary-focused (V+) instruction seeking to increase the effects of meaning-based outcomes (e.g., morphological knowledge, vocabulary knowledge) whereas other interventions sought to increase the effects of other non-meaning based outcomes (i.e., phonological awareness, word fluency) and were coded for as (2) vocabulary-related instruction (V). In addition, the degree to which instructors highlighted particular words by directing learners’ attention to words and their meanings differed across studies. In so doing, instructors are able to reveal specific aspects of lexical information that students may commit to memory through their word-focused, explicit teaching. As such, instructors could have used (1) intentional vocabulary instruction through explicit instructional strategies; (2) intentional instruction through implicit instructional strategies; (3) intentional instruction with both explicit and implicit instruction; or 4) incidental word learning through exposure, wherein no explicit or implicit teaching takes place. The goal of explicit instruction is to direct a learner’s attention to words and their meanings whereas implicit instruction may attract a learner’s attention to meanings while minimizing interruptions to the communication of meaning (Doughty & Williams, 1998).
Third, since word-meaning acquisition may depend on students’ level of participation, the group size to which the treatment was being delivered was coded. As expected, the greater number of students in a group could constrain the allocation of resources. Groups were therefore coded as the following: whole (i.e., 11 or more); small (i.e., 2 to 10); mixed groups (i.e., consisting of more than one group type) and individual student.

Fourth, instruction targeting specific words varied by the level of vocabulary exposure. Single exposures (SE) were coded for when the words were presented to students in one occurrence, with fewer opportunities to refine or reinforce semantic representations of words. Multiple exposures (ME) were coded when word meanings were elaborated by the use of various definitional strategies and/or providing new contextual information (e.g., using realia to represent a word’s visual representation as well as making a connection using students’ prior knowledge). Multiple reinforcement (MR) was coded for instructional conditions that provided the same contextual and/or definitional information through repeated exposure.

Fifth, through the use of L1 bridging strategies, instructors are able to support bilingual learners in accessing existing knowledge that may not be evident through their English lexicon. Thus, interventions were coded as having (1) no L1 support, (2) minimal L1 support (i.e., providing only definitional information with little contextual support through the use of the native language), and (3) significant L1 support (i.e., providing much contextual support with the use of the native language). Examples of providing minimal L1 support to teach about words in decontextualized ways include using concurrent translations to define target words during a shared book reading or
presenting form-meaning pairs with support from the students’ native language.

Significant L1 support, however, may be represented by the use of instructional time that affords opportunities to encounter word meanings in students’ native language through previewing or reviewing content in connected text or in conversation. In addition, high levels of L1 support may also occur throughout naturally occurring bilingual interactions in which words are discussed in context-dependent ways through codeswitching (i.e., alternating between two languages intra/inter sentimentally).

Sixth, to assess the degree of contextual versus definitional support provided throughout most of the intervention, studies were coded on a five-point scale derived from Stahl and Fairbanks’ (1986) study. Treatment and comparison groups were classified in the following manner: (1) supplying only definitional information (e.g., dictionary definition, synonym) without using the word in context; (2) making an emphasis on definitional information whereby some exposure is given to the word in context, but the emphasis is on the child learning the definition; (3) a balanced approach whereby nearly the same amount of definitional and contextual information is given; (4) and making an emphasis on contextual information whereby most word learning is in context; and (5) context-only whereby the child is exposed only to each word in context with no attempt to have the child derive a definition.

Lastly, an aspect that can impact the effects of an intervention is the level of fidelity with which the intervention was implemented. Trained individuals may adhere to intervention protocol more faithfully than those who received no intervention training, thereby impacting the effectiveness of the intervention. Therefore, studies were coded by whether training was provided along with the type of individual delivering instruction.
(i.e., researcher, teacher or other such as child care provider). When author(s) of the study carried out the intervention, it was assumed that they were trained individuals and were coded as such, whereas any other interventionist type would only be identified as a trained individual when it was indicated in the study.

Levels of evidence for effectiveness. To address the first subquestion, the included studies were evaluated based on reported information relating to design and outcome characteristics. According to the What Works Clearinghouse’s (2011) Procedures and Standards Handbook, a study’s level of evidence for its effectiveness is based on three study design factors: 1) the use of randomization, 2) the attrition rates for randomized control trials (i.e., experimental studies), and 3) baseline equivalence for randomized control trials that showed high levels of attrition and for quasi-experimental studies. However, additional requirements in this study were incorporated to add a greater level of rigor when evaluating a study’s level of effectiveness based on additional reporting standards found in this report. For this review, the level of evidence of effectiveness were evaluated for each study based on those factors relating to study design (i.e., two factors: randomization and attrition rate) but also on factors relating to study outcomes (i.e., two factors: reliability of measures and outcome alignment). By incorporating outcome characteristics into the What Works Clearinghouse’s (2011) framework for evaluating quality standards, studies will not only be examined on the degree to which certain design-related decisions may provide biased impact estimates but also on the degree to which certain outcome characteristics may also lead to biased impact estimates due to using measures that did not meet adequate levels of reliability or ones that were over-aligned to a treatment condition.
As shown in Figure 3, studies’ level of evidence were examined at four levels, which will indicate whether a study provides strong evidence (Meets evidence standards), weaker evidence (Meets evidence standards with reservations), or insufficient evidence (Does not meet evidence standards). At the first level, the use of random assignment was examined when evaluating how researchers formed groups of study participants. Experimental studies can receive the highest rating of Meets evidence standards, whereas quasi-experimental studies can receive the highest rating of Meets evidence standards with reservations.

At the second level, a decision was made on whether attrition bias among experimental studies with random assignment may contribute to the estimated effect. If so, these studies received a lower rating and must have provided information relating to baseline differences in a similar manner to quasi-experimental studies. As aforementioned, the overall attrition as well as differential rates of attrition for the treatment and comparison groups were also examined. High overall attrition generally makes the results of a study suspect, and severe differential attrition, which refers to the situation in which the percentage of the original study sample retained in the follow-up data collection, is substantially different for the treatment and the comparison groups, may compromise the comparability of the study groups.

Due to differences between groups found in quasi-experimental studies as well as for experimental studies resulting in high attrition rates, these studies were required to demonstrate that the treatment and comparison groups were equivalent on observable characteristics. According to the What Works Clearinghouse (2014), even with
equivalence on observable characteristics, there may be differences in unobservable characteristics; therefore, the highest rating a well-implemented quasi-experimental study may receive is *Meets evidence standards with reservations*. As stated earlier, attempts to establish equivalence should have been demonstrated using meaning-based outcomes (e.g., oral language proficiency) or school-based classifications (e.g., reading levels, ESOL levels). Studies that have conditions that were not matched on any of these characteristics were considered a study that *Does not meet evidence standards*.

At the third level, a determination was made on whether outcomes were aligned to both conditions or over-aligned to a particular treatment condition. Outcome measures were designated as over-aligned when the measure relied on materials used in the
treatment condition that provide support that was incomparable to a comparison condition. Face validity—the extent to which the measure is clearly defined and measures the construct it was designed to measure—was initially required in order to make this determination. Experimental studies that demonstrated a low attrition level and equivalent groups but had a vocabulary measure that was over-aligned to one of the conditions received the highest rating of Meets evidence standards with reservations. Quasi-experimental studies that showed baseline equivalence but also had an outcome that was over-aligned to one of the conditions were identified as Does not meet evidence standards.

At the fourth level, the reliabilities of outcomes were evaluated. Reliability of an outcome measure may be established by meeting the following minimum standards: a) internal consistency (e.g., Cronbach’s alpha) of 0.50 or higher, b) test-retest reliability of 0.40 or higher, or c) inter-rater reliability (e.g., percentage agreement, correlation, kappa) of 0.50 or higher. An experimental study that has shown acceptable evidence relating to the previous factors in levels 1 through 3 but did not meet standards for reliability received the highest rating of Meets evidence standards with reservations. Similarly, experimental studies that had high attrition rates as well as quasi-experimental studies that met levels 2 through 4 received the highest rating of Meets evidence standards with reservations.

**Coding Procedures**

The author served as the primary coder and, thus, the author coded all of the included studies. Inter-rater reliability of study coding for variables was assessed by comparing the coding by the author with the independent coding of studies by a trained
doctoral student who majored in teaching and learning. A subset (i.e., 20%) of the included studies were randomly drawn and assigned to the second coder. For categorical variables, Cohen’s Kappa (κ)\(^3\) was used to determine the degree of coding agreement between the two independent coders, resulting in substantial inter-rater reliability with a mean Cohen’s Kappa of 0.73, while correlation coefficients were used to check inter-rater reliability for continuous variables, resulting in a mean agreement of 0.91. Coded variables on which both raters differed were discussed and revised; any differences were resolved by reaching consensus.

**Effect Size**

In order to address the second main research question, effect sizes were used to synthesize statistical findings across individual studies that met evidence standards (i.e., Studies coded as *Meets evidence standards* and *Meets evidence standards with reservations*). When effect sizes were not provided in these studies, they were calculated by standardizing the mean difference between the mean values of treatment and comparison group conditions on vocabulary outcomes at pretest and posttest. Standardized mean effect sizes using Cohen’s \(d\) (1998) were calculated separately for effect sizes at pretest and posttest as

\[
d = \frac{\bar{X}_t - \bar{X}_c}{S_{pooled}}
\]

where \(d\) is the difference between the means of the treatment and comparison groups, \(\bar{X}_t - \bar{X}_c\), divided by the pooled standard deviation. The pooled standard deviation is defined as

\[\kappa = \frac{P_O - P_E}{1 - P_E} \text{, where } P_o \text{ is the observed proportion agreement and } P_e \text{ is the expected proportion agreement.}\]
where the pooled standard deviation is the square root of the pooled variance, which include the treatment and comparison group samples, \( n_t \) and \( n_c \), and the conditions’ standard deviations, \( s_t \) and \( s_c \). When studies used covariates through analyses of covariance (i.e., ANCOVAs) to adjust for baseline differences and did not provide effect sizes, effect sizes at posttest were calculated using adjusted means but were divided by the unadjusted pooled standard deviation, \( S_{pooled} \), at pretest. To evaluate the effectiveness of studies, Cohen’s (1988) standards for effect size (\( d \)) were used to evaluate the effectiveness of treatment conditions in relation to comparison groups. Only posttest effect sizes were interpreted that showed trivial pretest effect sizes (i.e., ranging from 0 to 0.20) or were a result of a statistical adjustment due to baseline differences. Consequently, certain treatment-comparison group contrasts were excluded from the analysis when addressing the second research question. In sum, effect sizes were excluded from the final analysis for any of the following reasons: 1) an individual study did not meet evidence standards; 2) treatment-comparison group contrasts showed significant baseline differences with no statistical adjustments to minimize these differences; or 3) effect sizes for treatment-comparison group contrasts or descriptive statistics used to calculate effect sizes (e.g., means, standard deviations, sample sizes) were not provided (See Appendix H).

**Results**

Eighteen studies consisting of 17 journal publications and 1 dissertation study met the inclusion criteria. The studies in this review were published from 1996 to 2011 (\( M=2006, SD= 4.52 \)). Notably, most vocabulary-related research using experimental and
quasi-experimental designs with ELLs was published over the past decade, particularly showing an even greater interest in this area over the span of ten years from 2001 to 2011 ($k=14$).

The following sections address the research questions. First, a summary of the methodological characteristics across studies will be presented, which will detail the evidence available relating to: 1) design characteristics to understand the efforts taken to minimize bias in this area of study; 2) learner characteristics, which will provide the distribution of age, grade levels, L1/L2 proficiencies and socioeconomic backgrounds of ELLs across studies; 3) treatment characteristics to highlight the elements making up varying instructional methods; and 4) outcome characteristics to showcase how word-meaning acquisition was assessed across studies. Additionally, an evaluation in the reporting standards for each study will be discussed.

The second main research question will be addressed by evaluating the effectiveness of the included studies. Studies that met evidence standards, including those that met them with reservations, will be analyzed. In these studies, available effect sizes resulting from treatment-comparison group contrasts will be discussed according to the outcomes measured across all studies.

**Methodological Characteristics Found Across the Studies**

To explore the available evidence on word-meaning instruction among ELLs, this study, which has been informed by previous syntheses, will explore multiple variables relating to the characteristics in design, sample, treatment, and outcome. Subsequently,
studies will be further examined to investigate their level of effectiveness based on reporting standards that are guided by What Works Clearinghouse’s (2011) procedures and other criteria as discussed in the previous section.

**Design Characteristics**

To understand the extent to which studies were designed to avoid biased effect estimates, the use of randomization and comparison group types were examined. Rates of attrition in studies with random assignment (i.e., experimental studies) and baseline equivalence of quasi-experimental studies as well as experimental studies with high attrition rates were also be assessed. Additionally, comparison group types were also analyzed to gain a better understanding of the underlying differences between conditions.

**Random assignment.** A total of ten studies were identified as experimental study designs, which used random assignment in distinct ways to form two or more groups that differed in their instructional approach. For instance, before randomly assigning either a group or an individual to treatment conditions, certain studies included blocking the sample into groups by language status (e.g., Hmong, Spanish and English; Roberts & Neal, 2003); by a combination of characteristics (i.e., gender, grade and school; Schunk, 1999); or by matching L2 receptive scores (Collins, 2004). One study used stratified random sampling by first grouping students by gender and then rank ordering students on their L2 receptive scores (Uchikoshi, 2006). One multi-grade study (i.e., pre-Kindergarten through second grade) randomly assigned participants to one of two teachers who provided a treatment within a student’s respective grade level (Silverman & Hines, 2009), whereas two studies randomly assigned participants to two treatment conditions within each participating classroom to avoid nesting effects (Farver et al.,
The remaining eight studies used quasi-experimental designs in which participants were not randomly assigned to the treatment and comparison conditions but instead formed part of pre-existing intact groups (i.e., classrooms, schools).

**Attrition among experimental design studies.** The ten experimental studies included in this review were either classified as having low ($k=6$) or high attrition ($k=3$). Five studies classified as having low attrition (Collins, 2010; Schunk, 1999; Nelson et al., 2011; Silverman & Hines, 2009; Zhang & Schumm, 2000) were not impacted by the withdrawal of any participants (i.e., both differential and overall attrition rates were 0%). Although Farver et al. (2009) only disclosed that participants left the study without indicating their assignment to either the treatment or comparison condition (i.e., an overall attrition rate of 2.1% could only be calculated), the differential attrition rate of the study was sufficiently low to result in classifying this study as having low attrition. Studies having overall and differential attrition rates that surpass conservative boundaries in Figure 2 were classified as high-attrition studies (Carlo et al., 2004; Roberts & Neal, 2003). Although attrition rates were not reported in the Carlo et al. (2004) study, a report by the What Works Clearinghouse (2006) has characterized this particular study as having high attrition due to the withdrawal of one classroom. Due to the unique nature of individual growth modeling employed in one study (Uchikoshi, 2006), which can utilize repeated waves of data and analyze data sets regardless of any missing data, differential and overall attrition rates were deemed low, and the study was therefore categorized as a low-attrition study for this category.

**Baseline equivalence.** As a result of potentially lacking equivalence between treatment and comparison conditions, quasi-experimental studies should assess whether
observable characteristics are equivalent prior to the instructional period of the study for the analytic sample. If baseline equivalence is not established, different statistical techniques could be used, such as an ordinary least squares regression adjustment for the baseline covariates and analysis of covariance (ANCOVA), to adjust for remaining differences.

Among the nine quasi-experimental studies, five studies stated that attempts were made to establish equivalence prior to the instructional period. In these studies, participants in treatment and comparison conditions did not significantly differ on English achievement scores (Atay & Kurt, 2006) and on L2 receptive vocabulary assessment scores (Filippini, 2007; Giambo & McKinney, 2004, Lesaux et al., 2010; Mancilla-Martinez, 2010; Spycher, 2009) before instruction was delivered. A study that did not initially attempt to establish baseline equivalence employed statistical adjustments by regressing pretest L2 receptive scores on posttest scores (Ulanoff & Pucci, 1999). In contrast, a few studies that did not establish baseline equivalence did not disclose any information relating to differences between conditions (Avila & Sadoski, 1996) or made statistical adjustments using assessment scores that measured students’ nonverbal ability (i.e., a non-vocabulary related measure) as covariates (Dockrell, 2010).

Experimental studies that did not meet attrition standards (i.e., discussed in the previous section) demonstrated attempts to establish baseline equivalence or made statistical adjustments for the post-attrition analytic sample to ensure comparable groups were formed. Pretest scores for vocabulary outcomes were used as covariates to make statistical adjustments in Nelson et al. (2003). Also, upon reviewing disaggregated data

---

4 No information about the assessment used in this study to measure English achievement was reported.
from the Carlo et al. (2004) study, the What Works Clearinghouse (2014) was able to statistically adjust for baseline differences in the same manner. Roberts and Neal (2003) created equivalent groups using scores on an English language proficiency test (Pre-Idea Proficiency Test-Oral, 1989)

Comparison group type. Study authors most regularly used diluted comparison group types to examine a treatment condition’s effectiveness on word learning. Fourteen comparison groups differed by the slight variation(s) among instructional activities found in their respective conditions. In diluted comparisons, a central activity (e.g., shared reading) was left intact in two conditions, but the conditions may have differed by: the extent to which the L1 was used to supply vocabulary-related information (Avila & Sadoski, 1996; Farver et al.; Ulanoff & Pucci, 1999; Zhang & Schumm, 2000); the degree of interaction in a set of certain activities within multi-componential conditions (e.g., interactive tasks versus independent tasks; Atay & Kurt, 2006; Nelson et al., 2011); whether thematically-related content was introduced through multimedia (Silverman & Hines); whether explicit attention was drawn to word meanings (Spycher, 2009); the use of visual representations to support the recall words (Avila & Sadoski, 1996; Schunk, 1999; Zhang & Schumm, 2000); and whether instructional components highlighted semantic relations rather than morphological instruction improved word learning when packaged with decoding-related instruction (Filippini, 2007).

When comparison groups differed significantly in their approach, they were identified as alternate comparison groups. Thus, four comparison groups used alternate forms of treatment conditions. These comparison groups included comparing: exposures to shared reading versus a phonological awareness condition (Giambo & McKinney,
activities associated with teacher-directed shared book readings versus student-centered, non-print-based communication (Dockrell et al., 2010); comprehension-oriented condition versus a decoding-related condition (Roberts & Neal, 2003); and exposure to two distinct television programs (e.g., *Arthur* versus *Between the Lions*) that differed in content and instructional strategy (Uchikoshi, 2006).

“Treatment as usual” comparison groups were found in four studies. These studies compared packaged instructional approaches to the standard practices adopted by the schools in which instruction was taking place. In three of the “treatment as usual” comparison groups, researchers sought to enhance breadth and depth of academic word knowledge across content areas (Carlo et al., 2004; Lesaux et al., 2010; Mancilla-Martinez, 2010). In a separate study, two treatment conditions (i.e., a teacher-directed story reading condition and a student-centered, oral language condition) were compared to one in which a national preschool curriculum was being followed to increase learners’ general vocabulary.

Lastly, a few studies made comparisons between treatment conditions and conditions in which no instruction was provided. Consequently, these comparisons were coded as “no instruction” comparison groups. “No instruction” conditions were compared to two treatment conditions per study across four studies (Collins, 2004; Dockrell et al., 2010; Filippini, 20007; Uchikoshi, 2006), resulting in eight “no instruction” comparisons. The studies with “no instruction” conditions did not provide a clear indication of how the instructional time was used, in contrast with “treatment as usual” conditions, which provided a general overview.
Outcome Characteristics

Based on the authors’ descriptions, outcome measures used to assess word-meaning acquisition were coded in multiple ways. Descriptions of the outcome measures are provided according to the types of words assessed, either testing knowledge of general vocabulary or knowledge of specific vocabulary (i.e., targeted vocabulary). Their descriptions include the assessment types used (i.e., curriculum-based or standardized); ability types measured (i.e., receptive and/or receptive knowledge); the assessments’ objectives (i.e., to test breadth or depth of knowledge); type of word knowledge (i.e., definitional and/or contextual knowledge); and the modality used throughout the task (i.e., oral and/or written). A summary description of these various characteristics is shown in Appendix A. In addition, reliabilities of study outcomes are also described as well as how measures aligned with treatment-comparison conditions across studies.

Measures of general vocabulary knowledge. Eight of the eighteen studies used standardized, general vocabulary assessments to measure the breadth of learned words, with most of this subset exclusively examining students’ definitional knowledge (Appendix B). That is, a study distinctively used a multi-componential assessment that measured receptive knowledge of both contextual and definitional information relating to general vocabulary and provided a composite score for these components (Lesaux et al., 2010), while a greater number of studies, which strictly measured definitional knowledge of general vocabulary \((k=5\); Carlo et al., 2004, Dockrell, 2010, Farver et al., 2009; Giambo & McKinney, 2004; Uchikoshi, 2006), tested ELLs’ listening vocabulary (i.e., receptive knowledge). In these assessments, through oral instructions, participants were requested to select one of four pictures that best describe a word’s meaning. Conversely,
to assess ELLs’ expressive knowledge, certain studies \( k = 3 \) elicited an oral response indicating the name for a pictured item. Another means of measuring definitional knowledge was found in two studies that employed the use of a comprehensive receptive and expressive vocabulary measure for which ELLs were instructed to read a word and provide a related meaning (e.g., synonym, antonym) through an oral response (Nelson et al., 2011). In all of these assessments, ELLs were expected to have developed a form-meaning connection, which involves recalling a meaning when seeing or hearing a particular word, or having to recall a spoken or written form when wanting to express a meaning (Nation, 2012). In the included studies, no standardized measures were used to assess depth of knowledge.

**Measures of curriculum-based vocabulary knowledge.** As shown in Appendix C, curriculum-based measures that targeted specific words were used in fifteen studies to assess breadth of knowledge (i.e., measuring definitional knowledge: \( k = 12 \); measuring contextual knowledge: \( k = 3 \)) and in six studies to assess depth of knowledge (i.e., measuring definitional knowledge: \( k = 3 \); measuring a combination of definitional and contextual knowledge). Breadth of knowledge measures testing definitional knowledge consisted of picture-naming tasks (Collins, 2004; Roberts & Neal, 2003), similar to those found in standardized assessments measuring receptive (e.g., PPVT; Dunn & Dunn, 1997) and expressive vocabulary (e.g., British Ability Scales II; Elliot, 1996); cued recall tasks in which students were presented with a previously encountered word and were required to produce an associated meaning in the form of a translation equivalent (Avila & Sadoski, 1996) or synonym (Zhang & Schumm, 2000); a “true or false” task in which students agreed or disagreed with a given definitional statement of a target word.
(Filippini, 2012); and multiple-choice tasks in which ELLs were asked to select a target meaning of a given word from a range of possible distractor statements (Carlo et al., 2004; Lesaux et al., 2010; Nelson et al., 2011; Ulanoff & Pucci, 1999). Studies with breadth of knowledge measures testing contextual knowledge were fewer in number (Avila & Sadoski, 1996; Mancilla-Martinez, 2010; Zhang & Schumm, 2000) and solely evaluated ELLs’ receptive knowledge. These receptive vocabulary assessments included sentence completion tasks, which required students to comprehend words in context, by identifying the most appropriate word from a range of possible answers that would appropriately fit the context of a sentence for which a word has been omitted (Avila & Sadoski; 1996; Zhang & Schumm, 2000) or to replace a highlighted word with a synonym that may also apply.

Curriculum-based measures used to assess depth of knowledge were either used to evaluate the amount of semantic information known about a word (i.e., through measures of definitional knowledge) (Carlo et al., 2004; Lugo-Neris, 2007; Silverman & Hines, 2009) or knowledge of both definitional and contextual information about a word (i.e., through a comprehensive measure) (Atay & Kurt, 2006; Carlo et al., 2004; Mancilla-Martinez, 2010). These measures of definitional knowledge tested ELLs’ knowledge of paradigmatic relationships through a word association task in which students were asked to select a set of words most closely related in meaning of a target word (Carlo et al., 2004) or by correctly identifying as many definitional statements about a word providing a “Yes” or “No” response (Silverman & Hines, 2009). Combinations of both definitional and contextual information were tested in three studies: in Carlo et al.’s (2004) study through a polysemy production task in which
students generated as many example sentences possible conveying the different meanings of polysemous words; in Atay & Kurt’s (2006) study through a self-evaluative task in which students were provided with a target word and rated their degree of familiarity with the word (i.e., whether they have heard of it; whether they know something about it; or whether they know it well and can use it); and in Mancilla-Martinez’s (2010) study, which required participants to generate sentences, providing a definitional statement as well as its use in context, only after specifying they had a certain degree of familiarity when presented with the word.

Reliability of outcome measures. Different types of reliability were reported by studies to exhibit the degree to which outcome measures produced stable and consistent results. Test-retest reliability, solely used for a curriculum-based outcome measure in one study, showed scores were highly stable over time ($r = 0.92$; Roberts & Neal, 2003). Inter-rater reliability used for curriculum-based measures was reportedly high in two studies (Silverman & Hines, Ulanoff & Pucci, 1999). Most studies, which reported internal consistency reliability for outcome measures, ranged from 0.64 to 0.98 and met the minimum standards of reliability (What Works Clearinghouse, 2011); standardized measures generally showed higher levels of internal consistency than curriculum-based measures. Two studies that did not report reliabilities for outcome measures provided substantial information relating to outcome measures’ reliabilities. Mancilla-Martinez (2010) stated that the two curriculum-based vocabulary measures were significantly correlated with a standardized measure’s vocabulary subtest (i.e., Group Reading Assessment and Diagnostic Evaluation; Williams, 2002). Similarly, Spycher (2009) mentioned that several statistical analyses were conducted prior to the study with students
at the same school using methods for establishing validity and reliability of language assessments (e.g., internal consistency, concurrent validity with PPVT-III). Two studies did not report information relating to reliabilities of outcome measures (Avila & Sadoski, 1996; Schunk, 1999).

**Outcome alignment.** Outcome measures were also coded to determine their alignment to the instructional goals of the treatment and comparison conditions for each study. That is, measures were considered over-aligned when they were particularly tailored to a treatment condition, not providing an accurate indication of the effect of the condition. An outcome measure based on an assessment that relied on materials (i.e., using or exposing ELLs to target words) used in one treatment condition but not in the comparison condition was classified as being over-aligned. Although certain studies with curriculum-based measures that have “treatment as usual” comparison groups (Carlo et al., 2004; Lesaux et al., 2010; Mancilla-Martinez, 2010) intentionally exposed learners to target words during one condition, there was an expectation that learners in these comparison groups would be exposed to the same academic vocabulary as they received regular instruction in the same content areas as those in the treatment conditions. However, Roberts and Neal (2003), who used an alternate treatment condition, did not expose learners to target vocabulary instruction in the comparison group since the authors were seeking to examine the effects of two distinct skill sets—comprehension-related and decoding-related outcomes—that aligned with the treatment and comparison group, respectively. In addition, when measures are standardized, they are considered aligned to both conditions regardless of instruction since the same test is given in the same manner to all test takers and do not contain words specifically targeted by any one condition.
Sample Characteristics

A total of 1,708 ELLs participated across the 18 studies, with a range of 19 to 274 participants per study. The study authors described this population using various terms: limited English proficient (LEP) (Avila & Sadoski, 1996; Schunk, ; Zhang & Schumm, 2000); ELLs (Carlo et al., 2004; Collins, 2004; Dockrell, 2010; Farver et al., 2009; Filippini; Spycher, 2009); bilinguals (Carlo et al., 2004; Uchikoshi, 2006); non-speakers of English (Schunk, 1999); language minority students (Lesaux et al. 2004; Mancilla-Martinez, 2010); and English foreign language (EFL) or English as a second language (ESL) learners (Atay & Kurt, 2006; Schunk, 1999). With the use of these various terms, however, the samples in these studies similarly referred to learners whose L1 is not English and were identified as ELLs prior to the study (i.e., based on home language surveys or classroom placement) and/or did not meet certain levels of English proficiency on a standardized assessment during the recruitment of participants. According to authors’ descriptions, the samples in this review received instruction in a variety of program models, which include English foreign language classrooms (EFL) (Atay & Kurt, 2006); transitional bilingual classrooms (Ulanoff & Pucci, 1999); bilingual classrooms (Carlo et al., 2004), and mainstream classrooms (Carlo et al., 2004; Filippini, 2007; Mancilla-Martinez, 2010; Spycher, 2009). Appendix D provides an overview of the pertinent information coded for each study relating to sample characteristics.

Age and grade levels. The distribution of age groups for the included studies is fairly balanced when classifying ELLs by early childhood versus elementary grades, with 10 studies representing ELLs from the early childhood grades, 8 studies in the elementary grades, and two studies with a combination of early childhood and elementary grades (i.e.,
K through 2nd grade and prekindergarten through 2nd grade; Schunk 1999; Silverman & Hines, 2009). As shown in Appendix E, the preponderance of studies with ELLs in the earliest grades, spanning from prekindergarten to kindergarten, is indicative of a greater interest in vocabulary-related research among this age group. More specifically, when grades were considered separately (i.e., prekindergarten through 6th grade)—with the exclusion of multi-grade studies—prekindergarten and kindergarten accounted for approximately half of the published studies. The mean age of the samples in each study were solely reported by 10 studies.

**Language groups.** As seen in Appendix E, most studies included homogeneous language groups that predominantly comprised Spanish-speaking ELLs (Avila & Sadoski; Farver et al., 2009; Filippini, 2007; Giambo & McKinney, Nelson et al., 2011; Spycher, 2009; Uchikoshi, 2006; Ulanoff & Pucci, 1999; Zhang & Schumm, 2000). Portuguese- and Turkish-speaking ELLs also formed part of homogeneous groups found in two studies (Atay & Kurt, 2006; Collins, 2004). Heterogeneous language samples were included in 7 studies; the primary languages for participants in these studies were diverse across studies.

**L2 proficiency levels.** Several authors provided greater detail about the sample in the attempts to characterize ELLs within a study. However, when doing so, only a few differentiated the L2 proficiency levels within ELL study samples but did not conduct subgroup analyses within these studies. Although descriptive information may have been provided across studies, most samples comprised homogeneous low proficiency groups with few studies using heterogeneous ability groups (Appendix E). The distribution of varying L2 proficiency levels among ELLs was provided in two studies (Giambo &
McKinney, 2004; Roberts & Neal, 2003) using scores from the Preschool Language Assessment Scales-English (PreLAS, 2000; Duncan & De Avila, 2003) to ensure that ELLs of similar proficiency were proportionately distributed across treatment conditions. Similarly, two studies (Mancilla-Martinez, 2010; Zhang & Schumm, 2000) relied on school-based classifications (i.e., formerly LEP and LEP students; ESOL levels) to describe the distribution of ELLs across treatment conditions. As a result, these studies were all classified as using heterogeneous ability groups, which may include a combination of ELLs with low, medium and high L2 proficiency.

Other studies did not differentiate among their respective ELL samples based on L2 proficiency and were resultantly coded as homogenous L2 ability groups. Only a few studies specified the level of English proficiency shared by these homogeneous groups, but in those studies, which provided descriptive information relating to L2 proficiency (Dockrell et al., 2010; Lesaux et al., 2010; Roberts & Neal, 2003; Uchikoshi, 2006), their claims of low English proficiency were substantiated prior to the study through the use of scores from standardized assessments—measures of vocabulary (i.e., PPVT-III; Dunn & Dunn, 1997), reading comprehension (i.e., Gates-MacGinitie Reading Comprehension Test; 2000) and, more comprehensively, oral language (i.e., Pre-Literacy Language Assessment Scales; DeAvila & Duncan, 2000)—as well as through the use of home language surveys. Studies with homogeneous ability groups that did not report additional information relating to L2 proficiency were coded as low L2 proficiency since it is assumed that ELLs’ English skills have not met a minimum threshold.

**L1 proficiency levels.** The level of learners’ L1 proficiency was rarely considered in the studies included in this review (Appendix E). Only two studies reported
information regarding the samples’ L1 proficiency and was coded as a “homogeneous L1 ability” sample (Farver et al., 2009; Uchikoshi, 2006) since ELLs within the sample were left undifferentiated. Uchikoshi (2006) examined ELLs’ Spanish vocabulary scores on a parallel, untranslated version of the PPVT-III (i.e., Test de Vocabulario en Imagenes Peabody; Dunn, Padilla, Lugo, & Dunn, 1986), which revealed that the sample scored slightly lower than their monolingual Spanish counterparts on which this standardized test was normed, and consequently this sample was coded as having low L1 proficiency. Farver et al. (2009) used a Spanish vocabulary measure along with an English version of the same measure to determine whether instruction would have an impact on both Spanish and English receptive and expressive vocabulary; ELLs in this study were identified as having low proficiency at pretest.

**Socioeconomic backgrounds.** There were three ways in which information relating to participants’ socioeconomic backgrounds was reported. Study authors reported SES by supplying data: (1) at the city or district level \((k=6)\); (2) at the school level \((k=5)\); or (3) at the individual level \((k=6)\). Two studies did not report any information relating to ELLs’ SES (Nelson et al., 2011; Zhang & Schumm, 2000). As showcased in Appendix E, most studies consisted of samples with low SES \((k=12)\), with few that were coded as low and medium SES \((k=4)\), and yet a smaller proportion that did not provide any information relating to SES \((k=2)\).

Most studies that were coded at the city or district level consisted of low SES samples \((k=4)\) and varied in their descriptions of the schools’ surrounding communities, explaining that they were simply disadvantaged (Avila & Sadoski, 1996); “the most deprived borough in the community” (Dockrell et al., 2010); a high-poverty
neighborhood (Spycher, 2009). In addition, two particular studies were coded as both low and medium SES at the city or district level, resulting from the authors’ description of the schools’ surrounding demographics as middle- to low-income families (Collins, 2004) and working-class families (Carlo et al., 2004).

Five studies addressed samples’ SES at the school level. Two studies were coded as low SES based on a large proportion of the school population receiving free and reduced-price lunch (Filippini, 2007; Mancilla-Martinez, 2010; Uchikoshi, 2006; Ulanoff & Pucci, 1999), whereas Lesaux et al.’s (2010) study was coded as low and medium SES since schools in the sample of this study showed income disparities within this group (i.e., school populations ranged from 24.4% to 100% of students with low-income backgrounds; Lesaux et al., 2010).

Studies in which details were provided at the individual level were all considered to have samples with low SES with the exception of one study. Silverman & Hines (2009) indicated that 55% of the total sample received free and reduced-price lunch. As a result, this study was coded as a low and medium SES. In the remaining studies, study authors verified whether individual participants were recipients of the free and reduced-price lunch program (Giambo & McKinney, 2004; Schunk, 1999) or simply stated that participants were from low-income families without providing further evidence (Atay & Kurt, 2006; Farver et al., 2009; Roberts & Neal, 2003).

**Treatment characteristics**

Several treatment characteristics were coded as a means of understanding how instructional factors may impact word-meaning acquisition among ELLs. These characteristics include: the focus of instruction; words targeted during instruction;
explicitness of instruction; degree of contextual versus definitional support; instructional time; group size; level of exposure; level of L1 support; and the training provided to individuals delivering instruction. Appendices F & G provide a summary of the coding categories for the treatment characteristics found in this section at the study level and condition level, respectively.

**Focus of instruction.** The included studies were classified as being vocabulary-focused or vocabulary-related. Through vocabulary-focused instruction, thirteen studies principally investigated the effects of various approaches on word-meaning acquisition, which were seldom measured alongside other closely related, meaning-based outcomes (i.e., morphological awareness and reading comprehension). To primarily increase word learning, these studies implemented the following: generative word-learning approaches (Atay & Kurt, 2006; Avila & Sadoski, 1996; Schunk, 1999; Zhang & Schumm, 2000), packaged instructional approaches with multiple word-focused components (Carlo et al, 2004; Lesaux et al., 2006; Mancilla-Martinez, 2010), multimedia approaches (Silverman & Hines, 2009; Uchikoshi, 2006), the use of multiple explicit strategies versus incidental exposure (Collins, 2004; Spycher, 2009), and approaches that centered on the use of bilingual methodologies in context-dependent ways (Ulanoff & Pucci, 1999).

On the other hand, several studies were classified as providing vocabulary-related instruction, whereby teaching was particularly geared toward increasing the effects of other literacy skills with word learning as a by-product of the instructional activity. In addition to improving ELLs’ vocabulary knowledge, these studies examined the effectiveness of instruction on measures of morphosyntax (Dockrell et al., 2010), narrative skills (Dockrell et al., 2010); print knowledge (Farver et al., 2009); fluency
Words targeted during instruction. To determine the degree to which instruction may have enhanced ELLs’ word learning, studies were coded by the levels of conceptual difficulty for words targeted during instruction, using Beck & McKeown’s (2002) tier-word classifications. As explained by the study authors, tier 1 words in these studies were instructed due to being easily acquired by virtue of their concreteness and/or use in everyday communication (Atay & Kurt, 2006; Avila & Sadoski, 1996; Filippini, 2007; Nelson et al., 2011; Roberts & Neal, 2003; Schunk, 1999; Ulanoff & Pucci, 1999; Zhang & Schumm, 2000). When words were not intentionally targeted during instruction (e.g., learned through incidental exposure) or targeted words did not include descriptive information that would enable their classification by tier, outcome measures for studies were instead analyzed. Due to the descriptions of these measures, three studies were additionally classified as targeting tier 1 words since these measures relied on picture-
naming tasks with concrete words (Farver et al., 2006; Giambo & McKinney, 2004; Uchikoshi, 2006). In total, fourteen studies targeted tier 1 words and were therefore greater in number than studies which focused on tier 2 and tier 3 words.

Tier 2 words were exclusively targeted in four studies. The main reason for targeting tier 2 words, as indicated by the study authors, was due to their nature of being highly-functional, cross-disciplinary words. In particular, some studies in this subset demonstrated their frequency and utility by exposing ELLs to words in multiple contexts across academic disciplines, such as through historical accounts and fictional storybooks (Carlo et al., 2004; Filippini, 2007; Lesaux et al., 2010) as well as during mathematics and history instruction (Mancilla-Martinez, 2010).

Tier 3 words were targeted in fewer studies and were always targeted in conjunction with Tier 2 words. Consequently, these studies were coded as using a combination of Tier 2 and Tier 3 words. In Collins (2004), Tier 2 and Tier 3 words were taught due to their importance in understanding selected storybook readings. In these stories, Tier 2 words included words like shallow and unruly whereas Tier 3 words included fauna and aperture. To ensure ELLs’ comprehension during science instruction, instruction was strategically designed to teach general academic words to a greater proportion than content-specific words to ensure comprehension of overarching science concepts (e.g., habitat) in Spycher (2009).

**Explicitness of instruction.** As previously mentioned, all study conditions—both those considered treatment conditions and comparison conditions in each study—were coded to identify the level of explicitness undertaken throughout their respective instructional practices. After identifying the conditions in which intentional vocabulary
teaching took place, these studies were coded as using 1) explicit strategies, 2) implicit strategies or 3) both. When no intentional vocabulary instruction was implemented throughout a condition, they were coded as a condition providing 1) incidental exposure, wherein learners may have been exposed to a meaning-based activity and no efforts were made by the instructor to direct or attract learners’ attention to words, or they were coded as one with 2) not reported, wherein no meaning-based instruction (e.g., solely focused on phonological awareness; Filippini, 2007) or any details about instructional procedures (e.g., for treatment as usual conditions) were provided for this condition.

Intentional vocabulary teaching rather than solely exposing learners to contextual information predominated most instructional conditions. Of the 44 conditions across all studies, 18 conditions exclusively used explicit strategies, 2 only used implicit strategies, and 8 used both explicit and implicit strategies. Incidental word learning was promoted in 8 treatment conditions, and no instruction was provided for 8 conditions. As showcased in Appendix B, most studies compared explicit teaching conditions with ones providing opportunities for incidental word learning.

**Degree of contextual versus definitional support.** When examining how conditions in this review varied by their degree of contextual versus definitional support across the 44 conditions, the various typologies representing the contextual-definitional continuum were comparable in number with the exception of one. A larger proportion examined the effectiveness of conditions supporting word learning through contextual emphasis (i.e., these conditions narrowly provided definitional with a major emphasis on learning words in context). These conditions were either compared to another condition with contextual emphasis (Dockrell et al., 2010; Farver, 2009; Schunk, 1999), a condition
with context-only support (Collins, 2004; Roberts & Neal, 2003; Schunk, 1999; Uchikosi, 2006), a condition providing support through definitional emphasis (Atay & Kurt, 2006); or one in which no meaning-based support was provided (i.e., Conditions with no meaning-based support included those that strictly focused on decoding-related skills with no exposure to meaning and those in which no instruction was provided in the study, such as “treatment as usual” conditions).

At the ends of the continuum, definition-only conditions were generally compared to other conditions that were closer in degree of contextual support, whereas comparisons with context-only conditions tended to vary in contextual and definitional support. Accordingly, context-only conditions were compared with another context-only condition; conditions with contextual emphasis (Collins, 2004; Roberts & Neal, 2003; Schunk, 1999; Uchikoshi, 2006); conditions with definitional emphasis (Filippini, 2007); conditions with no instruction (Collins, 2004; Giambo & McKinney; Uchikoshi); and one with a balanced approach (Spycher, 2009). In contrast, definition-only approaches to word learning were implemented in a few studies and were also compared to other definition-only conditions that provided no contextual support (Avila & Sadoski, 1996; Zhang Schumm, 2000); these conditions consisted of instructing ELLs to use the keyword method versus using side-by-side translations of words to recall word meanings. Balanced approaches were frequently compared to conditions with either higher contextual or definitional support was provided and were a rarely compared with other balanced conditions. Silverman & Hines (2009) provided similar degrees of contextual
and definitional support throughout all the components of the two conditions being compared in the efforts to chiefly examine the effectiveness of one learning medium with another (i.e., storybooks versus video-based multimedia).

**Instructional time.** The amount of time devoted to instruction varied for each study based on duration, frequency and intensity. The duration—the number of weeks in which instruction was delivered—ranged from 1 week to 21 weeks ($M= 11.85$, $SD= 7.61$). Within each week, instruction was delivered between 2 to 5 days per week ($M= 3.41$, $SD= 0.97$). The intensity of instruction, defined as the number of minutes per day, ranged from 15 to 60 minutes ($M=24.09$, $SD=13.57$). When studies did not provide the total instructional time, the number of minutes was computed by finding the product of the duration, intensity and frequency. The total instructional time across studies ranged from 67.5 minutes to 3,240 minutes ($M=1197.83$, $SD=865.20$). Instructional time could not be determined for two studies that did not provide sufficient information involving the time-related characteristics (Spycher, 2009; Ulanoff & Pucci, 1999).

**Group size.** Instruction was mainly organized around whole-group ($k=8$) and small-group sessions ($k=8$) across studies. Mixed groups, which included any combination of individual, small-group, or whole-group sessions, were identified in two studies: Carlo et al. (2004) used a combination of whole and small groups to deliver instruction, whereas Lesaux et al. (2010) provided ELLs opportunities to engage in word learning through whole-, small-, and individual instruction (Atay & Kurt, 2006; Carlo et al., 2004; Lesaux et al., 2010). Dockrell et al.’s (2010) study was coded as having misaligned groups since both conditions differed by the size of the group in which instruction was delivered (i.e., whole versus small group).
**Level of exposure.** Instructional conditions were coded by the level of exposure provided to ELLs as they encountered targeted words. Conditions affording opportunities to ELLs in refining word meanings (i.e., “multiple exposure” studies), which implemented a variety of definitional strategies and/or contexts, consisted of 15 conditions (See Appendix B). From this subset, most utilized multiple instructional components through which ELLs were exposed to novel contextual information as definitional strategies were also employed to support word learning. On the other hand, three conditions in this group (Studies: 4a, 10a, 10b) aimed to test the effectiveness of definitional strategies and therefore did not provide additional learning tasks; word learning was mainly grounded in ELLs’ storybook experiences.

Conditions that repeatedly exposed ELLs to the same definitional and/or contextual information (i.e., “multiple reinforcement” studies) with no additional efforts to expand on word meanings comprised 9 conditions. In these studies, researchers examined how minor disparities in definitional strategy may improve word learning. Consequently, the same meanings for words were reinforced through review exercises to explore whether the addition of visual representations to written definitions (Avila & Sadoski, 1996; Zhang & Schumm, 2000) or gestures throughout a recitation activity (Schunk, 1999) would improve recall of word meanings. Additionally, several conditions exposed ELLs to the definitional and/or contextual information on a single encounter (i.e., “single exposure” studies) with a word; these included those that exposed ELLs to one reading of a storybook without supplying novel information outside the immediate context. A large proportion of conditions in the included sample of studies did not provide sufficient information to evaluate the level of exposure.
Level of L1 support. The use of L1 support was limited to a few studies ($k=5$). Significant L1 support was provided in multiple conditions that offered much contextual support through the L1. For instance, Carlo et al.’s (2004) study was designed to integrate ELLs’ L1 into multiple components of instruction by providing texts in Spanish as a means of previewing words prior to encountering them in English, by providing translation equivalents of target words, and throughout a cognate activity in which ELLs were deemed as resources during instruction. In the same vein, instructional activities were structured to enable learners to engage in code switching with others during their word-focused talk in one study (Atay & Kurt, 2006). Similarly, although the provision of L1 support was integral to the definitional strategies being examined (i.e., presenting the keyword method versus translation equivalents) in Avila and Sadoski (1996), instructors were encouraged to also use the L1 in conversations with students during the instructional period to facilitate their learning. Minimal L1 support was demonstrated by a few conditions that strictly adhered to supplying the translation equivalents for targeted words.

Training provided. Most studies included in this review reported that individuals delivering instruction were trained on the procedures of the study. More specifically, teachers were trained to deliver instruction in nine studies. Additionally, study authors explained that teacher training was conducted through initial sessions prior to the study (Avila & Sadoski, 2006; Silverman & Hines, 2009); through ongoing meetings or professional development activities (Carlo et al., 2004; Dockrell, 2010; Giambo & McKinney, 2004; Lesaux et al., 2010; Mancilla-Martinez, 2010; Silverman & Hines, 2009; Spycher, 2009); and through researcher-modeled lessons during
instructional time (Spycher, 2009). Studies with researcher-led instruction \((k=4)\) did not report any information relating to training on the instructional procedures of the study; however, it was assumed they were trained individuals due to their involvement in the study (Collins, 2004; Schunk, 1999; Zhang & Schumm, 2000). Other individuals \((k=4)\) who were also trained to deliver instruction included undergraduate students (Filippini, 2007; Roberts & Neal, 2003), graduate students (Farver et al., 2009; Filippini, 2007), and para-educator tutors (Nelson et al., 2011). Finally, no information was given regarding training in three studies (Atay & Kurt, 2006; Uchikoshi, 2006; Ulanoff & Pucci, 1999). However, instruction in Uchikoshi (2006) was wholly dependent on the use of multimedia support (i.e., exposure to television), requiring no instruction from another individual.

**Statistical Characteristics**

The statistical tests in the 18 studies included \(t\)-tests \((k=2)\), analysis of variance (ANOVA; \(k=3\)), analyses of covariance (ANCOVA; \(k=5\)), multivariate analysis of variance (MANOVA; \(k=4\)), multiple regression analysis \((k=1)\), multilevel modeling \((k=2)\), and individual growth modeling \((k=1)\). Although most design studies consisted of individual-level analyses that used data for each participant, cluster-level analyses were conducted for three studies (Carlo et al., 2004; Lesaux et al., 2010; Nelson et al., 2011) in which data for each cluster was an aggregation of data from individuals at a point in time. Two of the studies using \(t\)-tests compared the means of the conditions’ gain scores on vocabulary assessments (Atay & Kurt, 2006; Mancilla-Martinez, 2010). Conversely, Giambo & McKinney (2004) applied the \(t\)-test to instead test for significant differences in vocabulary scores from pretest to posttest for each condition.
Studies with ANOVA analyses were slightly more frequent. One study included testing the differences between two learning conditions (i.e., keyword vs. translation) with time (i.e., immediate and delayed) as a within-subject factor (Avila & Sadoski, 1996). Two-way ANOVAs were used in Spycher (2009) to determine whether there was any significant difference on vocabulary pretest scores and on the pre-posttest gain scores between two samples and between each language group (i.e., ELLs and non-ELLS) within and between samples.

In contrast to the previous ANOVAs, the more common ANCOVA tests used continuous predictors as well as categorical predictors. Data were analyzed in a series of univariate ANCOVA with instructional condition as the between-subjects factor, time as the within-subject factor and pretest scores of the outcome under of analysis as the covariate (Dockrell et al., 2010; Farver et al., 2009; Filippini, 2007; Schunk, 1999; Silverman & Hines, 2009). Additionally, Dockrell et al. (2010) used non-verbal ability scores in addition to vocabulary pretest scores as covariates. SES and language backgrounds (ELLs vs. non-ELLS) were also used as a covariate in Silverman & Hines (2009) to control for these learner characteristics. Similarly, one study used multiple predictor variables in a multiple linear regression to model the relationship between explanatory variables and the outcome. In this study, Collins (2004) used hierarchical regression analyses to predict target vocabulary scores from the instructional condition, baseline vocabulary scores and home reading variables (e.g., frequency of home reading).

Several studies used MANOVA tests to analyze data that involved more than one dependent variable. This was the case for Carlo et al.’s (2004) study, which used six dependent measures for multiple tasks (i.e., word mastery, word association, polysemy
production, cloze procedures, morphology) and incorporated as predictor variables site, language status and the condition. Similarly, Roberts and Neal (2003) used six dependent variables related to early literacy skills (i.e., letter writing, letter naming, rhyming, vocabulary, story event sequencing, and concepts of print knowledge). In Ulanoff and Pucci (1999), a MANOVA test was mainly used to explore differences between three conditions (i.e., concurrent translation, preview-review, and no exposure) across three time points (i.e., pretest, posttest and delayed), which was subsequently followed by ANOVAs to investigate where any differences would lie. Comparably, Zhang and Schumm (2000) used MANOVA to analyze the differences between groups using instructional condition as the between-subject factor (i.e., 3 levels: keyword Spanish method, keyword English method and translation method) and time (i.e., 2 levels: immediate and delayed) and test (2 levels: sentence completion and cued recall tasks) as the within-subject factor.

More robust statistical approaches were used for studies with larger samples. Multilevel models were used for studies in which participants were organized at more than one level. In Lesaux et al. (2010), through a multilevel model, children were nested within classes as well as nested within schools. Within this model, posttest scores for each measure (i.e., reading vocabulary, target word mastery, word association, word-meanings-in-context, morphological awareness, and reading comprehension) were regressed on the condition variable (i.e., academic vocabulary instruction versus treatment as usual), and several covariates were applied including pretest scores for each measure and a dummy variable for language minority status (i.e., native English speaker served as the reference category). Similarly, Nelson et al (2011) used a three-level model
in which student scores on three measures (i.e., root word vocabulary, reading vocabulary, and word reading) were nested within small groups, which were in turn nested within school, and the corresponding pretest for each outcome was used as a covariate. Individual growth modeling was used by Uchikoshi (2006) to examine the differences in the level and rate of change among individuals. Unlike other approaches, this analysis is able to include all participants regardless of missing data and is able to analyze longitudinal data with varying numbers of waves of data. Predictors of vocabulary growth (i.e., receptive and expressive oral language) for this study included the condition (i.e., exposure to *Arthur, Between the Lions, or no show*), indicators of home television viewing, gender, parents’ educational history, pre-kindergarten experience, number of siblings, years in the United States, children’s initial vocabulary levels in Spanish and English, number of children’s books in the home, and library exposure. Classroom differences were also investigated in this study.

**Levels of Evidence for Effectiveness Across Studies**

Before addressing the second research question, it was important to first evaluate all available evidence relating to word-meaning acquisition. Subsequently, evaluating studies’ level of evidence of their effectiveness is an appropriate means of identifying studies with effect sizes that may be the result of biased outcomes. As such, the studies’ levels of evidence for effectiveness were evaluated based on several aforementioned methodological characteristics that comprise three design-related characteristics, which include random assignment, attrition, and baseline equivalence, and three outcome-related characteristics that include face validity, reliability, and outcome alignment. Based on the framework found in Figure 3, studies were examined to determine whether
evidence standards were met. Appendix H provides an overview of the proportion of studies that met the various evidence standards at each level through the use of the adapted framework, and details on how individual studies met evidence standards by level are found in Table 1. Four studies met evidence standards (Collins et al., 2004; Farver et al., 2009; Silverman & Hines, 2009; Zhang & Schumm, 2000); 11 studies met evidence standards with reservations (Atay & Kurt 2006; Carlo et al., 2004; Filippini, 2007; Giambo, 2004; Lesaux et al., 2010; Mancilla-Martinez, 2010; Nelson et al., 2011; Schunk, 1999; Spycher, 2009; Uchikoshi, 2006; Ulanoff & Pucci, 1999); and 3 studies did not meet evidence standards (Avila & Sadoski, 1996; Dockrell et al., 2010; Roberts & Neal, 2003).

At the first level, only experimental studies, which used randomization, and demonstrated low attrition levels, received the highest rating of *Meets evidence standards*. This was the case for nine studies, which randomly assigned participants to treatment and comparison conditions. Next, seven experimental studies that showed low attrition met evidence standards at the second level (Avila & Sadoski, 1996; Collins, 2004; Farver et al., 2009; Schunk, 2009; Silverman & Hines, 2009; Uchikoshi, 2006; Zhang & Schumm, 2000). High-attrition studies had to establish baseline equivalence in order to meet evidence standards with reservations at the second level. As a result, the two studies with high attrition stated that baseline equivalence was established (Carlo et al., 2004; Roberts & Neal, 2003). Quasi-experimental studies that did not meet this requirement were considered to not meet evidence standards (Avila & Sadoski, 1996; Dockrell et al., 2010). At the third level, outcomes used in each study had to be well aligned with both treatment conditions. Only one study revealed that a vocabulary outcome was heavily aligned with
the treatment condition (Roberts & Neal, 2003). As an experimental study that failed to meet acceptable attrition rates at level 2 and standards for “outcome alignment” at level 3, the Roberts and Neal (2003) study was deemed to not meet evidence standards. At the final level, studies had to provide an adequate description of the measures used to assess participants’ vocabulary knowledge in addition to meeting acceptable levels of reliability for each measure whether it is through a measure’s internal consistency, test-retest reliability, or inter-rater reliability. The only study that did not meet this standard was Schunk (1999). In sum, 6 studies did not meet evidence standards, 9 met evidence standards with reservations, and 3 met evidence standards (see Table 1).
<table>
<thead>
<tr>
<th>Study name</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atay &amp; Kurt (2006)</td>
<td>No</td>
<td>---</td>
<td>Yes</td>
<td>English achievement scores</td>
<td>Aligned</td>
</tr>
<tr>
<td>Avila &amp; Sadoski (1996)</td>
<td>No</td>
<td>---</td>
<td>No</td>
<td>Not reported</td>
<td>Aligned</td>
</tr>
<tr>
<td>Carlo et al. (2004)</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
<td>Met according to the What Works Clearinghouse</td>
<td>Aligned</td>
</tr>
<tr>
<td>Collins (2004)</td>
<td>Yes</td>
<td>Low</td>
<td>---</td>
<td>---</td>
<td>Aligned</td>
</tr>
<tr>
<td>Dockrell et al. (2010)</td>
<td>No</td>
<td>---</td>
<td>No</td>
<td>Non-verbal and phonological ability</td>
<td>Aligned</td>
</tr>
<tr>
<td>Study</td>
<td>Low wealthiest (Differential: NR; Overall: 2.1%)</td>
<td>Comprehension (IC: 0.81) &amp; Naming Vocabulary (IC: 0.80)</td>
<td>PCTOPP subtests: Receptive vocabulary (IC: 0.87) &amp; Expressive vocabulary (IC: 0.98)</td>
<td>Target Vocabulary test (IC: 0.84)</td>
<td>PPVT (IC: 0.95)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Farver et al. (2009)</td>
<td>Yes</td>
<td>Aligned</td>
<td>H</td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>Filiippini (2007)</td>
<td>No</td>
<td>Yes</td>
<td>English receptive and expressive scores</td>
<td>Target Vocabulary test</td>
<td>M</td>
</tr>
<tr>
<td>Lesaux et al. (2011)</td>
<td>No</td>
<td>Yes</td>
<td>No differences at pretest on vocabulary measures</td>
<td>SAT-10</td>
<td>M</td>
</tr>
<tr>
<td>Mancilla-Martinez</td>
<td>No</td>
<td>Yes</td>
<td>Nonsignificant at pretest</td>
<td>Multiple-choice test</td>
<td>M</td>
</tr>
<tr>
<td>Study</td>
<td>Yes/No</td>
<td>Low</td>
<td>High</td>
<td>No</td>
<td>Not met</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>----</td>
<td>---------</td>
</tr>
<tr>
<td>Nelson et al. (2010)</td>
<td>Yes</td>
<td>Low</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Roberts &amp; Neal (2003)</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>Not met</td>
<td>Over-aligned</td>
</tr>
<tr>
<td>Schunk (2009)</td>
<td>Yes</td>
<td>Low</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Silverman &amp; Hines (2009)</td>
<td>Yes</td>
<td>Low</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Spycher (2009)</td>
<td>No</td>
<td>---</td>
<td>Yes</td>
<td>No differences at pretest on vocabulary measures</td>
<td>Aligned</td>
</tr>
<tr>
<td>Study</td>
<td>Yes/No</td>
<td>Low/High</td>
<td>Pretest Outcome</td>
<td>Alignment</td>
<td>Measure</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Uchikoshi (2006)</td>
<td>Yes</td>
<td>Low</td>
<td>---</td>
<td>Aligned</td>
<td>PPVT-III (0.95)</td>
</tr>
<tr>
<td>Ulanoff &amp; Pucci (1999)</td>
<td>No</td>
<td>---</td>
<td>Yes</td>
<td>Pretest outcome of vocabulary as covariate</td>
<td>Aligned</td>
</tr>
<tr>
<td>Zhang &amp; Schumm (2000)</td>
<td>Yes</td>
<td>Low (Differential: 0%; Overall: 0%)</td>
<td>---</td>
<td>Aligned</td>
<td>Cued recall and Sentence Completion Tasks (IC: 0.91)</td>
</tr>
</tbody>
</table>

NR= Not reported

a IC= Internal Consistency; TRR= Test-retest reliability

b L= Low (Does not meet evidence standards); M= Medium (Meets evidence standards with reservations); H= High (Meets evidence standards)
Studies with Evidence of Effectiveness and Their Impact on Word Learning

Studies included in this review that met acceptable levels of evidence of effectiveness will form part of the discussion in this section, which addresses the second main research question. Furthermore, only posttest effect sizes for treatment-comparison contrasts that showed trivial effects at pretest—an indicator that no statistical differences exist between two conditions—along with those contrasts for which statistical adjustments were made will be included. The following analysis will discuss the effect sizes grouped by the measures used as well as by the degree of definitional and contextual support used in these studies. In addition to explanations of the significant differences existing between these conditions, standardized mean effect size estimates for vocabulary outcomes will also be used to quantify study findings and to provide systematic evidence of the treatments’ effectiveness in the included studies. To discuss effect sizes for each treatment-comparison contrast, descriptive statistics including means, standard deviations and sample sizes were necessary for each group unless secondary calculations, such as effect sizes and summary statistics, were provided to express mean differences in gain score between groups. Therefore, due to the nature of statistical analyses employed in certain studies, information about the magnitude of effect sizes could not be gleaned as it related to ELLs in particular conditions.

The following sections will discuss the evidence relating to studies with measures examining 1) breadth of knowledge for general vocabulary, 2) breadth of knowledge for curriculum-based vocabulary, and 3) depth of knowledge in curriculum-based vocabulary. These assessments were categorized as receptive and/or receptive measures, which will
also discussed in the next sections. No measures were used to examine depth in knowledge of general vocabulary. Appendices I through K provide summary information relating to the effect sizes described in the following sections.

**Breadth in Knowledge of General Vocabulary**

A total of six studies used measures to assess ELLs’ growth in breadth of definitional knowledge of general vocabulary while no studies tested the depth of general vocabulary. Researchers examined general vocabulary growth of ELLs’ knowledge using receptive and/or expressive standardized measures. Among these studies, Farver et al. (2009) and Uchikoshi (2006) used both receptive and expressive measures, Carlo et al. (2004) and Silverman & Hines (2009) solely used receptive measures; and Nelson et al. (2011) used a measure that combined receptive and expressive scores. Studies aiming to improve ELLs’ general vocabulary knowledge were all experimental studies ($k=7$) classified as two studies meeting evidence standards (Farver et al., 2009; Silverman & Hines, 2009) and three meeting them with reservations (Carlo et al., 2004; Nelson et al., 2011; Uchikoshi, 2006). Across all studies, which relied on ELLs’ definitional knowledge to assess how word meanings, effect sizes ranged from $-0.11$ to $0.99$ ($M=0.16$, $SD=0.26$). Although there was variance among the studies as shown across the multiple treatment-comparison contrasts in Figure 4, the performance of ELLs on receptive measures was paralleled on expressive measures. More specifically, contrasts for Farver et al. (2009) on receptive measures resulted in small to moderate effect sizes for the same groups contrasted on expressive measures, whereas no treatment effects were found on receptive measures and, comparably, on expressive measures in Uchikoshi (2006). The studies showing positive benefits on word learning differed in their levels of contextual
and definitional support. The vocabulary-related studies in this group that targeted decoding and vocabulary skills showed small to moderate effect sizes (Farver et al., 2009; Nelson et al., 2011) on general vocabulary. In contrast, vocabulary-focused studies that were centered on enhancing meaning-based skills, varied in their effectiveness, demonstrating large (Silverman & Hines, 2009) to no improvements on general vocabulary (Carlo et al., 2004; Uchikoshi, 2006). All conditions with context-only approaches found in Uchikoshi (2006) showed no significant improvements in word learning.

**Balanced approaches and breadth of general vocabulary.** The studies that implemented a balanced approach to vocabulary instruction (Carlo, 2009; Lesaux et al., 2011; Silverman & Hines, 2009) revealed conflicting evidence about their effectiveness on receptive vocabulary outcomes ($M= 0.01$, $SD=0.60$, range: -0.11 to 0.99). Although these three studies provided rich opportunities to engage with words that could facilitate incidental word learning, only Silverman & Hines (2009) had a significant and positively large effect ($d=0.99$) on ELLs’ receptive definitional knowledge of general vocabulary measured by the PPVT-III (Dunn & Dunn, 1997). On the other hand, Carlo et al. (2004) did not reveal significant or practically meaningful effects on receptive measures of general vocabulary ($d=-0.11$). The Carlo et al. (2004) and Lesaux et al. (2011) studies differed from the Silverman and Hines (2009) study by their comparison group types (i.e., “treatment as usual” versus diluted form); grade levels (i.e., $5^{th}$ grade and $6^{th}$ grade versus a multiple grades); and group size during instruction (i.e., mixed grouping versus whole group).
Figure 4. Effect sizes and confidence intervals for studies assessing breadth of general vocabulary
**Contextual approaches and breadth of curriculum-based vocabulary.**

Studies with contextual approaches also provided inconsistent results ($M=0.20$, $SD=0.24$, range: 0.06 to 0.71). When comparing context-only conditions (Uchikoshi, 2006)—whereby exposure was given to words in context with no attempt to have the learner derive a definition—with those providing contextual emphasis that provided some definitional information throughout instruction (Farver et al., 2009), the ELLs in conditions with contextual emphasis tended to exceed the performance of ELLs that were only afforded opportunities to incidentally learn words with no explicit or implicit instruction. The largest effect sizes on receptive ($d=0.71$) and expressive measures ($d=0.75$) were revealed by a condition in which significant L1 support was provided to review previously taught words. Two underlying factors were explained to be the main contributors to the differences between conditions in Farver et al. (2009): the use of small group instruction and the opportunities to revisit targeted words used in a familiar context.

**Definitional approaches and breadth of general vocabulary.** Although there are no studies with definition-only conditions, which strictly focused on supplying definitional information to increase general vocabulary, that assessed breadth of ELLs’ general vocabulary, Nelson et al. (2011) revealed small and positive effects on a measure of receptive and expressive knowledge of general vocabulary for a condition with a definitional emphasis (i.e., supplemental intervention designed to develop root word vocabulary and reinforce decoding skills) that provided minimal exposure to words in context by using a predictable instructional format with short lessons of decoding-relating activities and decontextualized meaning-based tasks (e.g., sentence completion, matching words to pictures). This condition was compared with a condition providing contextual
emphasis using shared book readings. The study authors revealed no significant
differences between both conditions on a measure of both receptive and expressive
definitional knowledge of general vocabulary (Woodcock Reading Mastery Test-
Revised/Normative Update; Woodcock, 1987). Although not significant, the study
authors reported that there was a trend favoring the condition with definitional emphasis,
resulting in a small effect size of 0.38 at posttest.

**Breadth in Knowledge of Curriculum-Based Vocabulary**

In order to examine ELLs’ knowledge of targeted words, 9 studies used
curriculum-based measures. From this group, 6 studies used breadth of knowledge
measures to assess growth in ELLs’ definitional knowledge of specific words targeted
during instruction, and 1 study examined ELLs’ contextual knowledge (Mancilla-
Martinez, 2010). These studies most consistently used receptive measures (Carlo et al.,
2004; Collins; 2004; Filippini, 2007; Mancilla-Martinez, 2010; Spycher, 2009; Ulanoff &
Pucci, 1999), with only 1 study using a measure that assessed both receptive and
expressive knowledge (Nelson et al., 2011). No studies used an expressive curriculum-
based measure to exclusively test ELLs’ breadth of word knowledge. As shown in
Figure 5, effect sizes across all studies in this group were moderate to large (\(M=0.92,
SD= 0.25, \text{range: 0.54 to 1.25})\).
Figure 5. Effect sizes and confidence intervals for studies using curriculum-based measures to assess breadth of word knowledge.
Balanced approaches and breadth of curriculum-based vocabulary. Four vocabulary-focused studies (Carlo et al., 2004; Lesaux et al., 2011; Mancilla-Martinez, 2010; Spycher, 2009), used conditions with balanced approaches that measured breadth of definitional knowledge for specific words. In all conditions, large effect sizes were reported ($M=1.24$, $SD=0.12$, range: 1.03 to 1.25). Spycher (2009) uniquely compared the condition with the balanced approach to a context-only condition, whereas the other three studies used “treatment as usual” conditions as comparison groups (Carlo et al. 2004; Lesaux et al., 2011; Mancilla-Martinez, 2010). In Spycher’s (2009) study, when comparing a condition with a balanced approach with a context-only approach that presented learners with the same content using implicit exposure, no statistical differences were found between both groups on a receptive assessment of tier 2 (i.e., general academic vocabulary) and tier 3 words (i.e., academic science vocabulary). Although this was the case perhaps due to small sample sizes, the effect sizes were positive and large for the balanced condition ($d=1.25$). In Carlo et al. (2004), Lesaux et al. (2011), and Mancilla (2010), treatment conditions compared to “treatment as usual” conditions were statistically significant on receptive measures of tier 2 vocabulary. Conditions with balanced approaches in these three studies, which had a cross-content focus that used implicit and explicit strategies throughout instruction, revealed small to large effect sizes (i.e., $d=1.03$ in Carlo et al., 2004; $d=0.39$ in Lesaux et al., 2011; $d=1.24$ in Mancilla-Martinez, 2010).

Contextual approaches and breadth of curriculum-based vocabulary. One study used contextual approaches as a means of improving ELLs’ breadth of specific words (Ulanoff & Pucci, 1999). Ulanoff and Pucci (1999) used two conditions with
contextual emphasis, which were compared with a context-only condition. The conditions with contextual emphasis provided definitional information in distinct ways: one used a preview-review method, whereby important points and difficult vocabulary from a storybook were previewed prior to a shared reading and reviewed thereafter, while the other condition used a concurrent translation method to translate the text throughout the process of conducting the shared reading. ELLs in the context-only condition were simply exposed to the storybook content with no elaboration. Results from a repeated measures MANOVA analysis revealed no significant treatment effects at pretest, but follow-up analyses showed significant differences for the preview-review method and the concurrent translation method at posttest and delayed posttest, with the preview-method showing greater performance. Since effect sizes were not provided, they were calculated using standardized mean differences found in the study, which revealed differences at pretest between treatment-comparison contrasts with the exception of one: the preview-review condition versus the context-only condition. As such, the other effect sizes were not interpretable. Unlike the results reported for the other conditions in the study, the effect size for the preview-review condition was positive and moderate at posttest \( (d=0.73) \) for a receptive measure of definitional knowledge for tier 1 words.

**Definitional approaches and breadth of curriculum-based vocabulary.** Two vocabulary-related studies targeted words throughout instruction used conditions with a definitional emphasis during instruction to measure ELLs receptive, definitional knowledge (Filippini, 2007; Nelson et al., 2011), which revealed moderate to large effects \( (M=0.64, SD=0.26) \). Filippini (2007) compared three instructional conditions with: 1) phonological awareness and decoding (PAD) training; 2) PAD plus semantic
relations instruction and 3) PAD plus morphological awareness instruction. As stated by the study authors, ANCOVA with pretest scores as a covariate on the Tier 2 receptive vocabulary measure revealed that the mean differences were not statistically reliable since they were sufficiently large to protect against their influence. Although this was the case, the conditions in which vocabulary was instructed in conjunction with decoding instruction outperformed students that only received decoding instruction (i.e., PAD+morphology awareness: \(d=0.21\); PAD+semantic relations: \(d=0.54\)).

Nelson et al. (2011), which also used a condition with a definitional emphasis (i.e., a supplemental intervention designed to develop Tier 1 root word vocabulary and reinforce decoding skills), used a comprehensive, curriculum-based measure that combined receptive and expressive scores. When compared with a condition with contextual emphasis using shared book readings, significant treatment effects on Tier 1 vocabulary (i.e., root words) were found, indicating a positive and large impact on a receptive, curriculum-based measure of definitional knowledge (\(d=1.04\)).

**Depth in Knowledge of Curriculum-Based Vocabulary**

A limited number of 4 studies used depth of knowledge measures to test multiple meanings known for specific words, which also consisted of 3 studies using receptive measures (Carlo et al., 2004; Mancilla-Martinez, 2010; Silverman & Hines, 2009), 2 studies using expressive measures (Carlo et al., 2004; Silverman & Hines, 2009), and 1 study using a comprehensive measure of receptive and expressive knowledge (Atay & Kurt, 2006). As shown in Figure 6, depth of knowledge measures varied by the type of knowledge (i.e., definitional and/or contextual) and ability (i.e., receptive and/or expressive) assessed. Effect sizes for this group ranged from 0.15 to 0.59 (\(M=0.38\),
with marginal to moderate effects on the depth of words learned after instruction. All studies in this subset used a balanced instructional approach, with the exception of Atay & Kurt (2006), which used a condition with a definitional emphasis.

Among the studies using a “treatment as usual comparison” group, none showed statistical differences in the treatment-comparison contrasts; however, there are mixed results in performance among ELLs in these studies, with only Lesaux et al. (2011) not showing a meaningful impact on ELLs’ depth of knowledge. Although Carlo et al. (2004) and Lesaux et al. (2011), used a similar receptive measure of definitional knowledge on a word association task in which learners were asked to identify words that are most closely related to a target word and no significant differences were found in these two studies between treatment and “treatment as usual” conditions, Carlo et al. (2004) revealed a positive, albeit small, impact ($d=0.44$) while Lesaux et al. (2011) and did not increase ELLs’ depth of knowledge ($d=0.15$). An additional depth of knowledge measure was used in Carlo et al. (2004) to examine learners’ contextual knowledge by producing as many sentences conveying the different meanings of polysemous words. Findings show that treatment condition with balanced instruction also did not differ from the “treatment as usual” condition on this expressive measure of contextual knowledge but showed a moderate effect size (0.59). Mancilla-Martinez (2010), which also had a “treatment as usual” comparison group, assessed ELLs’ depth through a receptive measure that combined definitional and contextual knowledge. On this task, ELLs were asked to provide a rating on how well they know a word and whether they would be able
Figure 6. Effect sizes and confidence intervals for studies using curriculum-based measures to assess depth of word knowledge.
to use it in a sentence. Despite not showing significant differences between both conditions, results on this measure showed moderate effects of the condition with balanced support over the “treatment as usual” condition ($d=0.59$).

Silverman and Hines (2009), which compared a condition with balanced support to a similar condition, used an expressive measure of depth of knowledge in which definitional statements rather than examples sentences (i.e., using words in context) were produced. Findings in this study show significant differences between the treatment condition with a balanced approach (i.e., multimedia condition) and the comparison group, which used a similar approach (i.e., non-multimedia condition), on a receptive measure of definitional and contextual knowledge. The treatment effects for this measure were large ($d=0.97$). However, their performance on an expressive measure of definitional knowledge did not differ from the comparison group ($d=0.19$).

**Definitional approaches and depth of curriculum-based vocabulary.** Only one study used definitional approaches to deepen ELLs’ understandings of words (Atay & Kurt, 2006). In this study, the two conditions mainly differed by the level of interaction taking place in a follow-up activity, which was either achieved through a small-group session or individually, depending on the condition. The treatment condition that provided more interaction significantly differed from the diluted comparison group and showed a positive and small impact on ELLs’ definitional and contextual knowledge ($d=0.31$).
Chapter 6: Synthesis and Discussion of the Findings

Exposure to meaning-based activities and participation in word-focused interactions have shown to develop ELLs’ knowledge of word meanings in varied ways. However, in order to understand the impact of the instructional approaches under investigation and the extent of the opportunities afforded to ELLs to learn word meanings, multiple design, treatment, learner, and outcome variables should be considered. Thus, a central goal of this study was to appraise the available research on vocabulary learning in order to make a critical assessment of the variables that are currently represented in the field. Specifically, the research questions guiding this systematic review were: (1) How do intervention studies developing L2 word-meaning acquisition with ELLs in pre-kindergarten through 6th grade settings vary in their methodological characteristics (i.e., research design, learners, instructional design and outcome)?; (a) What are the studies’ levels of evidence for effectiveness based on their design and outcome characteristics?; and (2) What is the impact of studies that met requisite evidence standards on L2 word-meaning acquisition? To answer these research questions, a multi-layered investigation was conducted that involved identifying relevant studies that met several inclusion criteria, assessing the included studies’ methodological characteristics, investigating studies’ level of evidence for effectiveness, and analyzing the findings of studies demonstrating the best research evidence.

The results of this research help to add to the literature in areas that may not be fully informed by the findings from reviews of studies primarily consisting of monolingual English populations (Marulis & Neuman, 2010; National Institute of Child Health and Human Development, 2000; Stahl & Fairbanks, 1996). This review required
that included studies show that ELLs outnumber monolingual participants in studies with heterogeneous samples or must provide data that strictly relates to this population’s vocabulary outcomes (e.g., by providing disaggregated data), hence, findings will provide insights into effective instructional practices and those that are in need of further investigation due to the state of the field. This study aims to bring to the forefront the need to continuously develop, refine and report on interventions that impact learning outcomes with ELLs. In so doing, these investigations are able to focus the work of researchers who in collaboration are able to work toward an initiative that helps to resolve questions relating to these diverse learners’ development by building on prior theoretical and empirical work.

Through one of the most noteworthy reports on literacy research, which has had significant and widespread influence on education policy and practice, the findings of the National Reading Panel (NRP; National Institute of Child Health and Human Development, 2000) were a result of an evaluation of extant research to determine the best evidence-based methods to teach children to read. Upon examining over 100,000 studies that span over more than fifty years, they identified several principles underlying the best approaches to reading instruction with the conscious exclusion of second language research targeting ELLs’ language and literacy development. Upon narrowing their search for studies using vocabulary instruction to promote reading comprehension, only 50 studies were identified for further evaluation, which included studies that dealt with a relatively large number of variables. To supplement the evident gaps of research, a group of second language researchers convened to form the National Literacy Panel on
Language-Minority Children and Youth to address issues of learning and instruction in non-English speaking children by further investigating whether the approaches in the NRP report would apply to ELLs.

As pointed out in the report by the National Literacy Panel on Language-Minority Children and Youth, there is a critical need to move thinking about second-language literacy development beyond the simple frameworks that do not appropriately accommodate the complex processes that interact dynamically across grades as ELLs acquire literacy (August & Shanahan, 2006). In their search, these researchers found only three experimental studies of English vocabulary learning, in contrast with the 45 experimental studies of vocabulary teaching with first-language students (National Institute of Child Health and Development, 2000). Given the crucial role of vocabulary in reading and academic success, these findings along with the evidence provided in this synthesis reveal an enduring and overwhelming need for vocabulary interventions that could yield more evidence of effective practices with ELLs.

The discussion of the findings in this systematic review are aimed at expanding knowledge in this research area but also at intensifying the need to support further research in language and literacy development that is specific to ELLs. The following discussion will be guided by Stanley and Campbell’s (1963) typologies for independent variables found in experimental studies. In this manner, these dimensions of manipulability among independent variables will be considered: (1) manipulated variables, such as the instructional approach used, which are assignable by the person delivering instruction; (2) relatively unchanging and fixed variables (i.e., aspects of the learners’ environment), such as age, gender and SES, which are not under direct control
of the person delivering instruction but may serve as the basis for matching or stratification in the study; (3) response characteristics, such as scores at baseline and posttest. The primary interest in research is to determine how manipulated variables may impact an outcome while keeping other variables constant (i.e., variables 2 and 3). In the following section, an in-depth discussion of the findings will be presented in an effort to underscore the variables that were addressed effectively to meet evidence standards in addition to identifying current gaps of knowledge in the field.

**Fixed Variables Relating to Learner and Setting Characteristics**

Research findings based on a sample that is not well-defined holds less value than when studies provide detailed descriptions that articulate both observed and unobserved characteristics. Because learner variables are taken into greater consideration in well-executed experimental studies, external validity becomes increasingly possible. In the included studies though, only a small number detailed descriptive information relating to the study samples that would facilitate readers in determining the applicability of study findings with particular ELL populations. More specifically, ELL populations were generally not differentiated in a manner—such as dividing samples into subgroups for further analysis—that would highlight the effectiveness of particular treatment conditions that are effective with ELLs of varying bilingual proficiency as well as with diverse cultural and socioeconomic backgrounds.

**Demographic and contextual factors.** Although ELLs are frequently able to develop low-order skills relating to decoding at the same extent as their native English speakers in the earlier grades (August & Shanahan, 2006), factors relating to the instructional contexts and their surrounding communities have not been thoroughly
investigated to determine their influence on ELLs’ word-meaning acquisition. For children whose native language is the dominant language used in their communities, thereby using and developing language and at times literacy skills in their L1, they are likely to follow a different trajectory of language and literacy development than their native English peers. The intensity of exposure to English words may be contingent on demographic factors, such as SES, as well (Hart & Risley, 1995). In previous research with native English speakers, children from middle- and upper-income homes were more likely to benefit from vocabulary intervention than those from low socioeconomic backgrounds (Marulis & Neuman, 2010). Yet, the limited studies that have been conducted examining demographic factors on ELL literacy development have mainly shown correlational evidence relating SES and length of second-language schooling with L2 reading performance (August & Shanahan, 2006).

In this research synthesis, the included studies did not commonly address how these factors impact word learning throughout the instructional period, which have shown to have long-term effects on later achievement. Although certain studies included SES data in statistical analyses to determine their predictive value on vocabulary outcomes (Collins, 2004; Silverman & Hines, 2009; Uchikoshi, 2006), no studies conducted subgroup analyses based on these characteristics. As revealed in the available evidence, the reviewed studies vary greatly in the extent to which they reported information relating to the demographic context. Also, studies reported information relating to students’ SES based on the city or district in which they lived (Avila & Sadoski, 1996; Carlo et al., 2004; Dockrell et al., 2010; Spycher, 2009); on the school they attended (Filippinni, 2007; Lesaux et al., 2010; Mancilla-Martinez, 2010; Uchikoshi, 2006; Ulanoff & Pucci,
Since little research has been conducted examining the relationship between SES and L2 word-meaning acquisition, it remains inconclusive on how demographic indicators (i.e., relating to individuals, schools, and city) and different levels of SES impact on vocabulary development. As noted by White (1982), SES at the aggregate level has a stronger relationship to school achievement than at the individual level. Therefore, not only would subgroup analyses examining ELLs’ individual SES provide fruitful contributions to the field but also further exploring overall school and city performance may elucidate the impacts on their vocabulary development, which may be contingent on their in- and out-of-school experiences.

Additionally, since many of the classrooms were selected on the basis of convenience and little information was provided with regard to sampling procedures, researchers must infer a population to which the results might generalize. One large-scale study in this review (Carlo et al., 2004), however, uniquely selected multiple classrooms across the US that consisted of diverse groups of one target population (i.e., Spanish-speaking ELLs), thereby increasing population validity for this group. In other cases, small-sample studies improved generalizability by explicitly stating their sampling procedures, although this was an infrequent practice (Collins, 2004; Filippini, 2007). In these studies, descriptive data, such as grade level, SES, proficiency levels in both languages and age helped to make inferences relating to the target population.

**Linguistic and cultural factors.** When exposing students to vocabulary, instruction that is responsive to students’ linguistic and cultural needs is critical in
helping them bridge existing understandings and novel information in the classroom (Biemiller & Boote, 2006). Language proficiency for any learner may not only be indicative of their vocabulary knowledge but also may provide evidence as to how their current knowledge may facilitate later learning. As previously discussed, learners with high vocabulary proficiency tend to have denser and more organized semantic networks than those exhibiting lower proficiency. For ELLs who develop varying degrees of proficiency in both of their languages though, previous evidence has shown that the relationship between their vocabulary knowledge in their two languages differs at different levels of L1 and L2 proficiency (Kaivanpanah & Alavi, 2008; Lanauze & Snow, 1989). However, after evaluating the intervention studies in this synthesis, there is still inconclusive evidence relating to the interaction between instructional methods and varying L1/L2 proficiencies. This sample of studies generally provided holistic information rather than investigating how particular ELL subgroups improved on vocabulary outcomes due to specific features found throughout the treatment condition.

Additionally, much of the evidence in this review addresses the vocabulary learning needs of Spanish-speaking ELLs while certain studies that included ELLs who speak non-alphabetic languages typically formed part of heterogeneous language samples that did not receive differentiated instruction based on their unique cultural and linguistic backgrounds (Dockrell et al., 2010; Lesaux et al., 2010, Roberts & Neal, 2003, Schunk, 2009, Silverman & Hines). Instead, their data was aggregated in study findings with other ELLs who may differ on linguistic and cultural experiences. As explained by
Wong-Fillmore and Snow (2000), culturally responsive practice with ELLs involves familiarity with students’ home languages to apply appropriate word-learning strategies (Wong-Fillmore & Snow, 2000).

The interdependence of ELLs’ two languages illustrates the importance of both first-and second-language sources of influence on second-language development, particularly for academic and higher order cognitive purposes (Cummins, 1981). As such, culturally and linguistically responsive teaching may look different based on students’ backgrounds. For instance, Spanish-speaking ELLs, whose first and second languages show more structural and functional similarities than learners of typologically distinct languages (e.g., Hmong, Turkish), may benefit more greatly from instruction that highlights these relationships (e.g., cognate instruction). As shown by Hancin-Bhatt & Nagy (1994) for speakers of Romance languages, these learners may be able to access meanings of words in their L1 that are rarely spoken in English but are more frequent across academic texts. Even though L1 supports were implemented in a few studies, only one study targeted more difficult-to-learn academic vocabulary through cognate instruction (Carlo et al., 2004).

Furthermore, no intervention studies have addressed the unique ways in which younger ELLs that speak typologically distinct languages and how they may differentially acquire word meanings. Studies with older ELLs, not included in this sample, using effective vocabulary teaching methods (e.g., using a bilingual dictionary; Li, 2012) may perhaps provide insights into how to help these learners mediate their understandings between their two languages as they encounter L2 words. Due to the diversity in language-specific features in their languages and cultural knowledge among
ELLs, further investigations into addressing the vocabulary learning needs of this diverse population are needed as the teaching practices for native English speakers commonly used across the included studies may be relevant for ELLs but are insufficient to meet their unique needs.

**Manipulated Variables Relating to Vocabulary Instruction**

The central purpose of studies exploring the effectiveness of teaching approaches is to manipulate instructional variables in the efforts to improve learning outcomes among a target population. With this in mind, researchers commonly delineate treatment characteristics as a means of showcasing the instructional features that are under investigation. This review investigated the extent to which these features were articulated and how the interplay of treatment characteristics may influence vocabulary growth. Despite the results shown in the analysis of the findings, caution should be taken when reading the following discussion points and in generalizing these findings since only a limited number of studies met evidence standards and provided interpretable effect sizes.

**Vocabulary teaching as an instructional factor.** The overarching goal of this systematic review was to explore how conditions would compare on vocabulary outcomes. Generally, studies differed along two dimensions: whether they were vocabulary-related studies (Dockrell, 2010; Farver et al., 2009; Filippini, 2007; Nelson et al., 2011) examining word-meaning acquisition as a result of instruction that emphasized comprehension-oriented skills along with other literacy skills (i.e., low-order), such as decoding, while vocabulary-focused studies tested particular elements that made up word-focused instruction. As revealed by the National Reading Panel’s report (2000), the
most effective method for teaching children to read is by promoting both low-order, word-level skills and high-order, text-level skills. It remains unclear, however, how best to allocate instructional efforts with ELLs whose dynamic backgrounds may require differentiated support depending on their first and second language experiences.

Unlike monolinguals, ELLs—who still use their L1 in their out-of-school experiences while receiving formal instruction in their L2—face the dual task of learning to understand and use words in two languages. In addition to this, ELLs are expected to learn to read (i.e., by developing skills in phonological awareness, letter knowledge, phoneme-grapheme relationships, spelling rules, fluency) and read to learn as they advance through grade levels (i.e., by developing skills in vocabulary, cultural knowledge, comprehension strategies, purposes for reading, discourse structure). Certain factors, such as students’ previous L2 schooling experiences and word-level skills, may deter instructors from allocating instructional time to meaning-based activities in deference to decoding-related instruction (Biemiller, 2005). Nevertheless, word reading skills become quickly and well developed by the middle elementary grades for ELLs, whereas the vocabulary gap may remain between ELLs and their native English-speaking peers due to fewer opportunities for exposure with L2 vocabulary (August & Shanahan, 2006).

Greater efforts over the past decade have been taken, however, in identifying effective instructional methods in order to support both word-level skills and text-level skills (e.g., vocabulary) through instruction with ELLs. As mentioned in this review, vocabulary-related studies varied greatly on the lower-order skills they aimed to develop. Those vocabulary-related studies that met evidence standards in particular strove to enhance decoding skills (Farver et al., 2009; Filippini, 2007; Nelson, 2011), narrative
skills (Dockrell et al., 2010), and morpho-syntactic knowledge (Dockrell et al., 2010). In these studies, the benefits of delivering both reading and comprehension-oriented instruction revealed the possibility of enhancing vocabulary knowledge alongside other literacy skills. However, due to the limited number and the diverse literacy skills targeted in these studies, further evidence is needed to better understand the contributions of how the combinations of word-level and text-level skills may impact vocabulary learning with ELLs who vary in L1/L2 proficiency and require varying levels of support in developing word-level skills.

In the same vein, those studies identified as vocabulary-focused studies in this review also differed greatly in their approach. To increase word learning, these studies implemented the following: generative word-learning approaches (Atay & Kurt, 2006; Avila & Sadoski, 1996; Schunk, 1999; Zhang & Schumm, 2000), packaged instructional approaches with multiple word-focused components (Carlo et al, 2004; Lesaux et al., 2006; Mancilla-Martinez, 2010), multimedia approaches (Silverman & Hines, 2009; Uchikoshi, 2006), the use of multiple explicit strategies versus incidental exposure (Collins, 2004; Spycher, 2009), and approaches that centered on the use of bilingual methodologies in context-dependent ways (Ulanoff & Pucci, 1999). Zhang and Schumm (2000) published the only study that was an offshoot of the work initiated by Avila and Sadoski (1996). In so doing, these two studies uniquely provide greater a degree of understanding of the ways in which vocabulary learning may differentially improve through the keyword method that exclusively uses students’ L1 or L2 during instruction.

In both the vocabulary-centered and vocabulary-related studies included in the final analysis, similar benefits were seen across studies in these two groups on breadth of
general vocabulary and curriculum-based measures. This may lend support to the notion that vocabulary learning may be enhanced to the same degree when integrated with teaching other literacy skills. However, it is evident that more research is needed in these areas.

Factors relating to instruction in treatment-comparison group contrasts. At a macro level, studies that contrasted treatment-comparison conditions that predominantly shared similar features but slightly varied on a variable of interest provided more substantial evidence that differences in vocabulary learning were attributable to the underlying factor that distinguishes these groups. This was more compelling for those studies in which diluted forms of treatment conditions were used as comparison groups (Atay & Kurt, 2006; Avila & Sadoski, 1996; Collins, 2004; Farver et al., 2009; Filippini, 2007; Nelson et al., 2011; Silverman & Hines, 2009; Spycher, 2009). On the other hand, studies that used “alternative treatment” comparison groups may differ too greatly in their approach to identify features that may have contributed to word learning.

Studies with “treatment as usual” comparison conditions, which were less straightforward when describing the ways in which instructional practices differed from those found in the focal treatment condition, tended to compare multi-componential instructional approaches with classrooms that continued to receive their regular instruction. Since no supplementary information was provided relating to instructional procedures and how the instructional time was used in these comparison groups, it is difficult to ascertain whether extraneous variables had been controlled by the researcher, so that any observed effect can be attributed solely to the many instructional features
tested in these conditions. Although certain studies included treatment-comparison group contrasts that slightly differed on a treatment variable within a study, these uncommon approaches pose a challenge when synthesizing findings across studies. Along these lines, examining the degree to which contextual versus definitional support may facilitate word learning correspondingly remains unclear.

**Contextual and definitional support as a factor in word learning.** In the summary findings of the National Reading Panel’s report (2000) for vocabulary instruction with monolingual English speakers, a combination of definitional and contextual approaches was more effective than solely using any one method. To assess whether these findings apply to ELLs, the studies that met sufficient levels of evidence for effectiveness in this review were analyzed with respect to how contextual/definitional information supported learning, which resulted in inconclusive evidence partly as a result of the limited number of studies that have been conducted to increase word learning.

Even though large numbers of interacting variables were evident in the limited number of treatment-comparison group contrasts that were analyzed, of the few treatment-comparison group contrasts available, all definitional approaches showed promising benefits to word-meaning acquisition on all outcome measures. In contrast, balanced approaches were used more frequently but showed inconsistent on standardized measures assessing general vocabulary and curriculum-based assessing depth of word knowledge; however, on curriculum-based measures assessing breadth of word knowledge showed positive effects. Notwithstanding the evidence described for the effect sizes in the findings of this synthesis, the various approaches (i.e., definitional, balanced and contextual) cited are too infrequently investigated to draw any definitive conclusions
about their overall effectiveness. Additionally, although approaches may appear to be comparable in many ways (e.g., those using balanced approaches that include multiple exposures with tier 2 academic vocabulary and mixed group configurations in Carlo et al., 2004 and Lesaux et al., 2011) for particular outcome groupings (e.g., studies that measured the same construct and used a balanced approach), their comparison groups exhibit considerable disparities that may confound the results.

In the attempts to explore how particular instructional strategies supported learning within broader instructional approaches, the elements that comprised treatment and comparison conditions were contrasted. However, determining the impact of individual strategies becomes even less apparent when analyzing their differentiating characteristics at the micro-level for the included studies. Thus, the interplay of particular elements that form part of a treatment condition, such as the use of L1 bridging strategies, multiple exposures or the types of target words being instructed, are in need of further investigation through studies that contrast more comparable instructional conditions. Disparities in the treatment characteristics within individual studies were so commonplace that it rendered difficult the possibility of finding evidence to substantiate claims relating to best practices for ELLs.

Response Characteristics

Like the fixed variables relating to learner and setting characteristics, outcome variables serve to increase precision and reveal how generalizable the effects of manipulated variables are (Campbell & Stanley, 1963). Consequently, multiple criteria were used to identify studies that avoided possible confounds in order to isolate the effects of a treatment condition. An assessment of whether outcome measures were both
reliable and aligned with the objectives of the study coupled with the use of equivalent
groups at baseline helped to avoid threats to internal validity. Knowing that participants
in a study are equivalent at baseline and that outcomes were effectively measured helped
to support evidence that particular treatment variables may be the basis for students’
vocabulary learning. As discussed in this systematic review, only those treatment-
comparison groups contrasts were included that formed part of a group of studies that met
standard levels of evidence for effectiveness. Since this was the case, only a small
number of outcomes across studies were further analyzed.

**Baseline equivalence.** Treatment-comparison group equivalence is essential to
establishing the internal validity of a research study. In this review, several areas were
considered prior to examining the standardized mean differences on vocabulary outcomes
between treatment-comparison conditions. Among the levels of evidence for
effectiveness applied in this study, criteria relating to baseline equivalence were satisfied
through random assignment, matching learners on vocabulary-related outcomes (e.g.,
through measures of oral English proficiency), as well as ensuring that attrition rates were
low enough to not impact outcomes at posttest.

As evidenced by the findings in this review, failure to satisfy baseline equivalence
was a recurring issue that heavily impacted the findings in this review. Although study
authors provided evidence that treatment-comparison groups were matched on outcomes
that were distally related to vocabulary (e.g., English achievement scores; Atay & Kurt,
2006), conflicting profiles of the participants in both groups were later discovered upon
calculating effect sizes for the outcomes of interest (i.e., vocabulary) at pretest. Consequently, significant differences that existed at pretest would expectedly affect the study outcome.

**Study outcomes.** Outcomes showed to vary on multiple dimensions across studies, which resulted in a need to analyze them by the constructs study authors purported to measure. Most notably was the common use of measures of general vocabulary that tested ELLs’ definitional knowledge of tier 1 words, which accounted for growth in concrete, easily acquired vocabulary that is used in everyday communication. Given the extensive use of academic vocabulary (i.e., tier 2 and tier 3 words) in written texts as students progress through school (Beck et al., 2002; Graves, 2006), assessments examining ELLs’ growth of non-targeted words (i.e., general vocabulary) in these domains were underrepresented, with only one study investigating growth in tier 2 reading vocabulary (Lesaux et al., 2011). Additionally, standardized assessments that were regularly used to assess general vocabulary, such as the PPVT-III (Dunn & Dunn, 1997) and the SAT-10 (Harcourt, 2004) found in these studies, are normed on monolingual English populations and do not take into account issues relating to cultural sensitivity. As such, an ELLs’ existing conceptual knowledge may not manifest itself on these measures.

Confirming the report by National Reading Panel (2000), there was little overlap in the types of measures used, including when more common standardized measures were utilized. Like the results found in previous syntheses, standardized assessments were not consistently sensitive to the changes in vocabulary growth at post-test (National Reading Panel, 2000; Stahl & Fairbanks, 1986). Conversely, researcher-created, curriculum-based
measures were more susceptible to exhibiting the nuances and complexities involved in ELLs’ vocabulary acquisition. Uncommon in previous syntheses, descriptions of the types of words assessed was an element that was heavily considered throughout this study. A notable finding was the rise in the number of studies implementing the use of tier 2 vocabulary assessments.

Through curriculum-based measures, researchers were able to design assessments that were sensitive to ELLs’ growth in breadth and depth academic words (i.e., tiers 2 and 3) encountered throughout instruction. Most notably, depth of knowledge measures, which have not been considered in any review thus far relating to vocabulary instruction, are able to exhibit the degree of changes in learners’ ability to demonstrate their knowledge in various ways, such as showing understandings of paradigmatic versus syntagmatic associations between semantically related words, providing definitional statements that highlight words’ various meanings, and supplying example sentences that illustrate the more nuanced ways in which words are used in multiple contexts. Depth of knowledge measures were used to capture the amount of semantic information known for a word in only a few studies (Atay & Kurt, 2006; Carlo et al., 2004; Mancilla-Martinez, 2010; Spycher, 2009). Most surprising though were the inconsistencies across all outcomes in the ability levels measured, types of words assessed (i.e., depended on the instructional content of the study), modality used to exhibit their knowledge, and the type of knowledge examined (i.e., definitional and/or contextual). These interacting variables relating to outcome measures like the many expressed throughout the discussion of other methodological characteristics in these studies further obfuscate the objectives of a
research synthesis, which is to combine and make sense of the findings from which to draw conclusions while concurrently identifying conflicting evidence that may serve as an impetus for future research.

Implications

In order to ascertain the role of instructional practices on vocabulary learning, appropriate inferences of study findings are possible when closely related variables, relating to the design of the study, instructional conditions, learners, and outcome measures are compared. Due to the wide variations in these variables, further research is necessary to elucidate the benefits of particular instructional approaches. Since little research has been conducted in the form of experimental design with this population, research that controls for extraneous variables should be conducted to reveal task-specific influences on vocabulary learning.

The results of this study suggest several implications regarding the design of future studies. However, the overall funding structure for educational research should first be considered. Over the course of the past decade after receiving increasing support through funding by the Institute of Education Sciences (IES), researchers of language-minority students have ramped up efforts in support of education policy and practices that are responsive and sensitive to ELLs’ linguistic and cultural needs. Due to the relative advances in the area of vocabulary that have better focused researchers over the past decade (National Reading Panel, 2000) along with the more recent call for investigations with ELL samples (August & Shanahan, 2006), it is not surprising to find that much of the existing intervention studies with ELLs have been exploratory in nature or in their developmental phase. As described by the National Center for Education Research, one
of the four centers within the IES, applicants must adhere to a five-part goal structure (i.e., Exploration, Development and Innovation, Efficacy Trials, Scale-up Effectiveness Trials, Measurement) when submitting proposals for federally funded research. With this in mind, it is evident that due to the many methodological limitations faced by education researchers who have predominantly conducted interventions that provide empirical justifications for developing and refining interventions to produce impacts on student outcomes in authentic education delivery settings (Goal 2: Development and Innovation), there is a critical need to move toward conducting a greater number of efficacy and replication studies (Goal 3). In this effort, research through these longer-term projects could contribute to existing knowledge by enabling researchers to evaluate how a different set of conditions, such as a change in sample or on procedures for implementation, may impact a previous evaluation of less controlled, smaller-scale interventions.

Researchers could potentially build on previously conducted interventions included in this review and identify the organizational supports, tools and procedures that are needed when conducting more controlled experiments. Consequently, with additional resources, many of the discussion points revealed in this systematic review could be better addressed in future intervention studies. For instance, greater efforts can be taken to disentangle the effects of individual conditions, which were frequently packaged with multiple instructional features, by enabling researchers to identify the underlying differences between treatment and comparison conditions in previous studies before newly conducting the intervention at the evaluation stage.
Another point of contention is how to ensure generalizability given the methodological limitations relating to available contexts and research participants. Under ideal conditions, researchers who could also potentially conduct randomized-controlled trials would better eliminate biased estimates through random assignment. Since researchers have more control of sampling procedures in these cases, they may also be able to better establish baseline equivalence. As illustrated in this study when evaluating studies on their levels of evidence for effectiveness, although attempts were made to establish baseline equivalence using a measure that was analogous to the outcomes under analysis to form equivalent groups, initial differences frequently remained at posttest. With more resources, researchers may be able to control for these differences either by sampling the target population once again or by making statistical adjustments through covariates that are correlated with the outcome variable when initial differences are within a small range. However, when differences exceed certain baseline differences (i.e., 0.25 standard deviations at pretest), which was frequently the case, the samples in the treatment-comparison groups may be too divergent for meaningful generalizations to be made (What Works Clearinghouse, 2014). As such, researchers should be wary of sampling techniques used with ELL populations who may vary in proficiencies in their two languages and reveal significant differences in their language outcomes (i.e., vocabulary). When resources remain scarce and certain challenges are faced that disallow the use of random assignment, future research that falls into the quasi-experimental category could avoid threats to internal validity through researchers’
explicit articulation of sampling procedures as a means of better informing the field of how students may have differed from one another on unobservable and observable characteristics.

This study also aims to inform the field of the ways in which future researchers should consider multiple variables relating to learner characteristics that represent the dynamic nature of ELLs’ linguistic and cultural knowledge. Perhaps the greatest challenge in U.S. classrooms is in addressing the learning needs of these students. Due to the inherent flaws in the procedures used to identify ELLs, which may vary across schools, districts and states, focusing on how to support these learners’ long-term development tends to go unheeded. A recurring issue stems from the unstable nature of the ELL category, as higher-performing ELLs are able transition out of the ELL category upon meeting certain achievement criteria while only lower-performing ELLs persist in this group. In effect, efforts to narrow the achievement gap between ELLs and those who meet achievement criteria become implausible as the sources of difficulty faced by these students are seldom directly addressed. Although certain students become reclassified due to more rapidly meeting proficiency requirements in the earlier grades, a greater concentration of students struggle to meet proficiency standards on content area assessments in the upper grades. As shown through national assessment data, students who were formerly designated as ELLs do not fare much better in the long-term despite having previously met requisite standards that would enable them to exit the ELL category (National Center for Education Services, 2009).

Accordingly, more studies are needed that take a more comprehensive approach in identifying the needs of these students who follow unique development pathways, have
diverse backgrounds (e.g., SES), and vary on their levels of proficiency in their L1 and L2. Through subgroup analyses, future researchers may be able to distinguish how levels of proficiency and other demographic factors may differentially impact word-meaning acquisition. Additionally, although most studies in this review were conducted with Spanish-speaking ELLs, few studies were sensitive to the linguistic and cultural needs of the diverse ELL population. Little evidence in the way of experimental research has been conducted to explore instruction that highlights specific features that are characteristic of the languages of ELL populations. As such, researchers could potentially explore instructional practices that may be supportive of vocabulary learning and better attribute these gains to a target ELL population’s needs.

Although the findings in this review reveal inconsistencies across studies, researchers should consider applying existing understandings to assure that results are reliable and valid in future studies. Replication of all studies, including those that did not meet evidence standards, have the potential to empirically support the results of original studies and better disentangle the effects of particular instructional features that may contribute to students’ performance. Researchers should also consider the use of multiple vocabulary outcome measures to provide a more comprehensive look at how curriculum-based as well as general vocabulary words alike may be impacted throughout the instructional period. Additionally, to consider the degrees to which words are known, depth of knowledge measures and using both receptive and expressive measures will bring forth more evidence of the richness of ELLs’ vocabularies. This, in turn, will provide a better gauge of how to best support ELLs and given their current understandings of particular words.
Limitations of Study

While this systematic review provides a base for future research in vocabulary with ELL populations, there are several limitations that restrict the usefulness or generalizability of the results. The results of this study only allowed for a description of the variables of interest based on the available information reported by study authors across the included studies. Consequently, there is a possible lack of reliability of these descriptions. Such variations in the level of details provided may have led to understating the benefits of certain methodological characteristics for particular studies while more thorough accounts of intervention procedures in other studies enabled a deeper analysis of the findings.

Due to this relatively underrepresented area of research in the field, further investigations are needed to better examine and control for the variables associated with instruction, learners, settings, and outcomes. Given the wide range and dimensions of outcome measures used in these studies, it is apparent that generalizing these findings must be done with caution. As previously shown, the assessments used to explore learners’ word-meaning acquisition differed greatly on the objectives of the measures, the abilities used by ELLs to convey their word knowledge, and the types of words assessed. Due to the narrow descriptions of study samples and intervention procedures, the discussion of the findings are also limited and do not fully represent the dynamic ways in which words and concomitant concepts are developed for ELLs through listening, speaking, reading and writing in their two languages. After more evidence in vocabulary
instructional practices with ELLs is accumulated, future reviews on the topic could better inform the field on the role of specific characteristics on word-meaning acquisition with this vast and diverse population.

Although this study made attempts to synthesize all the high-quality research evidence available, drawing inferences of overall cause-and-effect relationships on general and curriculum-based vocabulary knowledge was challenging due to the variations existing across methodological characteristics across studies. When evaluating studies’ levels of evidence for effectiveness, it was necessary to apply less stringent requirements than those used by the What Works Clearinghouse (2014) when examining available statistical information. Rather than disqualifying treatment-comparison group contrasts that showed some baseline differences (i.e., pretest effect sizes ranging from 0 to 0.20), albeit trivial, from the final analysis, this study included these groups in order to make a critical assessment of the existing yet limited research in this area.

Also, due to much of the available data being incomparable and only found in a limited number of studies, more comprehensive statistical techniques, such as through a meta-analysis, were not undertaken to combine samples of individual studies, investigate possible sources of heterogeneity, and conduct moderator analyses, which may better inform the field on the most responsive learning approaches to the needs of ELLs who exhibit varying language and literacy needs. Given the limitations in this review, the broad yet in-depth analysis of the available evidence conducted in this synthesis can be useful in clarifying some of the issues with respect to vocabulary instruction with ELLs and identifying future directions for researchers.
Conclusion

This chapter presented a synthesis of the findings of the systematic review and a discussion of the results as they relate to prior and future research. These findings highlight the need for increased research in the area of vocabulary instruction with ELL populations. By revisiting existing practices showcased in this review, researchers could potentially add to our understandings relating to vocabulary practices with ELLs by substantiating the inferences drawn from original studies with future studies that meet higher levels of evidence standards. Additionally, this study points to specific issues in current research and provides recommendations for researchers seeking to address the significant knowledge gap in research that is centered on ELL vocabulary learning. Future research should take into account the multi-faceted influences in the design of a study, instructional features, diversity of ELL populations, and the multiple dimensions of outcome measures.
References


Pre-IPT-Oral. (1989). Culver City, CA; Ballard and Tighe—A division of Educational IDEAS, Inc.


Appendix A: Outcome Characteristics for Included Studies

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Name of measure</th>
<th>Outcome objective</th>
<th>Type of measure&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Type of ability&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Modality</th>
<th>Word tier assessed</th>
<th>Type of knowledge&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Atay &amp; Kurt (2006)</td>
<td>Vocabulary Knowledge Scale</td>
<td>Depth</td>
<td>CB</td>
<td>R &amp; E</td>
<td>Written</td>
<td>1</td>
<td>D &amp; C</td>
</tr>
<tr>
<td></td>
<td>Sentence Completion</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Target Mastery</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Word Association</td>
<td>Depth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Polysemy Production</td>
<td>Depth</td>
<td>CB</td>
<td>E</td>
<td>Written</td>
<td>2</td>
<td>D &amp; C</td>
</tr>
<tr>
<td>Test Description</td>
<td>Level</td>
<td>Format</td>
<td>Test Type</td>
<td>Test</td>
<td>Domain</td>
<td>Version</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>-------</td>
<td>--------</td>
<td>-----------</td>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>British Ability Scales II: Naming Vocabulary</td>
<td>Breadth</td>
<td>G</td>
<td>R</td>
<td>Oral</td>
<td>1</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>9. Lesaux et al. (2010) SAT-10 Reading Vocabulary</td>
<td>Breadth</td>
<td>G</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D &amp; C</td>
<td></td>
</tr>
<tr>
<td>Target Word Association</td>
<td>Depth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>Target Word Mastery</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Author(s) &amp; Year</td>
<td>Test Description</td>
<td>Breadth</td>
<td>Format</td>
<td>Depth</td>
<td>Format</td>
<td>Core &amp; Grade</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
<td>------------------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>10</td>
<td>Mancilla-Martinez (2010)</td>
<td>Multiple Choice (MC)</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Vocabulary Self Check (VSC)</td>
<td>Depth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>2</td>
<td>D &amp; C</td>
</tr>
<tr>
<td>11</td>
<td>Nelson et al. (2011)</td>
<td>Woodcock Reading Mastery Test-Revised/Normative Update</td>
<td>Breadth</td>
<td>G</td>
<td>R &amp; E</td>
<td>Oral &amp; Written</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>Schunk (1999)</td>
<td>Adapted from PPVT-R</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Oral</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Science Concepts Knowledge</td>
<td>Depth</td>
<td>CB</td>
<td>E</td>
<td>Oral</td>
<td>3</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Target Vocabulary Assessment (TVA)</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Oral</td>
<td>2</td>
<td>D</td>
</tr>
<tr>
<td>15</td>
<td>Spycher (2009)</td>
<td>Emergent Science Vocabulary</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Oral</td>
<td>2 &amp; 3</td>
</tr>
<tr>
<td></td>
<td>Author(s) and Year</td>
<td>Test/Procedure</td>
<td>Breadth Type</td>
<td>Domain</td>
<td>Grade Level</td>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>----------------</td>
<td>--------------</td>
<td>--------</td>
<td>-------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Woodcock Language Proficiency Battery-Revised Picture Vocabulary subtest</td>
<td>Breadth</td>
<td>G</td>
<td>E</td>
<td>Oral</td>
<td>1</td>
</tr>
<tr>
<td>17.</td>
<td>Ulanoff &amp; Pucci (1999)</td>
<td>Book vocabulary</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sentence Completion Task</td>
<td>Breadth</td>
<td>CB</td>
<td>R</td>
<td>Written</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* NR = Not reported

*a* CB = Curriculum-based vocabulary; G = General vocabulary

*b* R = Receptive; E = Expressive

*c* D = Definitional; C = Contextual
Appendix B: Figures Displaying the Proportion of Studies by Outcome Characteristics

Number of studies using general vocabulary measures assessing breadth of word knowledge
Appendix C: Curriculum-Based Measures

Number of studies using curriculum-based vocabulary measures assessing breadth of word knowledge
Number of studies using curriculum-based measures assessing depth of word knowledge
### Appendix D: Learner Characteristics for Included Values

<table>
<thead>
<tr>
<th>Study</th>
<th>Grade</th>
<th>Age</th>
<th>Language group (Language)</th>
<th>L2 proficiency</th>
<th>L1 proficiency</th>
<th>SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Atay &amp; Kurt (2006)</td>
<td>6\textsuperscript{th}</td>
<td>$M=11$</td>
<td>Homogeneous (Turkish)</td>
<td>Low</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td>2. Avila &amp; Sadoski (1996)</td>
<td>5\textsuperscript{th}</td>
<td>$M=11.2$ (SD=0.97)</td>
<td>Homogeneous (Spanish)</td>
<td>Low</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td>3. Carlo et al. (2004)</td>
<td>5\textsuperscript{th}</td>
<td>NR</td>
<td>Heterogeneous (Spanish-speaking ELLs and native English speakers)</td>
<td>Low</td>
<td>NR</td>
<td>Low &amp; Medium</td>
</tr>
<tr>
<td>4. Collins (2004)</td>
<td>Pre-K</td>
<td>Range: 4 to 5;2</td>
<td>Homogeneous (Portuguese)</td>
<td>Low</td>
<td>NR</td>
<td>Middle to Low</td>
</tr>
<tr>
<td>Study</td>
<td>Grade</td>
<td>Language</td>
<td>Education Level</td>
<td>Low</td>
<td>NR</td>
<td>Medium</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>------</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td>Dockrell et al. (2010)</td>
<td>Pre-K</td>
<td>Range: 3.5-3.6</td>
<td>Heterogeneous</td>
<td>Low</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Bengali,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Syletti,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Turkish,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amharic,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Somali)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farver et al. (2009)</td>
<td>Pre-K</td>
<td>54.51 months, SD=4.72</td>
<td>Homogeneous</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filippini (2007)</td>
<td>1st</td>
<td>NR</td>
<td>Homogeneous</td>
<td>Low</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesaux et al. (2010)</td>
<td>6th</td>
<td>M=11;11 (includes native English speakers)</td>
<td>Heterogeneous (native English speakers; Spanish, Vietnamese, Hmong, Somali, Pilipino/Tagalog)</td>
<td>Low</td>
<td>NR</td>
<td>Low &amp; Medium</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td>Grade</td>
<td>Type</td>
<td>Population</td>
<td>Method</td>
<td>Control</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------</td>
<td>-------</td>
<td>------</td>
<td>------------------------------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>10</td>
<td>Mancilla-Martinez (2010)</td>
<td>5th</td>
<td>NR</td>
<td>Heterogeneous (Spanish-speaking ELLs &amp; native English speakers)</td>
<td>Low &amp; High</td>
<td>NR</td>
</tr>
<tr>
<td>11</td>
<td>Nelson et al. (2011)</td>
<td>K</td>
<td>NR</td>
<td>Homogeneous (Spanish)</td>
<td>Low</td>
<td>NR</td>
</tr>
<tr>
<td>12</td>
<td>Roberts &amp; Neal (2003)</td>
<td>Pre-K</td>
<td>M=52.8 months (SD=4.74)</td>
<td>Heterogeneous (Hmong &amp; Spanish)</td>
<td>Low</td>
<td>NR</td>
</tr>
<tr>
<td>13</td>
<td>Schunk (1999)</td>
<td>K to 2nd</td>
<td>NR</td>
<td>Heterogeneous (Spanish, Hmong &amp; Lao)</td>
<td>Low</td>
<td>NR</td>
</tr>
<tr>
<td>14</td>
<td>Silverman &amp; Hines (2009)</td>
<td>Pre-K to 2nd</td>
<td>M=6.5; Range: 4.5-8.5</td>
<td>Heterogeneous (Haitian Creole, Portuguese, Mandarin &amp; Spanish)</td>
<td>Low</td>
<td>NR</td>
</tr>
<tr>
<td>15</td>
<td>Spycher (2009)</td>
<td>K</td>
<td>NR</td>
<td>Homogeneous (Spanish)</td>
<td>Low</td>
<td>NR</td>
</tr>
<tr>
<td>16</td>
<td>Uchikoshi</td>
<td>K</td>
<td>M=5.6</td>
<td>Homogeneous (Spanish)</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>NR</td>
<td>Type</td>
<td>Level</td>
<td>NR</td>
<td>Level</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------</td>
<td>----</td>
<td>---------------------------</td>
<td>-------------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>Ulanoff &amp; Puchi (1999)</td>
<td>2006</td>
<td>NR</td>
<td>Homogeneous (Spanish)</td>
<td>Low</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td>Zhang &amp; Schumm (2000)</td>
<td>2000</td>
<td>NR</td>
<td>Homogeneous (Spanish)</td>
<td>Low, Medium &amp; High</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

*Note.* NR = Not reported
Appendix E: Proportions of Studies for Each Coding Category Relating to Sample Characteristics

Proportion of studies coded by differing grade levels: Early childhood, Elementary and Multi-grade

Proportion of studies coded by differing language groups: Homogeneous Spanish-speaking ELLs, Other homogeneous ELLs and Heterogeneous samples
Proportion of studies coded by differing levels of L2 proficiency: Homogeneous low proficiency groups and Heterogeneous ability groups

Proportion of studies coded by differing levels of L1 proficiency: Homogeneous low proficiency groups and Unreported
Proportion of studies coded by differing levels of SES: Unreported, Low and Low & Medium
<table>
<thead>
<tr>
<th>Study</th>
<th>Focusa</th>
<th>Targeted words (by tier)</th>
<th>Training received</th>
<th>Duration</th>
<th>Frequency</th>
<th>Intensity</th>
<th>Total Instructional time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Atay &amp; Kurt (2006)</td>
<td>V+</td>
<td>Tier 1</td>
<td>No</td>
<td>6 weeks</td>
<td>2 days</td>
<td>120 minutes</td>
<td>1440 minutes</td>
</tr>
<tr>
<td>2. Avila &amp; Sadoski (1996)</td>
<td>V+</td>
<td>Tier 1</td>
<td>Yes</td>
<td>1 week</td>
<td>3 days</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>3. Carlo et al. (2004)</td>
<td>V+</td>
<td>Tier 2</td>
<td>Yes</td>
<td>15 weeks</td>
<td>4 days</td>
<td>30-45 minutes</td>
<td>1286 minutes</td>
</tr>
<tr>
<td>4. Collins (2004)</td>
<td>V+</td>
<td>Tier 2 &amp; 3</td>
<td>Yes</td>
<td>3 weeks</td>
<td>3 days</td>
<td>20 minutes</td>
<td>180 minutes</td>
</tr>
<tr>
<td>5. Dockrell et al. (2010)</td>
<td>V</td>
<td>Tier 1</td>
<td>Yes</td>
<td>15 weeks</td>
<td>3 days</td>
<td>20 minutes</td>
<td>450 minutes</td>
</tr>
<tr>
<td>6. Farver et al. (2009)</td>
<td>V</td>
<td>Tier 1</td>
<td>Yes</td>
<td>21 weeks</td>
<td>2 days</td>
<td>15 minutes</td>
<td>1680 minutes</td>
</tr>
<tr>
<td>7. Filippini (2007)</td>
<td>V</td>
<td>Tier 1 &amp; 2</td>
<td>Yes</td>
<td>8 weeks</td>
<td>4 days</td>
<td>15 minutes</td>
<td>435 minutes</td>
</tr>
<tr>
<td>9. Lesaux et al. (2010)</td>
<td>V+</td>
<td>Tier 2</td>
<td>Yes</td>
<td>18 weeks</td>
<td>4 days</td>
<td>45 minutes</td>
<td>3240 minutes</td>
</tr>
<tr>
<td>Study</td>
<td>Type</td>
<td>Tier</td>
<td>Effort</td>
<td>Duration</td>
<td>Frequency</td>
<td>Time</td>
<td>Total Duration</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>Mancilla-Martinez (2010)</td>
<td>V+</td>
<td>Tier 2</td>
<td>Yes</td>
<td>20 weeks</td>
<td>5 days</td>
<td>15 minutes</td>
<td>1500 minutes</td>
</tr>
<tr>
<td>Nelson et al. (2011)</td>
<td>V</td>
<td>Tier 1</td>
<td>Yes</td>
<td>20 weeks</td>
<td>5 days</td>
<td>20 minutes</td>
<td>2000 minutes</td>
</tr>
<tr>
<td>Roberts &amp; Neal (2003)</td>
<td>V</td>
<td>Tier 1</td>
<td>NR</td>
<td>16 weeks</td>
<td>3 days</td>
<td>20-25 minutes</td>
<td>1080 minutes</td>
</tr>
<tr>
<td>Schunk (1999)</td>
<td>V+</td>
<td>Tier 1</td>
<td>NR</td>
<td>1-2 weeks</td>
<td>3 days</td>
<td>15 minutes</td>
<td>68 minutes</td>
</tr>
<tr>
<td>Silverman &amp; Hines (2009)</td>
<td>V+</td>
<td>Tier 2</td>
<td>Yes</td>
<td>12 weeks</td>
<td>3 days</td>
<td>45 minutes</td>
<td>1620 minutes</td>
</tr>
<tr>
<td>Schunk (1999)</td>
<td>V+</td>
<td>Tier 2 &amp; 3</td>
<td>Yes</td>
<td>5 weeks</td>
<td>NR</td>
<td>20-25 minutes</td>
<td>NR</td>
</tr>
<tr>
<td>Uchikoshi (2006)</td>
<td>V+</td>
<td>Tier 1</td>
<td>NR</td>
<td>NR</td>
<td>3 days</td>
<td>30 minutes</td>
<td>1620 minutes</td>
</tr>
<tr>
<td>Ulanoff &amp; Puchi (1999)</td>
<td>V+</td>
<td>Tier 1</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>90 minutes</td>
</tr>
<tr>
<td>Zhang &amp; Schumm (2000)</td>
<td>V+</td>
<td>Tier 1</td>
<td>Yes</td>
<td>1 days</td>
<td>2 days</td>
<td>60 minutes on first day; 30 minutes on second day</td>
<td>90 minutes</td>
</tr>
</tbody>
</table>

Note. NR= Not reported.
*V+= Vocabulary-focused studies; V= Vocabulary related-studies
Appendix G: Treatment Characteristics for Included Studies at the Condition Level.

<table>
<thead>
<tr>
<th>Study / Study conditions</th>
<th>Explicitness of instruction</th>
<th>Contextual / definitional support(^a)</th>
<th>Group size</th>
<th>Exposure(^b)</th>
<th>L1 Support(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) post-reading tasks focused on interaction</td>
<td>Both Explicit &amp; Implicit</td>
<td>4</td>
<td>Small</td>
<td>ME</td>
<td>2</td>
</tr>
<tr>
<td>b) post-reading tasks with independent work</td>
<td>Explicit</td>
<td>2</td>
<td>Small</td>
<td>ME</td>
<td>2</td>
</tr>
<tr>
<td>a) Keyword method</td>
<td>Explicit</td>
<td>1</td>
<td>Whole</td>
<td>MR</td>
<td>2</td>
</tr>
<tr>
<td>b) Rehearsal method</td>
<td>Explicit</td>
<td>1</td>
<td>Whole</td>
<td>MR</td>
<td>2</td>
</tr>
<tr>
<td>a) VIP curriculum</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Mixed</td>
<td>ME</td>
<td>2</td>
</tr>
<tr>
<td>b) Treatment as usual</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
<td>0</td>
</tr>
<tr>
<td>a) Rich explanations</td>
<td>Explicit</td>
<td>4</td>
<td>Small</td>
<td>ME</td>
<td>0</td>
</tr>
<tr>
<td>b) No explanations</td>
<td>Incidental</td>
<td>5</td>
<td>Small</td>
<td>SE</td>
<td>0</td>
</tr>
<tr>
<td>Study</td>
<td>Intervention</td>
<td>Type</td>
<td>Sample Size</td>
<td>Method</td>
<td>Experimental Group(s)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>--------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Dockrell et al. (2010)</td>
<td>a) Talking Time</td>
<td>Implicit</td>
<td>4</td>
<td>Whole</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>b) Story Reading</td>
<td>Explicit</td>
<td>4</td>
<td>Whole</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>c) Treatment as usual</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Farver et al. (2009)</td>
<td>a) Spanish-to-English small-group curriculum</td>
<td>Implicit</td>
<td>4</td>
<td>Small</td>
<td>MR</td>
</tr>
<tr>
<td></td>
<td>b) English-only small-group curriculum</td>
<td>English</td>
<td>4</td>
<td>Small</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>c) English whole-group curriculum</td>
<td>Not Reported</td>
<td>0</td>
<td>Whole</td>
<td>NR</td>
</tr>
<tr>
<td>Filippini (2007)</td>
<td>a) Phonological and decoding</td>
<td>Explicit</td>
<td>5</td>
<td>Small</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>b) Phonological and decoding + semantic relations</td>
<td>Explicit</td>
<td>2</td>
<td>Small</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td>c) Phonological and decoding + morphological instruction</td>
<td>Incidental</td>
<td>2</td>
<td>Small</td>
<td>ME</td>
</tr>
<tr>
<td>Study</td>
<td>Intervention Area</td>
<td>Method</td>
<td>Effect Size</td>
<td>Level of Evidence</td>
<td>Strength</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>b) Phonological awareness</td>
<td>Not reported</td>
<td>0</td>
<td>Small</td>
<td>NR</td>
</tr>
<tr>
<td>9. Lesaux et al. (2010)</td>
<td>a) ALIAS curriculum</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Mixed</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td>b) Treatment as usual</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>b) Treatment as usual</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>11. Nelson et al. (2011)</td>
<td>a) Supplemental Root Word Intervention</td>
<td>Explicit</td>
<td>2</td>
<td>Small</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td>b) Story Reading Activities</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Small</td>
<td>ME</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Decoding-related instruction</td>
<td>Not Reported</td>
<td>5</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Singing paired with signs</td>
<td>Explicit</td>
<td>4</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Speaking paired with signs</td>
<td>Explicit</td>
<td>4</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Singing only</td>
<td>Incidental</td>
<td>5</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Speaking only</td>
<td>Incidental</td>
<td>5</td>
<td>Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Multimedia</td>
<td>Explicit</td>
<td>3</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Nonmultimedia</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Spycher (2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Intentional approach</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Implicit approach</td>
<td>Incidental</td>
<td>5</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) <em>Arthur</em></td>
<td>Incidental</td>
<td>5</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) <em>Between the Lions</em></td>
<td>Explicit</td>
<td>4</td>
<td>Whole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) No exposure</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Singing paired with signs</td>
<td>Explicit</td>
<td>4</td>
<td>Small</td>
</tr>
<tr>
<td>b) Speaking paired with signs</td>
<td>Explicit</td>
<td>4</td>
<td>Small</td>
</tr>
<tr>
<td>c) Singing only</td>
<td>Incidental</td>
<td>5</td>
<td>Small</td>
</tr>
<tr>
<td>d) Speaking only</td>
<td>Incidental</td>
<td>5</td>
<td>Small</td>
</tr>
<tr>
<td>a) Multimedia</td>
<td>Explicit</td>
<td>3</td>
<td>Whole</td>
</tr>
<tr>
<td>b) Nonmultimedia</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Whole</td>
</tr>
<tr>
<td>15. Spycher (2009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Intentional approach</td>
<td>Both Explicit &amp; Implicit</td>
<td>3</td>
<td>Whole</td>
</tr>
<tr>
<td>b) Implicit approach</td>
<td>Incidental</td>
<td>5</td>
<td>Whole</td>
</tr>
<tr>
<td>a) <em>Arthur</em></td>
<td>Incidental</td>
<td>5</td>
<td>Whole</td>
</tr>
<tr>
<td>b) <em>Between the Lions</em></td>
<td>Explicit</td>
<td>4</td>
<td>Whole</td>
</tr>
<tr>
<td>c) No exposure</td>
<td>Not Reported</td>
<td>0</td>
<td>NR</td>
</tr>
</tbody>
</table>
Implicit | 4 | Whole | ME | 2 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Preview-Review</td>
<td>Explicit</td>
<td>4</td>
<td>Whole</td>
<td>SE</td>
<td>1</td>
</tr>
<tr>
<td>b) Concurrent Translation</td>
<td>Explicit</td>
<td>4</td>
<td>Whole</td>
<td>SE</td>
<td>1</td>
</tr>
<tr>
<td>c) Implicit Exposure</td>
<td>Incidental</td>
<td>5</td>
<td>Whole</td>
<td>NR</td>
<td>0</td>
</tr>
<tr>
<td>a) Spanish keyword method</td>
<td>Explicit</td>
<td>1</td>
<td>Whole</td>
<td>MR</td>
<td>0</td>
</tr>
<tr>
<td>b) English keyword method</td>
<td>Explicit</td>
<td>1</td>
<td>Whole</td>
<td>MR</td>
<td>0</td>
</tr>
<tr>
<td>c) Rehearsal method</td>
<td>Explicit</td>
<td>1</td>
<td>Whole</td>
<td>MR</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) 0= No instruction or Not reported; 1= Definition-only; 2= Definitional emphasis; 3= Balanced; 4= Contextual Emphasis; 5= Context-only

\(^b\) SE= Single Exposure; MR: Multiple Reinforcement; ME: Multiple Exposure

\(^c\) 0= No support; 1= Minimal support; 2= Significant support

NR= Not reported
Appendix H: Proportion of Studies Meeting Evidence Standards by Level
Appendix I: Effect Sizes for Studies Meeting Evidence Standards Using General Vocabulary Measures to Assess Definitional Knowledge

<table>
<thead>
<tr>
<th>Study / Treatment-Comparison Group Contrast</th>
<th>Measure</th>
<th>Type of ability&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Word tier assessed</th>
<th>Statistical Significance</th>
<th>95% Confidence interval&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Posttest effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlo et al. (2004): ALIAS curriculum vs. Treatment as usual condition</td>
<td>PPVT-R</td>
<td>R</td>
<td>1 &amp; 2</td>
<td>No</td>
<td>-0.46</td>
<td>0.24</td>
</tr>
<tr>
<td>Farver et al. (2009): English curriculum vs. whole group</td>
<td>P-CTOPPP</td>
<td>R</td>
<td>1</td>
<td>Yes</td>
<td>-0.10</td>
<td>0.90</td>
</tr>
<tr>
<td>Farver et al. (2009): Transition curriculum vs. whole group</td>
<td>P-CTOPPP</td>
<td>R</td>
<td>1</td>
<td>Yes</td>
<td>0.20</td>
<td>1.22</td>
</tr>
<tr>
<td>Farver et al. (2009): Transition curriculum vs. whole group</td>
<td>P-CTOPPP</td>
<td>R</td>
<td>1</td>
<td>Yes</td>
<td>-0.27</td>
<td>0.73</td>
</tr>
</tbody>
</table>

<sup>a</sup> Type of ability: R = Reading

<sup>b</sup> 95% Confidence interval: LL = Lower Limit, UL = Upper Limit
<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Measure</th>
<th>Age</th>
<th>Use</th>
<th>Effect Size</th>
<th>p-Value 1</th>
<th>p-Value 2</th>
<th>p-Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farver et al. (2009): English curriculum vs. whole group</td>
<td>P-CTOPPP</td>
<td>E</td>
<td>1</td>
<td>Yes</td>
<td>-0.90</td>
<td>0.91</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Farver et al. (2009): Transition curriculum vs. whole group</td>
<td>P-CTOPPP</td>
<td>E</td>
<td>1</td>
<td>Yes</td>
<td>0.24</td>
<td>1.26</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Farver et al. (2009): Transition curriculum vs. English mode</td>
<td>P-CTOPPP</td>
<td>E</td>
<td>1</td>
<td>Yes</td>
<td>-0.11</td>
<td>0.89</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Silverman &amp; Hines (2009): Multimedia vs. Nonmultimedia</td>
<td>PPVT-III</td>
<td>R</td>
<td>1 &amp; 2</td>
<td>Yes</td>
<td>0.23</td>
<td>0.75</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Uchikoshi (2006): Arthur vs. No viewing</td>
<td>PPVT-III</td>
<td>R</td>
<td>1 &amp; 2</td>
<td>No</td>
<td>-0.35</td>
<td>0.47</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

1 & 2 No -0.50 0.30 0.10
<table>
<thead>
<tr>
<th>Study</th>
<th>Test</th>
<th>Type</th>
<th>Condition</th>
<th>Language Proficiency Battery-Revised- Picture Vocabulary subtest</th>
<th>Effect Size</th>
<th>LL</th>
<th>UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uchikoshi (2006): Between the Lions vs. No viewing</td>
<td>PPVT-III</td>
<td>R</td>
<td></td>
<td></td>
<td>-0.22</td>
<td>0.54</td>
<td>0.16</td>
</tr>
<tr>
<td>Uchikoshi (2007): Arthur vs. Between the Lions</td>
<td>PPVT-III</td>
<td>R</td>
<td>1 &amp; 2</td>
<td>No</td>
<td>-0.22</td>
<td>0.54</td>
<td>0.16</td>
</tr>
<tr>
<td>Uchikoshi (2006): Between the Lions vs. No viewing</td>
<td>Woodcock</td>
<td>E</td>
<td>1 &amp; 2</td>
<td>No</td>
<td>-0.36</td>
<td>0.44</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>Language Proficiency Battery-Revised- Picture Vocabulary subtest</td>
<td></td>
<td></td>
<td></td>
<td>-0.28</td>
<td>0.44</td>
<td>0.13</td>
</tr>
</tbody>
</table>

*a R: Receptive; E: Expressive.

b LL: Lower limit; UL: Upper limit.
Appendix J: Effect Sizes for Studies Meeting Evidence Standards Using Curriculum-Based Measures to Assess Breadth of Word Knowledge

<table>
<thead>
<tr>
<th>Study / Treatment-comparison group contrast</th>
<th>Measure</th>
<th>Type of knowledge(^a)</th>
<th>Type of ability(^b)</th>
<th>Word tier assessed</th>
<th>Statistical significance</th>
<th>95% Confidence interval</th>
<th>Posttest effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlo et al. (2004): VIP vs. Treatment as usual</td>
<td>Word Mastery</td>
<td>D</td>
<td>R</td>
<td>2</td>
<td>Yes</td>
<td>-0.04</td>
<td>2.10</td>
</tr>
<tr>
<td>Collins (2004): Rich explanations vs. No explanations</td>
<td>Target Vocabulary Test (TVT)</td>
<td>D</td>
<td>R</td>
<td>1</td>
<td>Yes</td>
<td>0.89</td>
<td>1.89</td>
</tr>
<tr>
<td>Filippini (2007): PAD+SR vs. PAD</td>
<td>Target Vocabulary Knowledge (VOC)</td>
<td>D</td>
<td>R</td>
<td>2</td>
<td>No</td>
<td>-0.08</td>
<td>1.16</td>
</tr>
<tr>
<td>Filippini (2007): PAD+MA vs. PAD</td>
<td>Target Vocabulary Knowledge (VOC)</td>
<td>D</td>
<td>R</td>
<td>2</td>
<td>No</td>
<td>-0.06</td>
<td>1.34</td>
</tr>
<tr>
<td>Lesaux et al. (2010): ALIAS curriculum vs.</td>
<td>Target Word Mastery</td>
<td>D</td>
<td>R</td>
<td>2</td>
<td>Yes</td>
<td>-0.31</td>
<td>1.09</td>
</tr>
<tr>
<td>Study</td>
<td>Treatment</td>
<td>Measure Description</td>
<td>D</td>
<td>R &amp; E</td>
<td>Yes/No</td>
<td>Low Limit</td>
<td>High Limit</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------------</td>
<td>---</td>
<td>-------</td>
<td>--------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>Nelson et al. (2011):</td>
<td>Supplemental root word intervention vs. Story reading activities</td>
<td>Definitional vocabulary curriculum-based measure</td>
<td>D</td>
<td>R &amp; E</td>
<td>1</td>
<td>Yes</td>
<td>0.73</td>
</tr>
<tr>
<td>Mancilla-Martinez (2010): Word Generation vs. Treatment as usual</td>
<td>Multiple Choice Test</td>
<td></td>
<td>C</td>
<td>R</td>
<td>2</td>
<td>Yes</td>
<td>0.33</td>
</tr>
<tr>
<td>Spycher (2009):</td>
<td>Emergent Science Vocabulary Assessment (ESVA) adapted from PPVT-III</td>
<td>D</td>
<td>R</td>
<td>2, 3</td>
<td>No</td>
<td>0.27</td>
<td>2.23</td>
</tr>
<tr>
<td>Ulanoff &amp; Pucci (1999): Preview-Review vs. No explanation</td>
<td>Book vocabulary</td>
<td>D</td>
<td>R</td>
<td>1</td>
<td>No</td>
<td>0.07</td>
<td>1.39</td>
</tr>
</tbody>
</table>

a D= Definitional; C= Contextual
b R= Receptive; E= Expressive
c LL= Lower limit; UL= Upper limit
## Appendix K: Effect Sizes for Studies Meeting Evidence Standards Using Curriculum-Based Measures to Assess Depth of Word Knowledge

<table>
<thead>
<tr>
<th>Study: Treatment-comparison group contrast</th>
<th>Measure</th>
<th>Type of knowledge&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Type of ability&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Word tier assessed</th>
<th>Statistical significance</th>
<th>95% Confidence interval</th>
<th>Posttest effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atay &amp; Kurt (2006): Post-reading with interaction vs. Post-reading with independent activities</td>
<td>Vocabulary Knowledge Scale</td>
<td>D &amp; C</td>
<td>R &amp; E</td>
<td>1</td>
<td>Yes</td>
<td>-0.30</td>
<td>0.92</td>
</tr>
<tr>
<td>Carlo et al. (2004): VIP curriculum vs. Treatment as usual</td>
<td>Word Association</td>
<td>D</td>
<td>R</td>
<td>2</td>
<td>No</td>
<td>0.09</td>
<td>0.79</td>
</tr>
<tr>
<td>Carlo et al. (2004): VIP vs. Treatment as usual</td>
<td>Polysemy production</td>
<td>C</td>
<td>E</td>
<td>2</td>
<td>No</td>
<td>0.24</td>
<td>0.94</td>
</tr>
<tr>
<td>Lesaux et al. (2011): ALIAS</td>
<td>Word Association</td>
<td>D</td>
<td>E</td>
<td>2</td>
<td>No</td>
<td>-0.04</td>
<td>0.34</td>
</tr>
<tr>
<td>Study</td>
<td>Domain</td>
<td>Treatment</td>
<td>N</td>
<td>p</td>
<td>95% CI</td>
<td>Estimation</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------</td>
<td>------------</td>
<td>---</td>
<td>------</td>
<td>--------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>Mancilla-Martinez (2010): Word Generation</td>
<td>Vocabulary Self Check</td>
<td>D &amp; C</td>
<td>R</td>
<td>2</td>
<td>No</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Silverman &amp; Hines (2009): Multimedia vs.</td>
<td>Science Concepts Knowledge</td>
<td>D</td>
<td>E</td>
<td>3</td>
<td>No</td>
<td>-0.61</td>
<td></td>
</tr>
<tr>
<td>Nonmultimedia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.15</td>
<td></td>
</tr>
</tbody>
</table>

*a D= Definitional; C= Contextual*

*b R= Receptive; E= Expressive*

*c LL= Lower limit; UL= Upper limit*