Edinburgh Research Explorer

Assessing English language learners' collocation knowledge

Assessing English Language Learners' Collocation Knowledge: A Systematic Review

of Receptive and Productive Measurements

Abstract

Since collocation knowledge is integral to second language vocabulary depth, it necessitates a careful examination of various measurement approaches. To this end, the current paper provides an overview and evaluation of extant collocation measurements used in empirical studies on L2 English ($N = 154$) published between 1980 and 2023 indexed in the SSCI, SCIE, AHCI, SCOPUS, and ERIC databases. Six instruments, seven item formats, and three other assessment tools were identified and reviewed for the assessment of receptive and productive collocation knowledge. The review focused on the collocation knowledge measured by each tool, the instrument and/or item format employed, item design, reported reliability, and potential drawbacks of employing each instrument and item format in research or practice. The review proposes several theoretical and practical considerations for future assessments of and research on English collocation knowledge.

Keywords: collocation; receptive and productive knowledge; collocation measurement; instruments; item formats

1 Introduction

In recent decades, research has seen a surge of interest in multiword units — also called *formulaic sequences* (Schmitt 2010; Wray 2000) — due to their pivotal role in second or foreign language acquisition, and accounting for a considerable proportion of the English language (Altenberg 1998; Erman and Warren, 2000; Vilkaité, 2016). These

units confer processing advantages as prefabricated chunks or being highly automatized (e.g., Myles and Cordier 2017; Wray 2002). Nonetheless, their acquisition remains a formidable challenge, even for advanced language learners (Boers et al. 2014; Boone et al. 2022; Nesselhauf, 2003). Despite the growing literature on this topic, there still appears to be no consensus on the definition of formulaic sequences (Brown 2018; Gablasova et al. 2017; Siyanova-Chanturia and Omidian 2020; Wood 2020). This is largely due to the intricacy of formulaicity, conflicting terminology, interpretation of how those units are processed semantically or psycholinguistically, and the diverse, often conflicting or underdeveloped operationalization and classification conventions applied in distinct but related fields (Nesselhauf 2003; Wray 2012). As a result, researchers from the disciplines of phraseology, lexicography, second language acquisition (SLA) and pedagogy, corpus linguistics, and psycholinguistics define this construct differently (see Siyanova-Chanturia and Omidian 2020), with definitions loosely falling in line with three major approaches: the frequency-based approach, the phraseological approach, and the psycholinguistic approach (e.g., Gablasova et al. 2017; Siyanova-Chanturia & Omidian 2020; Uchihara et al. 2021). In the frequency-based approach, the habitual juxtaposition of words, mostly forming meaningful lexical units, is deemed as a fundamental feature for defining multiword units (e.g., Biber et al. 2002; Lei and Liu 2018; Nesselhauf 2003; Sinclair 1991). The phraseological approach uses a continuum of idiomaticity. At one end of the continuum lie idioms whose constituent words are restricted and can only appear in a particular order; at the other end lie free or non-compositional word combinations without any restrictions (Baldwin and Kim

2

2010; Cowie 1994; Howarth 1998). Third, the psycholinguistic approach is based on the assumption that formulaic sequences refer to linguistic units that are stored and processed in the mental lexicon as wholes (Pawley and Syder 1983; Sinclair 1991; Wray 2002).

As part of the broad family of formulaic sequences, collocations have been defined in various ways according to the approach used. A review of the recent literature shows the tendency for collocation studies to have applied a hybrid approach by combining different criteria from mainly the frequency-based and phraseological approaches to select pedagogically meaningful collocations as learning or research targets (Fazlali and Shahini 2019; Kim 2017; Snoder 2017). For the current purposes, we will adopt this hybrid approach and consider collocations as habitually co-occurring lexical partnerships that have relatively transparent meanings (e.g., *make an effort, heavy rain, pay attention*), contrary to idioms (e.g., *make a long story short)*, for example.

Collocation knowledge is an important part of the depth of vocabulary knowledge (i.e., how much one knows about words), which complements the breadth of vocabulary knowledge (i.e., how many words one knows) (Schmitt 2014). To define the concept of collocation knowledge, we will use the most established and accepted vocabulary knowledge framework by Nation (2022). Nation (2022) has postulated that knowing a word includes knowledge of its form, meaning, and use (see Figure 1). Knowledge of word *use* includes the grammatical functions, collocations, and constraints on use of a word. Focusing on collocations, a distinction between receptive and productive collocation knowledge can be drawn (see Figure 1). Receptive collocation knowledge

refs to the ability to recognize collocations when they are encountered in listening or reading, while productive collocation knowledge is the ability to produce collocations accurately in speaking and writing (Nation 2022).
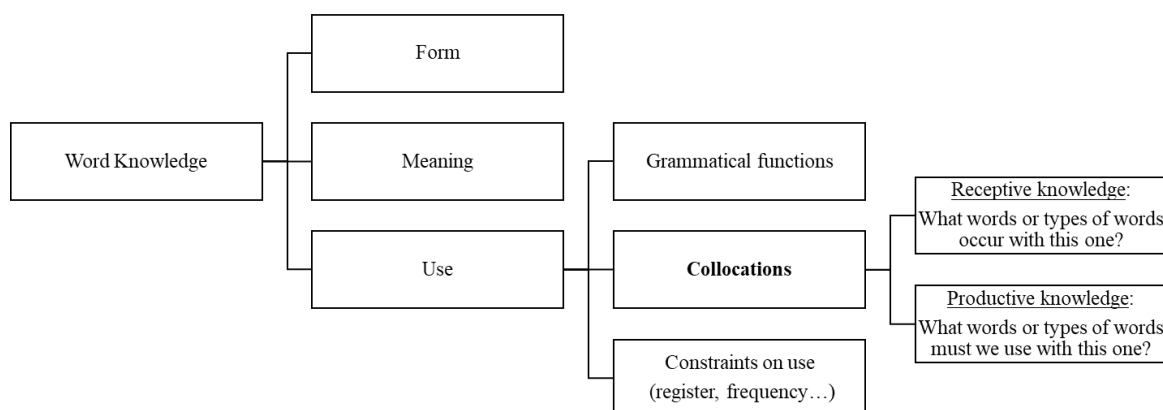
Figure 1:

Word knowledge framework (Nation 2022).

Until today, there have been different approaches to measuring collocation knowledge. Interestingly, it seems that the formats to assess collocation knowledge have largely consisted of simple adaptations of the formats used to assess individual word knowledge. The reasoning behind it might be that collocations are prone to being viewed as one of many word knowledge aspects rather than an independent construct (Nation 2022). For example, often used formats for assessing both single-word and collocation knowledge are *meaning recognition* and *form recognition* (for receptive knowledge) and *meaning recall* and *form recall* (for productive knowledge) (e.g.,

González-Fernández and Schmitt 2020; Lee and Shin 2021; Peters 2016). For the assessment of learners' receptive collocation knowledge, multiple-choice, word combination, matching, and Yes/No judgment tasks are commonly used. When it comes to testing learners' productive collocation knowledge, fill-in-the-blank or translation tasks are common ways of testing (Lee and Shin 2021).

Because of the diversity in tools to assess collocation knowledge, the current study aims to systematically review the available collocation knowledge measurements. In doing so, we aim to provide language instructors and researchers with an overview of these extant tools. Only when stakeholders have been provided with validated and reliable measurements with which to assess collocation knowledge can their curriculum planning or research project properly incorporate the assessment of collocation knowledge.

2 Method

2.1 Key concepts: collocation as a research construct, instrument, and item format

Prior to the description of the systematic review procedures, the concepts of *collocation as a research construct, instrument* and *item format* need to be established. Considering the complexity of collocations and the difficulty of producing a definition, we focused our analyses. In our investigation, *collocation* was defined in a hybrid approach which considered the following key aspects: 1) two- to three- word combinations constituting lexical (i.e., composed of content words) or grammatical

collocations (i.e., composed of a content word and a preposition or grammatical structure)[1] (Benson et al. 1986); 2) free-word combinations and collocations based on the phraseological classification framework (Howarth 1998) and 3) frequent word combinations for which the constituent words have a high tendency to co-occur. This operationalization allowed us to narrow our focus while also including a relatively comprehensive range of collocation instruments and item formats.

An *instrument* was a specially designed tool that could measure a construct through one or more items and a type or subcomponent (i.e., different node-collocate relationship, frequency level, or level of semantic transparency) of the collocation construct.

The term *item format* referred to how questions/items or tasks appeared within an instrument. For example, multiple-choice items, although having many renditions, have the following basic structure: stem, several distractors, and a key.

2.2 Phases of the systematic review

Following established practices of executing a systematic review in the field of applied linguistics (see Gough et al. 2012; Plonsky and Kim 2016; Plonsky and Oswald 2015), the present study included several phases to clarify the inclusion criteria, study selection, coding scheme, and analysis of synthetic data.

---

[1]In addition to the frequency-based and phraseological approaches, the syntactic relation among words within a collocation is also considered as a significant criterion. Despite its significance, this aspect has often been overlooked in L2 collocation study (Xia, et al. 2022).

2.2.1 Inclusion criteria

The following inclusion criteria were used to locate collocation measurements for review: (a) the source article reported either empirical studies or literature reviews; (b) the main focus was measuring collocation knowledge; (c) the collocation knowledge measurement was described in enough detail to allow for review; (d) the study or review was related to the field of L2 English teaching and learning; and (e) the article was written in English.

By focusing on the field of L2 English teaching and learning, this systematic review ensures a comprehensive and accessible pool of literature for analysis. In addition, the paper's contents are relevant and applicable to individuals involved in teaching and researching L2 English. Although we acknowledge that collocation research is not limited to L2 English, limiting the present review to this language helps ensure a manageable scope for the article, making it feasible to conduct a thorough and rigorous review within reasonable resource constraints. This approach also allows for a better comparison of studies and thereby enhances the systematic review's coherence and relevance for a specific audience (i.e., English as a second/foreign language researchers/teachers).

2.2.2 Study selection

To identify potential studies to be included in the review, we queried the databases of Educational Resources Information Center (ERIC), SCOPUS, and the Web of Science (WOS) (limited to SSCI, SCIE, and AHCI indexes). As searching across the

entire WOS database can be time-consuming and yield an overwhelming number of studies, we focused on specific databases, allowing us to retrieve more relevant and targeted papers. Each of the selected databases is widely recognized for its coverage of educational research, linguistic studies, and interdisciplinary content, which aligns closely with our review subject. Specifically, we aimed our searches at WOS databases that contained journal articles and not those that indexed conference papers and books.[2] The literature search covered studies published from January 1, 1980 to March 11, 2023. Research into L2 collocation acquisition can be dated back to the 1980s (e.g., Ellis 1983), when most relevant terminologies and approaches emerged (e.g., Cowie 1994; Howarth 1998; Sinclair 1991; Wray 2002). We intended to cover a comprehensive, if not exhaustive, sample of collocation knowledge measurements. The key search terms used were (collocation* OR phras* OR multiword* OR formulaic* OR n-gram*[3]) AND (measur* OR instrument* OR test* OR assess* OR examin* OR acqui* OR track* OR evaluat* OR develop* OR us*)[4].

To narrow down the search results, a series of strategies were deployed. First, document types were limited to articles and reviews. Second, English was selected as the publication language. Third, the subject areas of the databases were defined.

---

[2]Peer-reviewed journal articles undergo expert evaluation, enhancing credibility. Consistency in reporting aids data extraction, whereas non-standardized reporting in conference papers and books can hinder systematic review processes. Availability of comprehensive information and robust peer review in journals contribute to their preferred inclusion.

[3]The search terms encompassed n-grams due to their inherent inclusion of two-word combinations that align with our defined concept of collocation. Multiword units and formulaic sequences serve as overarching terms that inherently encompass collocation. However, in the case of lexical bundles and chunks, they deviate from our precise definition of collocations. These terms can encompass very lengthy word combinations, often forming relatively complete structures. Encompassing these terms within our study parameters would results in an influx of literature and necessitate substantial human filtering efforts, yet yield a limited quantity of highly relevant literature that aligns with our criteria.

[4]us* was replaced with use* for Web of Science searches as three characters are required before wild card searches.

Specifically, in searching WOS, the subject was limited to all related areas of linguistics, education, and psychology. For the SCOPUS database, the subject was confined to arts and humanities, psychology, social sciences, and neuroscience. The subject could not be refined in the ERIC database since the subject area is education. As our searches were limited to databases that mostly indexed journal articles, we wanted to allow for seminal works to be included as well. This was done by searching the reference lists of the 128 obtained articles (see Figure 2). These seminal works were found repeatedly in the reference lists but did not show up in the search results, often because they were published as conference papers or book chapters.
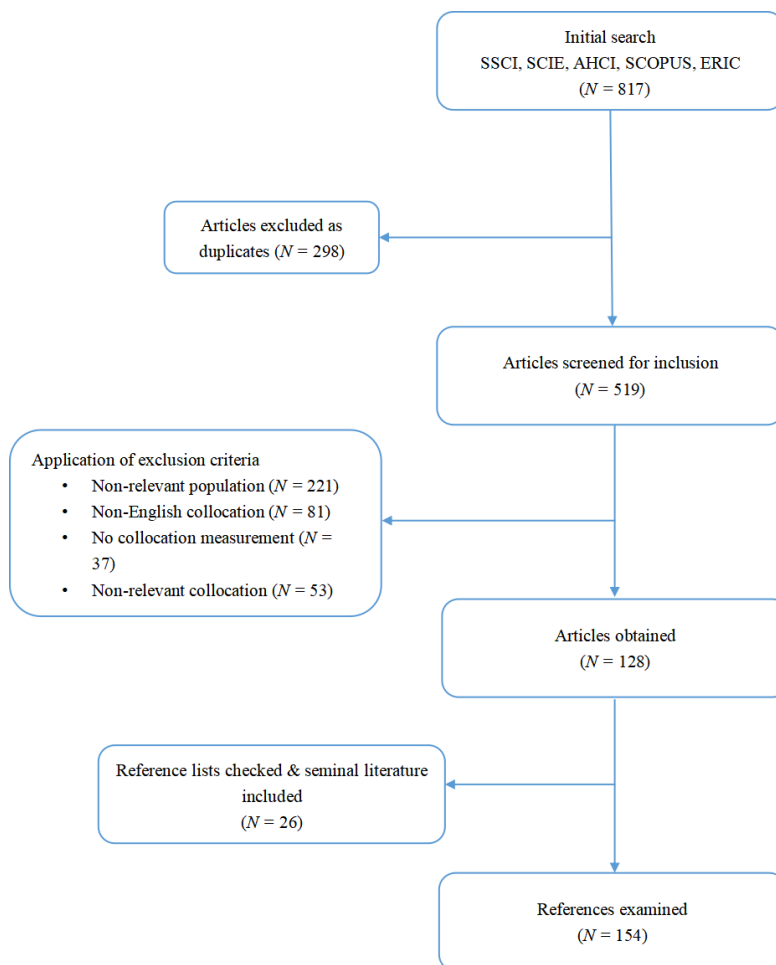
Figure 2:

Flow diagram for the search phases and results.

2.2.3 Coding and analysis

As our primary interest was exploring and evaluating existing collocation measurements, a coding scheme was developed to collect data from each sampled study. All 154 references were coded by the first author according to a coding scheme consisting of the following categories: (a) study identification (e.g., year of publication, author), (b) measurement targets and features (e.g., collocation definition, collocation types, approaches to collocation), (c) instruments (e.g., instrument name, item format, item number, scoring), (d) learners and contexts (e.g., L1 background, EFL or ESL context), and (e) reported reliability of the instrument. All coded data can be found in the supplementary materials，specifically within the "Coding Book".

Reliable coding was ensured by adhering to conventions suggested by Cooper (2017). The first and second authors co-designed the coding book and wrote definitions for the coding categories. Then, they worked through the coding of one reference together to ensure coding consensus. After that, to pilot the coding book, the first and second author independently coded three additional references and then discussed any differences found in their coding. Once the piloting of the coding book was complete, the first author independently extracted and coded data from the remaining references at two different times with a two-week period in-between. The intercoder agreement rate was calculated and reached over 99%. The inconsistent codes were discussed

between the first and second authors. Then, the second author double coded around 10% of the total retrieved references ($N = 15$), with an intercoder agreement of over 99%. The codes for "measurement target" were further scrutinized, as some researchers used different terminology to refer to the same type of vocabulary knowledge (e.g., meaning recall labeled as productive knowledge). The first author then did a third round of blind coding for this category to further ensure reliability of the codes for this category.

3 Results

We identified a set of six instruments, seven item formats and three other assessment tools, which have been extensively documented in the 154 articles reviewed. To provide a clear overview, concise compilations of the 16 measurements are presented in subsequent tables, as was done in similar review studies on vocabulary assessment (Boers 2011, 2013). Table 1 categorizes the instruments based on their ability to assess either receptive or (semi-)productive collocation knowledge, specifying the type of collocate-node relationship they evaluate. The reliability from relevant studies is also included in Table 1. Additionally, Table 1 outlines the format variations used by each instrument, accompanied by corresponding item examples. Table 2 summarizes information on seven item formats specifically targeting collocation knowledge. The remaining three tools – eye-tracking, writing task, and speaking task – are presented separately in Table 3 as they may not exclusively assess collocation knowledge. The supplementary materials, specifically within "Description of Instruments and Item Formats Reviewed" offer a detailed description of each instrument and item format.

Table 1:

Instruments assessing collocation knowledge.

Name	Type of collocation knowledge	Collocate-node relationship	Reliability	Format variation	Example item
Word Associates Test (WAT) (Read 1993)	Receptive collocation knowledge	Adjective-noun collocations	Cronbach's alpha of 0.9	Depth of Vocabulary Knowledge Test (DVK) (Qian 2002)	Instruction: From the two boxes, select four words that you think are relevant to the stimulus word (i.e., *sudden* in this example): *sudden* | beautiful **quick surprising** thirsty || **change** doctor **noise** school |
COLLEX (Gyllstad 2009)	Receptive collocation knowledge	Verb-noun collocations composed of delexicalized verbs	Cronbach's alpha of 0.89		Instruction: Each question consists of three word sequences. Your task is to choose one of the three word sequences marked with a), b), and c). Select the word sequence that you consider to be the most natural and commonly used by placing a clear *cross* under the corresponding letter in the box in the right column. a. drive a business b. run a business c. lead a business
COLLMATCH (Gyllstad 2009)	Receptive collocation	Verb-noun collocations	Cronbach's alpha of		*Instruction*: Your task is to decide whether the word combinations are used in the English language or

	knowledge	composed of delexicalized verbs	0.89	not. If you think a word combination is used in the English language, tick the yes box. If you don't think a word combination is used in the English language, tick the no box. catch a cold draw a limitation □ yes □ yes □ no □ no

PHRASE Test (Martinez 2011)	Receptive collocation knowledge	Phrasal verbs	Cronbach's alpha of 0.87 as reported in Kremmel et al. (2017)	Instruction: Find the phrase that has the closest meaning to the vocabulary in the sentence:
				go away: They will *go away*.
				    a. travel fast
				    b. leave
				    c. be tired
				    d. play

Frequency-based Collocation Test	Receptive collocation knowledge	Verb-noun and adjective-noun collocations	Cronbach's alpha of 0.77 (Nguyen and Webb 2017)	Instruction: Choose the most suitable verb (or adjective) that can be combined with each of the following nouns. There is only one correct answer among four options a, b, c, d. The correct answer is the verb (or the adjective) that can directly combine with the given noun without any other word between them: 1. money (1,000 frequency) a. check b. drop c. make d. miss 2. piano (2,000 frequency) a. grand b. traditional c. familiar d. entire
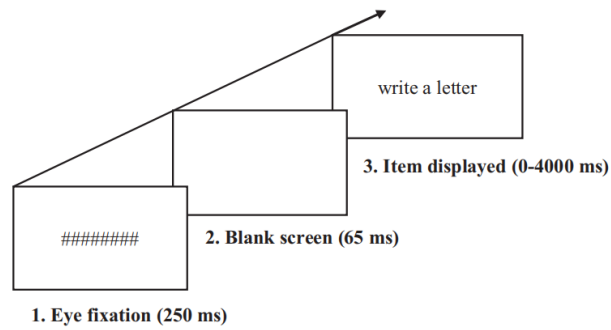
CONTRIX	Semi-productive collocation knowledge	Verb object-noun transparent, semi-transparent and non-transparent collocations	Cronbach's alpha of 0.89 (Revier 2009)	Instruction: Fill in the gap in the sentence by matching the verb with the determiner and the noun as options.

Instruction: Fill in the gap in the sentence by matching the verb with the determiner and the noun as options.

The quickest way to win a friend's trust is to show that you are able to _____ .	tell	a/an	joke
	take	the	secret
	keep	---	truth

Table 2:
Item formats assessing collocation knowledge.

Name	Type of collocation knowledge	Collocate-node relationship	Format variation	Example item
Multiple-choice	Receptive collocation knowledge	All lexical and grammatical collocations	Item presented in isolation	See example item in COLLEX
			Item presented in sentential context	See example item in Chou (2019) (providing contextual information in stem). The seasoning really ___ the flavor of the meat. A. brings out B. gets off C. pulls up D. goes back

				See example item in Book (2000) (providing contextual information in options).
				Instruction: Each of the four sentences is using the underlined verb in a different way. One of them is not really a correct usage of that word. Circle the letter corresponding to the least acceptable sentence.
				a. You are <u>covered</u> by your doctor's diagnosis.
				b. Another soldier <u>covered</u> Peter with a machine gun as he started forward.
				c. With this insurance policy, we will <u>cover</u> you in case of any accident.
				d. Wow, we've been driving fast! We <u>covered</u> almost three hundred miles in one day.
Acceptability judgment task	Receptive collocation knowledge	All lexical and grammatical collocations	Item presented online	See example item in Gyllstad and Wolter (2016). Participants are asked to press the "yes" or "no" key if they feel an item (e.g., write a letter) presented on the screen is acceptable collocation.

				write a letter 3. Item displayed (0-4000 ms) ######## 2. Blank screen (65 ms) 1. Eye fixation (250 ms) Presentation sequence for items in the semantic judgement task
			Item presented offline	See example in Zhang and Wen (2019). Learners are asked to rate the polysemous meanings of phrasal verbs used in different sentential contexts on a 5-point Likert scale from "very easy to guess" (very transparent) to "very difficult to guess" (very opaque).
Matching	Productive collocation knowledge	All lexical and grammatical collocations		See example item in CONTRIX.
Fill-in the blanks	Productive collocation knowledge	All lexical and grammatical collocations	Item with different cues provided	See example item in Boers et al. (2014) (no contextual information being provided). free ____ (ride) See example item in Bahns and Eldaw (1993) (providing contextual information as the cue with no form variation of the target). When she was a teenager, she used to _____ a diary. (keep)

				See example item in Bonk (2000) (providing contextual information as the cue with form variation of the target). He was accused of _____ rumors about other employers. (spreading) See example item in Fernández and Schmitt (2015) (providing contextual information/L1 translation and the first letters as the cue with form variation of the target). L_____ a_____, I liked playing with all kind of dolls (Cuando tenia 12 anos me gustaba mucho jugar con munecas, pero eso fue hace ya mucho tiempo./When I was 12 years old, I liked to play with dolls, but that was a long time ago.) (Long ago) See example item in Sonbul and Schmitt (2013) (providing contextual information/L1 translation and the number of letters as the cue with form variation of the target). Lung infections might lead to different conditions such as the (_ _ _ _ _ _ _ _ lung) syndrome. (decrease in the lung's size as shown in an X-ray) (vanishing)
Translation	Productive collocation knowledge	All lexical and grammatical collocations	L1 to L2 translation and L2 to L1 translation	See example item in Webb et al. (2013). 满足需要 (meet demands) Meet demands (满足需要)

			Item with different cues provided	See example item in Peters (2014).
				Geldelijke verplichting, schuld die binnen het jaar betaald moet worden (L1 German)
				L2 definition: a financial obligation, debt, claim, or potential loss due for payment within one year:
				c_____ l_____
Sentence generation	Productive collocation knowledge	All lexical and grammatical collocations		See example item in Schmitt (1999). Instruction: Produce three sentences per target word based on its semantic fields. *Massive* Semantic fields: 1) war, 2) finance or the economy, and 3) change Potential answers: 1) War: attack, damage, destruction; 2) Economics or finance: amount, billion, budget; and 3) Change: cause, changes, increase, influx
Error correction	Productive collocation knowledge	All lexical and grammatical collocations		See example item in Ha (2017). *Instruction*: Identify if these sentences are written appropriately (A) or inappropriately (I) by checking the corresponding column. Make correction if necessary. After fooling around for most of the semester, now he has to stay the piper and study over vacation. A I (√) correction (pay the piper)

Table 3:
Other tools assessing collocation knowledge.

Name	Type of collocation knowledge	Collocate-node relationship	Format variation	Example item
Eye-tracking	Receptive collocation knowledge	All lexical and grammatical collocations		See example item in (Choi 2017). Learners read the experimental text with target collocations while their eye movements were recorded by an eye-tracker.
Writing task	Productive collocation knowledge	All lexical and grammatical collocations		See example item in Nesselhauf (2003). Learners' argumentative essays of about 500 words were collected from the German subcorpus of the International Corpus of Learner English for the analysis of learners' collocation use.
Speaking task	Productive collocation knowledge	All lexical and grammatical collocations		See example item in Boers et al. (2006). Learners took part in a one-on-one interview with two parts. 1. Learners had a prepared conversation with the interviewer after reading a short article and reflecting on it. 2. Learners had a spontaneous conversation with the interviewer about an unprepared but familiar topic.

4 Discussion

The present systematic review has indicated that there are six instruments (i.e., WAT, COLLEX, COLLMATCH, the PHRASE Test, the Frequency-based Collocation Test, and CONTRIX), seven item formats (i.e., multiple-choice, judgment task, matching, fill-in the blanks, translation, sentence generation, and error correction) and three other assessment tools (i.e., writing tasks, speaking tasks, and eye-tracking) language instructors and researchers may select from when aiming to measure receptive and productive collocation knowledge.

Regarding the six instruments, the most notable finding is that five out of six instruments assess receptive collocation knowledge, and only one (i.e., CONTRIX) measures learners' productive knowledge of verb object-noun collocations in a semi-productive format. A decade ago, Schmitt (2014) noted that compared with receptive collocation knowledge measurements, fewer validated productive collocation knowledge measurements were available. Today, validated instruments to measure learners' productive collocation knowledge are still lacking, presenting a significant gap that hinders the enhancement of fluency, accuracy, and complexity in language production (e.g., Boers et al. 2006; Erma and Warren, 2000; Nesselhauf 2003).

Concerning the item formats, almost all of them were used in the reviewed studies to measure learners' productive knowledge. Examples are fill-in the blanks (e.g., Bahns and Eldaw, 1993), translation (e.g., Snoder and Reynolds 2019), sentence generation (e.g., Schmitt 1998b) and error correction (e.g., Ha 2017). A typical item format for assessing learners' receptive knowledge was found to be the multiple-choice format

(e.g., Bonk 2000). We believe that these item formats are user-friendly tools for teachers who want to assess collocations in the classroom. Teachers can select the collocations that they want their learners to know, turn them into learning targets, and use the targets in exercises or in a test through the item formats reviewed here. Of course, it is imperative that there is no discrepancy between the focus in class (e.g., on receptive knowledge through reading or listening) and the collocation assessment (e.g., productive knowledge through fill-in the blanks). Teachers should also critically think about the pros and cons of those formats. For example, the multiple-choice format provides cues in the contexts of the stems that could facilitate informed guessing but the format also requires more processing time for learners to consider and comprehend each option within sentential context(s) (Schmitt 2010). Although translation tasks on the sentential level may prove to be more difficult for learners compared to receptive tasks (e.g., MCQ), they may be a meaningful way to elicit target collocations, if the task is well designed. Otherwise, the absence of using particular collocations in translation may reveal little about learners' productive collocation knowledge. In the fill-in-the-blank formats, teachers should ask themselves whether using cues such as the first letter results in an overestimate of learners' productive collocation knowledge. Teachers should exercise vigilance regarding the varying cues employed to assess different degrees of collocation knowledge. Specifically, when utilizing the learners' L1 as a cue to elicit the corresponding L2 collocation, the assessment primarily centers on learners' productive form aspect of collocation knowledge.

With the other assessment tools (i.e., speaking tasks, writing tasks, and eye-tracking) rich data can be gathered about learners' receptive or productive collocation knowledge, processing, or use. However, it should be noted that those tasks are not exclusively used to measure collocation knowledge. Speaking and writing tasks, for example, are often used to investigate language complexity, accuracy, and fluency (e.g., Housen and Kuiken 2009), and eye-tracking to examine vocabulary processing (e.g., Roberts and Siyanova-Chanturia 2013). Furthermore, those tasks also bring some challenges. For speaking tasks, for example, McGuire and Larson-Hall (2017) questioned the reliability of counting hesitation-free word strings as collocations as the rater may fall into the trap of "circularity in measuring" (p. 12). This happens when learners are trained to use more collocations, and as a result their language use appears to be more fluent. Measures that consider hesitation-free strings as an index of collocation knowledge need to explore an objective standard such as the time between two-word intervals for determining whether to treat word combinations as collocations. When using writing tasks, researchers must decide whether they will use manual or automatic extraction. Extracting collocations manually is a very time-consuming process and may be less systematic, whereas the automated extraction tools, such as the AntConc Software (Anthony 2018) are faster in their search for and extraction of collocations. However, machine generation could lead to word combinations of semantically non-meaningful word combinations, such as *of the.* Eye-tracking, which provides a more detailed account of the timing and level of attention devoted to linguistic input (Choi 2017), serves as a complementary tool to traditional offline

measures for assessing a learner's vocabulary processing during reading (Conklin et al. 2018). However, the practical application of eye-tracking in classroom assessment by teachers is very limited.

When it comes to researching collocations, we believe that all reviewed collocation measurements can contribute to the assessment of L2 learners' receptive and/or productive collocation knowledge. However, it should be remarked that a first problem – well known in collocation research – lies in the definition or classification of collocations (e.g., Gablasova et al. 2017; Siyanova-Chanturia and Omidian 2020; Uchihara et al. 2021). The reviewed studies indicate that there is still no consensus on the definition or classification of collocations, which makes comparing the results of previously used measurements difficult. For example, most studies either adhere to the frequency-based approach or the phraseological approach—both of which have their own respective strengths and limitations (Nguyen and Webb 2017).

Second, although the reviewed measurements have covered different types of collocation knowledge or used different types of collocations (e.g., adjective-noun, verb-noun), the use of only one measurement is limited in that it will be unable to reveal learners' complete collocation knowledge, an inherent limitation of all vocabulary tests (Nation and Webb 2011). Additionally, the reviewed instruments and item formats tend to have a limited number of items that may not adequately measure collocation knowledge (Nguyen and Webb 2017). In this light, it is worth noting that the scoring also varies among the different measurements. The WAT, for example, uses a complicated scoring system and considers partial knowledge, whereas other

measurements use the *all-or-nothing approach* (Schmitt 2010), in which only responses having all the correct answers selected are counted. Given these variations, it is advisable that future research focuses on the development of more granular scoring methods, incorporating factors such as frequency indices or other word difficulty metrics, to enhance the precision and nuance of the assessment process.

Third, another point that might be relevant for researchers is that the reviewed collocation assessment tools, in general, lack a clear description of their development procedures and test features. Except for studies aimed at developing specially designed collocation instruments (e.g., Gyllstad 2009; Martinez 2011), most publications that included collocation measurements did not justify the selection of the item format or explain how items that comprised the instrument were developed. This review's findings are in line with Schmitt et al. (2020), who showed a lack of a more systematic and rigorous procedure for test development. This presents a hinderance for meaningful development of collocation assessment research.

Fourth, the present review revealed that participants in the studies were mostly intermediate to advanced university-age L2 learners. This gives a rather limited picture on L2 collocation learning, since language learning is something that happens across the lifespan (e.g., Pfenninger et al. 2023; Xia et al. 2022). Finally, contrary to single-word vocabulary knowledge (McLean et al. 2015), no validated instruments measuring aural collocation knowledge are currently available. Aural collocation knowledge contributes to listening, as Stæhr (2009) found the receptive depth of vocabulary knowledge measured by the WAT can predict the listening comprehension of advanced

EFL learners. However, the WAT was designed to measure receptive written word knowledge and not receptive spoken word knowledge. There should be some relationship between the two, but researchers have yet to take up the challenging task of developing an aural collocation instrument with a high level of face validity. As far as we are aware, besides the limited number of studies targeting young learners found when we conducted this systematic review, no studies have been conducted in which listening alone was used to measure the receptive knowledge of collocations. Even an instrument that required learners to orally produce their choice (Smith and Murphy 2015) intended to measure written collocation knowledge.

However, the shortcomings of the reviewed assessment tools are important challenges that might drive future research. First, it is of vital importance to operationalize a construct before designing efficient, reliable, and valid instruments (Beglar and Nation 2007). Thus, future measurements need to clearly present the operationalization of collocations. We advocate an integration of the frequency-based and phraseological approaches as this may provide a more nuanced understanding of the concept of collocation knowledge (Bestgen 2017; Siyanova-Chanturia 2015). Second, we advocate for considering the use of a more complete approach, with measurements that not only tap into one subcomponent of collocation knowledge but aim to measure various levels of collocation knowledge simultaneously, as was done in the study of Lee and Shin (2021). Not only triangulation of both traditional offline measurements (e.g., a fill-in the blanks exercise) and online measurements (e.g., through eye-tracking), but also the combination of quantitative data (e.g., free writing)

and qualitative data (e.g., interviews) might provide a fuller picture of a learner's receptive and productive collocation knowledge. Future measurements also need to include different collocation types and an adequate number of items belonging to different frequency levels, since frequency is known to affect collocation knowledge. Other item-related factors that might affect learners' collocation knowledge are, to name just a few, semantic compositionality, transparency of meaning, congruency, and imageability/concreteness (e.g., Boers 2020). Clarity over these methodological decisions and the justification for them will help avoid misuse of the assessment tools or misinterpretation of the results, while increasing the quality of data gathered. Third, a clear description of item development should be reported, including, for instance, how sentences are developed to elicit the use of target collocations in a translation task. In agreement with Schmitt et al. (2020), we emphasize that future collocation instrument development should include a clear description of "the test's purpose, intended testees and educational context, the particular aspects of vocabulary knowledge which are being measured, and the way in which the test scores should be interpreted" (p. 1). In turn, this will contribute to increased quality in standardized testing, more transparency, and, eventually, more rigorous validation evidence. Fourth, it would be interesting to assess collocation knowledge of L2 learners of all ages, going through the stages of childhood, adolescence, adulthood, and third age. Therefore, development of suitable collocation measurements to address for example younger learners is sorely needed. Smith and Murphy (2015) already offer a good example of how English as an additional language (EAL) primary school learners' receptive knowledge of verb-noun

collocations was measured using a matching format. However, not only age should be taken into consideration. Another challenge when developing assessment tools is to consider other learner-related factors (e.g., phonological short-term memory, prior L2 learning experience, prior L2 vocabulary knowledge), because collocations that pose few problems to one learner may be quite difficult to another (e.g., Boers 2020). Finally, because most receptive collocation knowledge instruments reviewed required learners to read, future studies targeting collocation listening beyond the use of a listening task to measure receptive collocation knowledge will be highly valued additions.

5 Conclusion

This systematic review has revealed that 16 assessment tools have been employed within empirical studies that focus on receptive and productive collocation knowledge. The tools range from specially designed instruments to more general teacher-designed classroom-based item formats using more controlled tasks (e.g., sentence generation) to free production tasks (e.g., essay writing). When selecting a collocation instrument, one needs to be cautious as to the purpose of the measurement, what information it can yield, what concerns need to be addressed and the potential learners' characteristics. After reviewing the available collocation instruments and item formats, we expect to see future development of measurements that provide a clear operationalization of the collocation(s) intended to be measured and a detailed description of item development, including item/option selection. We also hope that multiple measurements can be combined to obtain a fuller picture of learners' collocation knowledge. As collocation knowledge is not only an important part of vocabulary learning but also contributive to

other language skills, such as listening, reading, speaking, and writing, we feel the use of more robust measurements in the language classroom that carefully control for measuring multiple language skills will have a positive washback effect on both collocation teaching and learning.

References

Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis and applications*, 101-122. Oxford: Oxford University Press.

Anthony, Laurence. 2018. *AntConc* (Version 3.5.7) . Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Bahns, Jens & Eldaw Moira. 1993. Should we teach EFL students collocations? *System* 21(1). 101-114. doi:10.1016/0346-251X(93)90010-E

Baldwin, Timothy & Kim Su Nam. 2010. Multiword expressions. In Nitin Indurkhya & Fred. J. Damerau (eds.), *Handbook of natural language processing,* 267-292. Boca Raton: CRC Press.

Beglar, David & Nation Paul. 2007. A vocabulary size test. *The Language Teacher* 31(7). 9-13.

Benson, Morton, Benson Evelyn & Ilsen F. Robert. 1986. *The BBI combinatory dictionary of English: A guide to word combinations.* Amsterdam, Netherlands: John Benjamins.

Bestgen, Yves. 2017. Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System* 69. 65-78. doi:10.1016/j.system.2017.08.004

Biber, David. 1993. Representativeness in corpus design. *Literacy and Linguistic Computing* 8(4). 243-257.

Biber, Douglas, Conrad Susan & Leech Geoffrey. 2002. *Longman student grammar of*

spoken and written English. Harlow: Longman.

Boers, Frank. 2011. Cognitive semantic ways of teaching figurative phrases: An assessment. *Review of Cognitive Linguistics* 9(1). 227-261. doi:10.1075/ml.9.1.11boe

Boers, Frank. 2013. Cognitive Linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching* 46(2). 208–224. https://doi.org/10.1017/S0261444811000450

Boers, Frank. 2020. Factors affecting the learning of multiword items. In Stuart Webb (ed.), *The Routledge handbook of vocabulary studies*, 143–157. New York: Routledge.

Boers, Frank, Eyckmans June, Kappel Jenny, Stengers Hélène & Demecheleer Murielle. 2006. Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research* 10(3). 245-261. doi:10.1191/1362168806lr195oa

Boers, Frank, Lindstromberg Seth & Eyckmans June. 2014. Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics* 11(1). 41-62.

Boone, Griet, De Wilde Vanessa & Eyckmans June. 2022. A longitudinal study into learners' productive collocation knowledge in L2 German and factors affecting the learning. *Studies in Second Language Acquisition* 45(2). 503–525. https://doi.org/10.1017/S0272263122000377

Bonk, William J. 2000. Testing ESL learners' knowledge of collocations. In Tom

Hudson & J. D. Brown (eds.), *A focus on language test development*, 113–142. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Brown, Dale. 2018. *Developing a measure of L2 learners' productive knowledge of English collocations*. Wales, UK: Cardiff University dissertation.

Chapelle, Carol A. 2011. Validity argument for language assessment: the framework is simple. *Language Testing* 29(1). 19-27.

Choi, Sungmook. 2017. Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research* 21(3). 403-426. doi:10.1177/1362168816653271

Chou, Mu-Hsuan. 2019. Task-supported language teaching to enhance young EFL adolescent learners' comprehension and production of English phrasal verbs in Taiwan. *Education 3-13* 48(4). 455-470. doi:10.1080/03004279.2019.1617328

Collins, Allan M. & Loftus Elizabeth F. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82(6). 407-428. doi:10.1037/0033-295X.82.6.407

Conklin, Kathy, Pellicer-Sánchez Ana & Carrol Gareth. 2018. Eye-tracking: *A guide for applied linguistics research*. Cambridge, UK: Cambridge University Press.

Cooper, Harris (ed.). 2017. *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, California: Sage Publications, Inc.

Cowie, Anthony. P. 1994. Phraseology. In Ronald E. Asher (ed.), *The encyclopedia of language and linguistics*, 3168–3171. Oxford, UK: Pergamon.

Crossley, Scott A., Salsbury Tom & McNamara Danielle S. 2015. Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics* 36(5). 570-590. doi:10.1093/applin/amt056

Davies, M. 2008. The corpus of contemporary American English: 520 million words, 1990–present. Retrieved from http://corpus.byu.edu/coca

Dronjic, Vedran & Helms-Park Rena. 2012. Fixed-choice word-association tasks as second-language lexical tests: What native-speaker performance reveals about their potential weaknesses. *Applied Psycholinguistics* 35(01). 193-221. doi:10.1017/s014271641200029x

Durrant, Philip. 2014. Corpus frequency and second language learners' knowledge of collocations A meta-analysis. *International Journal of Corpus Linguistics* 19(4). 443-477. doi:10.1075/ijcl.19.4.01dur

Durrant, Philip & Schmitt Norbert. 2009. To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching* 47(2). 157-177. doi:10.1515/iral.2009.007

Durrant, Philip & Schmitt Norbert. 2010. Adult learners' retention of collocations from exposure. *Second Language Research* 26(2). 163-188. doi:10.1177/0267658309349431

Ellis, Rod. 1983. Formulaic speech in early classroom second language development. In Jean Handscome, Richard A. Orem & Barry P. Taylor (eds.), *On TESOL '83: The question of control*, 23-28. Washington, DC: TESOL.

Erman, Britt & Warren Beatrice. 2000. The idiom principle and the open choice

principle. *Text-Interdisciplinary Journal for the Study of Discourse* 20(1). 29-62.

Fan, May. 2009. An exploratory study of collocational use by ESL students - A task based approach. *System* 37(1). 110-123. doi:10.1016/j.system.2008.06.004

Fazlali, Behnaz & Shahini A. 2019. The effect of input enhancement and consciousness-raising techniques on the acquisition of lexical and grammatical collocation of Iranian EFL learners. *TESL-EJ* 24(2). 1-21.

Fernández, González & Schmitt Norbert. 2015. How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL - International Journal of Applied Linguistics* 166(1), 94-126.

Foster, Pauline. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In Martin Bygate, Peter Skehan & Merrill Swain (eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing,* 75-94. London: Longman.

Frenck-Mestre, Cheryl & Prince Paul. 1997. Second language autonomy. *Journal of Memory and Language* 37(4). 481-501. https://doi.org/10.1006/jmla.1997.2526

Gablasova, Dana, Brezina Vaclav & McEnery Tony. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning* 67(S1). 155–179.

Garner, James, Crossley Scott & Kyle Kristopher. 2019. N-gram measures and L2 writing proficiency. *System* 80. 176-187. doi:10.1016/j.system.2018.12.001

González-fernández, Beatriz & Schmitt Norbert. 2020. Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics* 41(4). 481–505. https://doi.org/10.1093/applin/amy057

Gough, David, Oliver Sandy & Thomas James (eds.). 2012. *An introduction to systematic reviews*. Thousand Oaks, California: SAGE.

Gries, Stefan. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: some necessary clarifications. *Studies in Language* 36(3). 477-510.

Groom, Nicholas. 2009. Effects of second language immersion on second language collocational development. In Andy Barfield & Henrik Gyllstad (eds.), *Researching collocations in another language: Multiple interpretations*, 21-33. Basingstoke: Palgrave.

Gyllstad, Henrik. 2009. Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In Andy Barfield & Henrik Gyllstad (eds.), *Researching collocations in another language: Multiple interpretations*, 21-33. Basingstoke: Palgrave.

Gyllstad, Henrik & Wolter Brent. 2016. Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning* 66(2). 296-323. doi:10.1111/lang.12143

Ha, Do Thi. 2017. The impact of different instructions on Vietnamese EFL students' acquisition of formulaic sequences. *English Language Teaching* 10(8). 18-31.

Henriksen, Birgit. 2013. Research on L2 learners' collocational competence and

development－a progress report. In Camilla Bardel, Christina Lindqvist & Batia Laufer (eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis,* 29-56. EUROSLA: The European Second Language Association.

Housen, Alex & Kuiken, Folkert. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30(4). 461–473. https://doi.org/10.1093/applin/amp048

Hou, Jun Ping , Loerts Hanneke Loerts & Verspoor Marjolijn H. 2018. Chunk use and development in advanced Chinese L2 learners of English. *Language Teaching Research* 22(2). 148-168. doi:10.1177/1362168816662290

Howarth, Peter. 1998. Phraseology and second language proficiency. *Applied Linguistics* 19(1). 24-44. doi:10.1093/applin/19.1.24

Hughes, Arthur. 2005. *Testing for Language Teachers*. Cambridge, UK: Cambridge University Press.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge, UK: Cambridge University Press.

Jaén, María Moreno. 2007. A corpus-driven design of a test for assessing the ESL collocational competence of university students. *International Journal of English Studies* 7(2). 127-147.

Kane, Michael. 2006. Validation. In Robert Brennen (ed.), *Educational Measurement* (Vol. 50), 1-73. Westport, CT: Greenwood.

Kim, Sugene. 2017. Minimum input, maximum output, indeed! teaching collocations

through collocation dictionary skills development. *Lexikos* 27. 265-286.

Knell, Ellen. 2018. Role of oral language in the development of L2 literacy skills. In John Liontas (ed.), *The TESOL Encyclopedia of English Language Teaching*, 1-8. Wiley Blackwell.

Kremmel, Benjamin, Rrunfaut Tineke & Alderson Charles. 2017. Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics* 38(6). 848-870.

Laufer, Batia & Girsai Nany. 2008. Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics* 29(4). 694-716. doi:10.1093/applin/amn018

Lee, Senyung. 2019. L1 transfer, proficiency, and the recognition of L2 verb-noun collocations: A perspective from three languages. *IRAL —International Review of Applied Linguistics in Language Teaching* 47(2). 181-208. doi:10.1515/iral-2018-0220

Lee, Senyung & Shin Sun-Young. 2021. Towards improved assessment of L2 collocation knowledge. *Language Assessment Quarterly* 18(4). 419–445. https://doi.org/10.1080/15434303.2021.1908295

Lei, Lei & Liu Dilin. 2018. The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics* 23(2). 216–243.

Li, Senyung. 2017. Using corpora to develop learners' collocational competence. *Language Learning and Technology* 21(3). 153-171.

Lundell, Fanny Forsberg & Lindqvist Christina. 2014. Lexical aspects of very advanced

L2 French. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes* 70(1). 28-49. doi:10.3138/cmlr.1598

Lundell, Fanny Forsberg, Lindqvist Christina & Edmonds Amanda. 2018. Productive collocation knowledge at advanced CEFR levels: Evidence from the development of a test for advanced L2 French. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes* 74(4). 627-649. doi:10.3138/cmlr.2017-0093

Marsden, Emma, Thompson Sophie & Plonsky Luke. 2018. A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics* 39(5). 861-904. doi:10.1017/S0142716418000036

Martinez, Ron. 2011. Putting a test of multiword expressions to a test. Paper presented at the IATEFL Testing, Evaluation and Assessment Conference (TEA SIG), University of Innsbruck, 16, September.

Martinez, Ron & Murphy Victoria A. 2011. Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly* 45(2). 267-290.

Martinez, Ron & Schmitt Norbert. 2012. A phrasal expressions list. *Applied Linguistics* 33(3). 299-320.

McGuire, Michael & Larson-Hall Jenifer. 2017. Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal* 34(3). 1-25. doi:10.18806/tesl.v34i3.1271

McLaughlin, Judith, Osterhout Lee & Kim Albert. 2004. Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature*

Neuroscience 7(7). 703-704. doi:10.1038/nn1264

McLean, Stuart. 2017. Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics* 39(6). 823-845. doi:10.1093/applin/amx003

McLean, Stuart, Kramer Brandon & Beglar David. 2015. The creation and validation of a listening vocabulary levels test. *Language Teaching Research* 19(6). 741-760. doi:10.1177/1362168814567889

McNamara, Timothy P. 2005. *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

McNeill, Arthur. 1996. Vocabulary knowledge profiles: Evidence from Chinese-speaking ESL teachers. *Hong Kong Journal of Applied Linguistics* 1(1). 39-64.

Morimoto, Shun & Loewen, Shawn. 2007. A comparison of the effects of image-schema-based instruction and translation-based instruction on the acquisition of L2 polysemous words. *Language Teaching Research* 11(3). 347-372. doi:10.1177/1362168807081181

Myles, Florence & Cordier Caroline. 2017. Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition* 39(1). 3–28. https://doi.org/10.1017/S027226311600036X.

Nation, Paul. 2013. *What Should Every EFL Teacher Know*. Seoul: Compass Publishing.

Nation, Paul. 2022. *Learning vocabulary in another language*. Cambrigde: Cambrigde University Press.

Nation, Paul & Webb Stuart. 2011. *Researching and Analyzing Vocabulary*. Boston,

MA: Heinle, Cengage Learning.

Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24(2). 223-242.

Nguyen, Thi My Hang & Webb Stuart. 2017. Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research* 21(3). 298-320. doi:10.1177/1362168816639619

Paquot, Magali. 2018. Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly* 15(1). 29-43. doi:10.1080/15434303.2017.1405421

Paquot, Magali. 2019. The phraseological dimension in interlanguage complexity research. *Second Language Research* 35(1). 121-145. doi:10.1177/0267658317694221

Paquot, Magali & Granger Sylviane. 2012. Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32. 130-149. doi:10.1017/S0267190512000098

Pawley, Andrew & Syder Frances Hodgetts. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Jack C. Richards & Richard W. Schmidt (eds.), *Language and Communication*, 29–59. London, UK: Longman.

Peters, Elke. 2014. The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research* 18(1). 75-94. doi:10.1177/1362168813505384

Peters, Elke. 2016. The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research 2*0(1). 113–138. https://doi.org/10.1177/1362168814568131

Pfenninger, Simone E., Festman Julia & Singleton David. 2023. *Second language acquisition and lifelong learning*. New York: Routledge. https://doi.org/10.4324/9781003168935

Phakiti, Aek, DeCosta Peter, Plonsky Luke & Starfield Sue. 2018. *The Palgrave Handbook of Applied Linguistics Research Methodology*. UK: Springer.

Plonsky, Luke & Kim Youjin. 2016. Task-based learner production: a substantive and methodologocal review. *Annual Review of Applied Linguistics* 36. 73-97. doi:10.1017/S0267190516000015

Plonsky, Luke & Oswald Frederick. 2015. Meta-analysing second language research. In Luke Plonsky (ed.), *Advancing quantitative methods in second language research*, 106–128. New York: Routledge.

Qian, David. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52(3). 513-536. doi:Doi 10.1111/1467-9922.00193

Qian, David & Schedl Mary. 2004. Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 21(1). 28-52.

Read, John. 1993. The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10(3). 355-371. doi:10.1177/026553229301000308

Read, John. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Read, John & Chapelle Carol. A. 2001. A framework for second language vocabulary assessment. *Language Testing* 18(1). 1-32. doi:10.1177/026553220101800101

Revier, Robert Lee. 2009. Evaluating a new test of whole English collocations. In Andy Barfield & Henrik Gyllstad (eds.), *Researching collocations in another language: Multiple interpretations*, 125-138. Basingstoke: Palgrave.

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

Reynolds, Barry Lee & Wible David. 2014. Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly* 48(4). 843-861.

Roberts, Lee & Siyanova-Chanturia Anna. 2013. Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition* 35(2). 213-235. doi:10.1017/S0272263112000861

Schmitt, Norbert. 1998a. Measuring collocational knowledge: Key issues and an experimental assessment procedure. *International Journal of Applied Linguistics* 119(1). 27-47.

Schmitt, Norbert. 1998b. Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning* 48(2). 281-317. doi:10.1111/1467-9922.00042

Schmitt, Norbert. 1999. The relationship between TOEFL vocabulary items and meaning, association, collocation, and word-class knowledge. *Language*

Testing 16(2). 189-216.

Schmitt, Norbert. 2010. *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, Norbert. 2014. Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64(4). 913-951. doi:10.1111/lang.12077

Schmitt, Norbert, Nation Paul & Kremmel Benjamin. 2020. Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching* 53(1). 109-120.

Schmitt, Norbert, Ng Janice Wun Ching & Garras John. 2010. The Word Associates Format: Validation evidence. *Language Testing* 28(1). 105-126. doi:10.1177/0265532210373605

Schmitt, Norbert, Schmitt Diane & Clapham Caroline. 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18(1). 55-88. doi:10.1177/026553220101800103

Shin, Dongkwang & Nation Paul. 2007. Beyond single words: the most frequent collocations in spoken english. *ELT Journal* 62(4). 339-348.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, Anna. 2015. Collocation in beginner learner writing: A longitudinal study. *System* 53. 148-160. doi:10.1016/j.system.2015.07.003

Siyanova-Chanturia, Anna & Omidian Taha. 2020. Key issues in researching multiword items. In Webb Stuart (ed.), *Routledge handbook of vocabulary*

studies, 529–544. London: Routledge.

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

Skrzypek, Agnieszka. 2010. *The relationship between phonological short-term memory and the development of L2 vocabulary and collocational knowledge: A study of adult polish learners of English at the A2 and B1 levels of English proficiency.* Dublin, Ireland:Trinity College Dublin dissertation. Retrieved from http://hdl.handle.net/2262/90603

Skrzypek, Agnieszka & Singleton David. 2013. Productive knowledge of English collocations in adult polish learners: The role of short-term memory. *Vigo International Journal of Applied Linguistics* 10(1). 105-129.

Smith, Sara A. & Murphy Victoria A. 2015. Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing* 28(3). 347-369. doi:10.1007/s11145-014-9527-y

Snoder, Per. 2017. Improving English learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Canada Journal* 34(3). 140-164.

Snoder, Per & Reynolds Barry Lee. 2019. How dictogloss can facilitate collocation learning in ELT. *ELT Journal* 73(1). 41-50. doi:10.1093/elt/ccy024

Sonbul, Suhad & Schmitt Norbert. 2013. Explicit and implicit lexical knowledge:

Acquisition of collocations under different input conditions. *Language Learning* 63(1). 121-159. doi:10.1111/j.1467-9922.2012.00730.x

Stæhr, Lars Stenius. 2009. Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31(4). 577-607. doi:10.1017/S0272263109990039

Uchihara, Takumi, Webb Stuart, Saito Kazuya & Trofimovich Pavel. 2021. The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition* 44(2). 357–380. https://doi.org/10.1017/S0272263121000218

Vilkaité, Laura. 2016. Formulaic language is not all the same: Comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji Kalbotyra* 8, 28-54. https://doi.org/10.15388/TK.2016.17505

Wang, Hung-Chun & Shih Su-Chin Shih. 2011. The role of language for thinking and task selection in EFL learners' oral collocational production. *Foreign Language Annals* 44(2). 399-416. doi:10.1111/j.1944-9720.2011.01135.x

Webb, Staurt, Newton Jonathan & Chang Anna. 2013. Incidental learning of collocation. *Language Learning* 63(1). 91-120. doi:10.1111/j.1467-9922.2012.00729.x

Wesche, Marjorie & Paribakht T. Sima. 1996. Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review* 53(1). 13-40. doi:10.3138/cmlr.53.1.13

Wiliam, Dylan. 2011. What is assessment for learning? *Studies in Educational Evaluation* 37(1). 3-14. doi:https://doi.org/10.1016/j.stueduc.2011.03.001

Wood, David. 2020. Classifying and identifying formulaic language. In Webb Stuart (ed.), *Routledge handbook of vocabulary studies*, 30-45. London: Routledge.

Wolter, Brent & Gyllstad, Henrik. 2011. Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics* 32(4). 430-449. doi:10.1093/applin/amr011

Wray, Alison. 2000. Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics* 21(4). 463–489.

https://doi.org/10.1093/applin/21.4.463

Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, Alison. 2012. What do we (think we) know about formulaic language? an evaluation of the current state of play. *Annual Review of Applied Linguistics* 32. 231-254. doi:10.1017/S026719051200013X

Xu, Jing. 2018. Measuring "spoken collocational competence" in communicative speaking assessment. *Language Assessment Quarterly* 15(3). 255-272. doi:10.1080/15434303.2018.1482900

Zhang, Xian & Lu Xiaofei. 2015. The relationship between vocabulary learning strategies and breadth and depth of vocabulary knowledge. *The Modern Language Journal* 99(4). 740-753. doi:10.1111/modl.12277

Zhang, Xiaopeng & Wen Ju. 2019. Exploring multiple constraints on second language development of English polysemous phrasal verbs. *Applied Psycholinguistics* 40(5). 1073-1101. doi:10.1017/S0142716419000146

Xia, Detong, Chen Yudi & Pae Hye K. 2022. Lexical and grammatical collocations in beginning and intermediate L2 argumentative essays: a bigram study. *International Review of Applied Linguistics in Language Teaching* 61(4). 1421-1453. doi: https://doi.org/10.1515/iral-2021-0188