

Measuring Productive Vocabulary Depth: Current Practices and New Perspectives

Wassila TEBA¹, Yasmine BOUKHEDIMI²

¹ University of Abou el Kacem Saâdallah, Algiers 2 (Algeria),
tebbaa.wassila@ensb.dz

² University of Abou el Kacem Saâdallah, Algiers 2 (Algeria),
yboukhedimi@yahoo.fr

Received: 16/2/2023 Accepted: 2/9/2023 Published:31/3/2024

Abstract:

Second and foreign language depth of vocabulary knowledge is a challenging issue where no consensus has yet been reached amongst researchers as concerns its conception. The multidimensional nature of word depth meaning made it difficult to design accurate measures in assessing vocabulary. This article reviews the incremental definitions of deep of lexical knowledge and its existing established measures. It concludes that previous studies focused on the assessment of some aspects of depth (e.g. collocations, synonyms, associations and using words at sentence level context), but they mostly neglected the entire language communicative contexts. These insights are important for language assessors who should consider more contextualized measures of vocabulary use; especially that language learning is communicatively oriented.

Keywords: contextualization; depth of vocabulary knowledge; vocabulary assessment.

1. INTRODUCTION

The domain of second/foreign language (SL/FL) vocabulary assessment has received an increasing interest since 1990s (Mukarto, 2005; Thékes, 2018; Anthony, 2020; Read & Dang, 2022). However, hardly or little attention is paid to productive depth of vocabulary knowledge since assessment methods of this attribute has not yet reached an extensively acceptable levels of reliability up until recently (Miralpeix, 2020). Despite the fact that studies conducted on vocabulary acquisition has covered various areas such as, vocabulary growth (Schmitt, 1998); the FL mental lexicon (Meara, 2009); and vocabulary learning strategies (Nation, 1990), considerable emphasis has now been put on the assessment of vocabulary size, mainly lexical richness and diversity (Webb, 2020) to produce reliable estimates of learners' overall lexical repertoire. (Daller, Milton, & Treffers-Daller, 2007)

Vocabulary size has long been investigated in various studies that led to fruitful, valid, and accurate multi-dimensional batteries of tests. Tests of this type either target students' receptive or productive vocabulary. They involve, among other measures, multiple-choice tests; Yes/No format tests (Meara & Buxton, 1987) for written receptive vocabulary size; AuralLex test (Milton & Hopkins, 2005) and a dictation test (Fountain & Nation, 2000) for aural receptive vocabulary size that provide an estimate of receptive word knowledge in learners.

There exist numerous test formats developed in quest of testing learners' productive vocabulary. These involve, for instance, Vocab Profiler or Lexical Frequency Profile (Cobb, 2006) and Lexical Diversity (McCarthy & Jarvis, 2007) tests that are used to analyze samples of writings intended to measure texts' lexical richness and lexical diversity respectively. In these assessment tools, learners' language proficiency is measured by means of estimating learners' overall productive vocabulary knowledge.

Mostly, size or breadth tests require learners to merely select responses from closed options or distractors, to match words to definitions or to fill in the blanks; the aim of which is to check meaning recognition. Note that even those measures intended for productive word use tend to count the number of words a learner knows to provide an overall estimate of vocabulary size. All of these receptive and productive batteries entail that superficial knowledge from the part of learners and they, adversely, do not practically measure how deeply target words are known. (Read, 1989)

In order to measure the quality of learners' knowledge of lexical items, different perspectives have been highlighted, and a number of tests have been developed in the literature of vocabulary assessment. Yet, research conducted on depth of vocabulary knowledge is said to be still sustained and immature (Webb, 2013; Read & Dang, 2022) and lacks reliable research tools. There exist a great deal of tests developed in search of evaluating L2 (second language) vocabulary development; yet, assessments are biased towards vocabulary breadth (Edmonds, Clenton, & Elmetaher, 2022). In this paper we survey the most established and extensively used measures on testing depth

of vocabulary knowledge, their merits and limitations upon which we suggest alternative perspectives for further remedial.

2. Operationalization of Deep Word Knowledge:

Vocabulary or vocabulary knowledge is a multi-dimensional theoretical construct as it comprises various aspects in knowing a word (Anderson & Freebody, 1981; Cronbach, 1942; Dale, 1965; Richards, 1976; Nation, 1990; 2001; 2013; Read, 2000; Daller et al., 2007; Webb, 2013; Read & Dang, 2022). In the literature of vocabulary knowledge, a handful of conceptualizations have been suggested, amongst the famous is drawing a borderline between vocabulary size and depth.

Anderson and Freebody (1981) first introduced the concept of depth as one brand of vocabulary knowledge that seems to be enormously different from breadth. It denotes the extent to which a learner knows a word, whereas vocabulary breadth refers to how many (the number) words a learner knows (Anderson & Freebody, 1981; Read, 1989; 2000; Qian, 1998; 1999; Qian & Schedl, 2004; Webb, 2013; Schmitt, 2014). In this dichotomy, breadth definition sounds quite accurate since it accounts for the relationship between form and meaning in a numeric sense as illustrated in the case of Vocabulary Levels Tests (Nation, 1990).

Similarly, when considering the previous definition of deep word knowledge (as equated to depth of knowledge), it seems a bit clear, but when its aspects are taken up it sounds rather superficial and in this case its definition differs and becomes much fuzzier. Therefore, depth of vocabulary knowledge has no unitary conceptualization among researchers (Webb, 2013) because of the incremental nature of vocabulary and the various aspects involved in depth of knowledge. These together made a convention on depth impossible. Despite this lack of consensus, many attempts have been made to identify depth of vocabulary knowledge along with its various components.

In the literature of vocabulary assessment, one common way to conceive of depth is having a relative command on the various facets of knowledge of vocabulary. A rampant view of depth lies in dividing it into distinctive features entailing word use (Schmitt, 2014). Knowledge of a word is rather complex and multi-componential (Nation, 1990; 2001; 2013). Its multidimensional aspects go far beyond word meanings, a one-dimensional component of breadth, to involve semantic relationships, collocations and syntactic behaviors of a word. (Bardakçı, 2016)

Long ago, Richards (1976) provided an initial framework on depth of vocabulary knowledge, set up in eight assumptions of different kinds of lexical knowledge a native speaker should develop. In his proposal, he stresses that word learning is an ongoing process in which he itemized the several facets of vocabulary knowledge that are critical to have a full mastery of a lexical item. These facets involve the various types of knowledge combined gradually in the mental lexicon during a language user's life.

More precisely, Richards' insight encapsulates numerous features of knowing a word. These briefly comprise recognition of spoken and written

forms and their connected words arranged in lexical sets and semantic relationships, appropriate use of words respecting both structural and situational patterns, and knowledge of word formation processes including derivatives and inflections as well as semantic connections (associations) a word might have with other words in a sentence.

Richard's approach emphasized the complex and multi-faceted nature of vocabulary knowledge (Read, 2000) because it involved more word properties like register and frequency level besides meaning recognition of word forms that represent a foundational stone of learners' lexical competence.

In her study on depth of vocabulary knowledge, Henriksen (1999) partitions vocabulary development into breadth, depth, and receptive and productive knowledge. Henriksen's notion of deep word knowledge is generally equated to Richards' basic assumptions of word knowledge and to that of Nation's (1990) components involved in knowing a lexical unit. Henriksen's view indicates that knowing a lexical item entails various aspects. She affirms that deep word knowledge, or what she terms "rich meaning" (Henriksen, 1999, p. 306) calls for meaning recognition in addition to other paradigmatic (synonymy, hyponymy), syntagmatic (collocations) sense relations, syntactic patterning and morphological features of a word.

Following this path, Read (2000) asserts that Henriksen's theoretical conception of depth seems to represent the best conceptualization of quality of word knowledge particularly in that it provides the specification of what construct dimensions a researcher on vocabulary should measure in his/her particular studies. Yet, he suggests a distinct conceptualization of depth in L2 vocabulary acquisition.

Read (2000) categorized vocabulary depth under three-faceted perspectives: precision of meaning, comprehensive word knowledge, and network knowledge. His influential proposition formulating the conception of depth of vocabulary knowledge includes three major approaches:

1. *Precision of meaning*: Read's precision of meaning is equated to Anderson and Freebody's (1981) definition of depth. In this first conception, Read (2000) distinguishes types of knowledge of meaning a learner has; a learners' meaning recognition ranges from having narrow or imprecise thought of a word to a more elaborated and precise understanding of its meaning.

To illustrate, polysemous words are inherently vague in that they carry various meanings to the extent that a learner may have a limited or general idea about especially when put out of context. The word 'book', for example, is a contextually dependent term that corresponds to a set of printed pages in one situation, and making a reservation in hotel in another. One more example would consider specialized or technical meanings of which a learner has accurate and advanced recognition. A learner enrolled in specialized linguistics courses precisely recognizes words like morpheme, paradigm, and syntagm; and a learner of literature would perfectly know words like prosody, simile, and pitch.

2. *Comprehensive word knowledge*: Though componential features of

word knowledge have long been comprehensively configured in Cronbach's (1942) and Richards' (1976) frameworks, this approach is paralleled to Nation's (2001) conception of what is involved in knowing a word. In this sense, comprehensive word knowledge requires more than grasping the surface meanings. It surpasses to incorporate further constituents: First, and foremost, word forms that are related to awareness of a words' spelling, pronunciation, and word parts. Second, word meanings that are associated with form and meaning relationships, concept and referents, and associations, and finally, word use that corresponds to learners' knowledge of grammatical functions, collocations, constraints on use i.e. register as well as frequency.

As far as assessment is concerned, Read highlights that researchers' attempt to measure all the comprehensive word knowledge facets yields in completely perplexed test design, because such facets involve multi-sub-knowledge features a researcher cannot have control over during test development.

3. *Network knowledge*: It represents a network approach which is based on the assumption that when a learner's vocabulary size increases, a learner incorporates newly learned words into a connected lexical network settled in the mental lexicon. An L2 learner develops the ability to accumulate the freshly attained knowledge of vocabulary to his/her prior knowledge of vocabulary, and recognizes whether linguistic items are semantically different or related to other words and how to connect these words in a language. This dimension is commonly likened to Meara's (1996) 'lexical organization' and Henriksen's (1999) 'depth of knowledge'. Subsequently, it is possible to theorize depth of vocabulary knowledge from the standpoint of the way words are stored in the mental lexicon in an organized structure by means of processes of semantic mapping, collocation, synonymy, super-ordination, and co-ordination, which, in turn, are considered aspects of depth of knowledge in the literature of vocabulary assessment.

A similar way of conceiving depth is promoted by Meara (1996). The researcher describes depth as a network of lexical items greatly structured in the mental lexicon. Instead of depth, he favors the term lexical organization. In other words, he stressed on how words are processed in the mind of individuals and not on the quality or the quantity of word knowledge. Lexical organization can be measured by means of assessing learners' ability to identify or give a large set of associations; however, depth conceived as lexical organization is considered as one measure of vocabulary breadth. (Schmitt, 2014)

The abovementioned three dimensions indicate that knowing a word involves many aspects and these aspects provide clear guidelines for vocabulary measurement but "each one is limited in the degree to which depth is measured, and overlap between the categories suggests that together they may still not provide an accurate assessment of depth" (Webb, 2013, p. 3). In a similar vein, Nation (2001) further elaborated on the above three aspects of depth in his famous table that summarizes several levels of word mastery.

Nation's (2013) hierarchical table subdivides the three main elements of depth of vocabulary knowledge (form, meaning, and use) into further constituents; and each of which is composed of three other sub-component aspects, and each aspect is further divided into receptive and productive knowledge categories. (Webb, 2013; Yanagisawa & Webb, 2020)

Yanagisawa and Webb (2020) confirm that depth of vocabulary knowledge is identified by the extent to which Nation's (2001, 2013) eighteenth (18) aspects involved in knowing a word are present, and these in turn, would indicate whether a lexical item is fully mastered or not.

Nation (2001; 2013) asserts that vocabulary depth comprises three basic components of form, meaning, and use. Knowledge of form primarily entails how well learners can recognize how a new word is spelled whenever seen and how is pronounced whenever heard. Within this sort of knowledge are word parts; learners can distinguish, for instance, between listen, listener, listening when met in a reading passage or listened to in a spoken discourse. Knowledge of meaning alludes to comprehending a word's associated synonyms, antonyms, connotations, and polysemy (words having more than one meaning). Knowledge of word use refers to learners' ability to correctly use a word in a given context without hindering the lexical, syntactical, semantic, and collocational relations.

Nation indicates that features involved in knowing a word underpin both receptive and productive use ability. Possibly, the best description of productive and receptive vocabulary based on Nation's (2001) framework is set up forth by Schmitt (2014). The author delineates receptive (or passive) vocabulary as learners' ability to recognize words when exposed to listening and reading tasks, and views productive (or active) vocabulary as their ability to use words in spoken and written discourse contexts.

Qian (1999) has previously conceptualized depth dimension from a similar perspective assimilating syntactic, semantic, morphemic and graphemic characteristics of word knowledge.

The aforementioned conceptualizations indicate how depth is a controversial and multifarious theoretical concept and due to this fact it is awfully tough to decide how to approach it from a theoretic outlook (Schmitt, 2014). However, the most elaborated descriptive framework of what is involved in deep word knowledge is that of Nation (Read, 2004; Daller, et al., 2007; Gyllstad, 2013; Webb, 2013). In sum, knowing a word does not require only knowing its form-meaning connection rather implies its semantic properties, its orthographic (spelling), phonological (pronunciation), morphological (word parts), syntactic (grammatical functions), collocational (word associations and the frequent co-occurrence of words in a linear string of elements in a sentence), and pragmatic (register) characteristics.

Now that depths of vocabulary knowledge have been elicited, its measurement is to be discussed in subsequent sections.

3. Approaches to Measuring Depth of Vocabulary Knowledge:

In the research literature, a number of attempts have been made to assess the quality of vocabulary knowledge. However, when compared to vocabulary breadth measures, depth measures are relatively few (Mukarto, 2005; Ebrahimi, 2017; Schmitt, 2014; Edmonds, Clenton, & Elmetaher, 2022). Arguably, breadth tests are biased because they are easy to administer and can cover large sets of words in a relatively short period of time as they need a single response associated with every lexical unit (Read, 1993). Eventhough, depth of vocabulary has received little attention, it has been approached differently and well-designed tests have been developed throughout the research endeavor.

Yanagisawa and Webb (2020) distinguish three approaches to measuring deep word knowledge: a developmental approach, a lexical network approach, and a components approach. Espousal of these methodologies diverges liable to how vocabulary depth is conceived; whether is operationalized as degrees of knowledge, as lexical mental connections, or as various sub-knowledge aspects.

The *Vocabulary Knowledge Scale* (Wesche & Paribakht, 1996) is the famous well-established assessment system of word knowledge profundity trendy in the developmental approach, the *Word Association procedure* (Read, 1989; 1993) and the lexical network test or *V-Links* (Meara, 1996; 2009) are popular in the lexical network approach, and the *Interview* test (Schmitt, 1998) is inherent in the components approach. In this subsequent section, we review and discuss these measures of depth together with others that have been taken up by previous research along the three approaches.

3.1. Developmental Approach:

In the developmental approach, developmental scales are utilized to identify the phases of mastery of a lexical item and, as such, it embodies the incremental nature of vocabulary learning. For Read (1993), it is due to the fact that vocabulary knowledge is frequently a gradual process that even native speakers of a language have an imprecise knowledge of meaning of words they know. Arguably, many researchers, when measuring vocabulary knowledge, devised different scales to represent different levels of word knowledge.

The scales range from no knowledge labelled as 0 to complete mastery symbolized as 5. Vocabulary depth is described in terms of levels of knowledge because learners go through developmental stages during their process of vocabulary acquisition. Succinctly, their knowledge of lexis grows gradually moving from wide-ranging word exoticism (unfamiliarity), through identification of a word's spoken and written forms and having superficial meaning of it, to the capacity to use the targeted word in well-formed sentences fittingly to the syntactical and semantic relationships (Wesche & Paribakht, 1996). This insight is congruous with Dale's (1965) earlier definition of depth of knowledge. The pioneering and widely used work

adopting this predisposition is demonstrated in the Vocabulary Knowledge Scale test (VKS) established by Paribakht & Wesche (1993).

The developmental approach implements only one assessment instrument, that of the VKS, which is also known as self-report of degrees of knowledge (Read, 2004). This test battery is a measure of productive as well as receptive vocabulary depth since it contains two different levels of scales; one involves learners to use particular words in a sentence and others require no creative use of the word, thus, tapping just the passive knowledge (word recognition).

The VKS has been described as a generic measure in that it can be applied to assess mastery of any set of words (Wesche & Paribakht, 1996; Read, 2000). It uses five scales; four are rather similar to Dale's (1965) in addition to a fifth stage under the statement 'I can use this word in a sentence' (i.e., besides recognizing a word's meaning and its appropriate place in context, a learner can further use it in sentence context).

Initially the scales chart four levels of word mastery especially developed by (Dale, 1965) who states that the incremental nature of vocabulary knowledge is due to four stages of vocabulary progress. He postulated the presence of no less than four progressive phases of word knowledge whereupon Wesche and Paribakht (1996) elaborated their most sophisticated five-point self-report elicitation scale as reflected in the research literature.

These scales indicate that learners' word knowledge include a variety of degrees of knowing ranging from no acquaintance due to poor exposure to words to fully knowing how a word is spelled but not knowing its meaning. Followed by certain level of guessing meaning of words, perused by the ability to provide a words' synonym or translation, completed by the ability to use a word in a sentence.

The five scales have been adopted as a measure in vocabulary assessment, especially assessing the quality of vocabulary knowledge. It is highly important to point out that the major difference between Dale's and Wesche and Paribakht's scales is that the former designed his scale for first language learners whereas the latter developed it for second language users.

The test was designed in the ESL (English as a second language) vocabulary development assessment paradigm. It was produced to demonstrate how receptive knowledge of some words would change to an initial productive knowledge over time. This growth or changes occur as a result of vocabulary instruction through reading and other activities.

The VKS has received great interest (Schmitt, 1998), nevertheless, it has been criticized for being time consuming research tool, eventhough it tests only a limited number of lexical items, as it tends to elicit spoken and written explanations of words from the part of learners (Read, 1993). One more limitation opposed to this scale is being a discrete point test that measures words out of larger communicative contexts despite the fact that one component of the scale measures how words are used within sentence boundaries (context).

3. 2. Lexical Network Approach:

This approach views deep word knowledge as all-encompassing organized lexical sets or networks in L2 learners' mental lexicon. This conception reflects Read's (2004) notion of deep word knowledge, the network knowledge in particular (the third type). Studies adopting this conceptualization tend to investigate the degree to which words are related to each other (Read, 1993; Meara, 1996; Henriksen, 1999). Successful developed measures adhering to this approach are the *Word Association Test* (Read, 1993) and the *Vocabulary/lexical networks* (Meara, 1996).

The Word Association Test (WAT) was first proposed by Read (1993) to receptively measure how well particular lexical items are well known. It is claimed that learners who are able to track the associations seem to have deeper lexical knowledge and those who are not have quite superficial knowledge. It is typically a test of lexical network where learners are required to recognize or supplement a number of associations. In this sense, word associations refer to "the links that connect or relate words in some manner in a person's mind." Schmitt & Meara (1997, p.19)

In this test format, test takers are instructed to identify how far words are related lexically or semantically to the prompt words that are jointly introduced with other four connected words, some of which are lexically or semantically related to the stimulus and some do not and act as distractors which either resemble stimulus words orthographically or morphologically. The test aims to elicit three types of semantic networks existing between the words; whether the lexical items are connected syntagmatically (collocations); paradigmatically (synonyms); or analytically (the associate partially represents the meaning of the stimulus word).

Example: denominator (stimulus word) *common develop divide eloquent fraction mathematics species western.* (Read, 1993, p. 366)
Here the participants are asked to pick out four words that are associates; either lexically or semantically related to, the stimulus word.

Overall, the test checks how deeply learners know 40 target words, including nouns, verbs and adjectives, by selecting four words out of eight words associated with the stimulus word. One advantage of this procedure is that it proved to be a reliable measure of depth of vocabulary knowledge. Its major limitation lies in being merely discrete and definitely receptive. It measures items in isolation and examinees are required to select correct answers from odd words instead of creating or supplying their own associates.

It has been mentioned in the previous section that there are two distinct perspectives to depth of vocabulary knowledge, one of which conceptualizes depth as a set of sub-knowledge components and the other, however, prefers the term organization instead. The latter approach typically refers to Meara (1996) who has introduced his idea of organization to mean that strong structured mental networks of lexical items. His approach to vocabulary testing is developed out of, and as a reaction to, vocabulary size and depth tests leading to purely systematized lexical networks assessment format.

The stimulus for this approach entails that testing knowledge of a large number of words needs extensively developed measures for testing individual words, and even including all aspects of word knowledge to depict how well 50 words, for instance, are known makes the process of assessment impossible and even unreliable as it requires designing dozen of subtests (600 item test). An alternative approach to these is to produce a single test that can assess depth meaningfully. Meara, thus, designed his most well-known *V. Links*, a test battery to measure lexical organization in 2009. It is a 20 items test that are randomly selected from the basic 1000 words in English core vocabulary and each of which comprises a selection of 10 target items. It is instantiated in a software program in which subjects are asked to recognize any sort of association pairs and show how far the association is strong on a four-point scale.

According to Nation (2014), this research endeavor is marginalized and still sophisticated. He, thus, emphasizes that large scale research is recommended to further investigate whether this approach can be optimized and widely realized.

Though it proved to be an effective measure of lexical organization as it extensively tests large sets of words (a total of 200 words) in a relatively short period of time (30 minutes), it neglected the basic communicative aspect of language because it tests words out of context.

3.3. Components/Dimensions Approach:

The dimensional approach views depth of vocabulary knowledge as comprising various aspects. The componential approach conception of vocabulary depth replicates that of Cronbach's (1942), Richards' (1976), Henriksen's (1999) and Read's (2004) comprehensive word knowledge, the second dimension of depth construct. Nation's (2013) description of vocabulary knowledge is the most comprehensive and extensively quoted reference cataloguing the different dimensions (Webb, 2013; Yanagisawa & Webb, 2020). Given that vocabulary knowledge is multidimensional, there is no such test battery involving all the aspects described by Nation and, hence, various assessment formats have been constructed.

The most influential tools used to assess vocabulary depth in the research literature involve interviews (Schmitt, 1998), *idioms method*, *Context-Dependent Collocation* (Milton, 2009), and procedures for *Meaningful Sentences* (Sharakhimov & Nurmukhamedov, 2021), and *Depth Tests* (Read & Dang, 2022).

In a longitudinal study, Schmitt (1998) produced an interview procedure to assess the quality and growth of learners' vocabulary knowledge. The study aim was to track the acquisition of 11 words of three proficient adult university FL learners coming from different backgrounds over a course of a year. Four brands of word knowledge, specifically spelling, connotations, syntactical and semantic relations have been assessed after test takers were exposed to the words during a course of study. Target words were selected from UWL (University Word List), they were all polysemous having more than three

senses to raise learners awareness of various senses. Schmitt designed a test of written form (spelling) and association measurement procedure that required learners to provide 3 responses matching the prompt word.

Eventhough, it enables vocabulary researchers to tap extended information needed to measure depth of lexical knowledge, this interview instrument is criticized for some pitfalls as it is time consuming and this, in turn, limits the number of target words as well as research participants. Furthermore, it limits the scope of contextualization to the sentence level instead of large situational communicative context of language use.

Among the various ways of assessing vocabulary production, Milton (2009) also proposes that the measures of *idiom knowledge* assess learners' knowledge of fixed English idioms. In this test format, test takers are required to supply one word for each gap in the 20 sampled idioms. Despite the fact that idioms are part of depth of vocabulary knowledge, Milton questioned their usefulness in testing L2 knowledge as they are likely to be rarely used in the actual native language usage because they are highly infrequent in normal speech.

Sharakhimov and Nurmukhamedov (2021) propose two other insightful tests for assessing aspects of word knowledge. These are *Context-Dependent Collocation* and *Meaningful Sentences* procedures. The former endeavors to assess test takers' ability to produce collocations. Learners are asked to produce two or more word combinations of words that frequently co-occur. Collocations are critical to vocabulary acquisition since learners are often not able to appropriately use words because they do not know the right collocates eventhough they recognize a word's meaning. (Webb & Sasao, 2013)

The second productive vocabulary depth procedure focuses on word use. Learners are instructed to produce meaningful utterances using target words. Sharakhimov and Nurmukhamedov (2021) suggest that this test format can be used in conversations during listening and speaking classes. This proposal is practically insightful and inspiring, but to the best knowledge of the writers, it has not yet been applied to prove its validity and reliability to the assessment of depth of productive vocabulary knowledge. Nevertheless, given that the *Context-Dependent Collocation* and *Meaningful Sentences* procedure focus on using words to construct word combinations like phrases, and to produce meaningful sentences respectively prevent depth of vocabulary from integrative, communicative, situational and contextualized assessment.

More recently Read and Dang (2022) developed a *Depth Test* meant to assess the degree to which examinees mastered some target words. It measures knowledge of synonyms, collocations, and word parts of known words. The test consists of three parts: Part A is intended to measure if leaners can match the target words with a synonym or short definition. It uses simple true/false and multiple-choice procedures. It measures knowledge of 30 words through presenting each target word with two synonyms and distractors. Part B, devoted to testing collocations, involves 30 whole phrases

and short sentences to assess knowledge of collocations using yes/no and not sure options in order to indicate correctness of word combinations. Part C requires examinees to find out the correct word form of 20 words using both inflected and derived affixes.

Despite being arguably reliable and insightful, these three depth tests do not consider context as an issue, which is important to elicit students' ability to communicate effectively via writing or speaking. They are, therefore, selected response rather than constructed response formats.

4. Assessing Word Use in Context:

All the aforementioned types of vocabulary depth tests that have been further discussed are, in Read's (2000) terms, discrete because they introduce target words in isolation regardless of context. They preliminarily eliminate the communicative aspect of language use. From the perspective of teaching and learning, it is quite ascertained that locating words within various discourse contexts is prerequisite to effective vocabulary instruction and development (Stahl & Bravo, 2010). Schmitt (2014) contends that learning words out of context may be trivial; it stimulates only knowledge of form-meaning linkage that would facilitate some words retention and, thus, having large vocabulary size but when coming to word use learners may not be able to use them in different settings.

On the importance of context in enhancing learners' quality of word knowledge, Webb (2013) states that context is a predicator of vocabulary depth development since it provides learners with opportunities to synthesize their knowledge via exposure to unknown words and using them in different contextual patterns. Webb further indicates that levels of knowledge in multiple situations would elicit learners' ability to learn to recognize the orthographical form of the word, the phonological distribution (pronunciation) of the word, the morphological traits (derivational and inflectional affixes), the semantic features of the word (e.g., sense relations, collocates and word associates), and, above all, the syntactical roles a word may play in a sentence. A full mastery of words is, therefore, apt to the application of previously discussed features fitting to the contextual use.

Following this line of thought, Ozturk (2012) sought to examine the impact of context on learners' achievement in a vocabulary test and assess the extent to which taught words have been mastered. To this end, he used two multiple choice tests; one is discrete (d-test) in which lexical items are presented in sentence context, the other is contextualized (c-test) where target words are integrated within meaningful texts or discourses to provide contextual clues for learners to determine meanings of words. Research findings indicated that students' performance was better in the contextualized test.

From an assessment standpoint, it is largely emphasized that vocabulary should be assessed in context (Read, 2000) because learners need stimulating contextual information to generate suitable responses (Read & Chapelle, 2001). In his framework for SL vocabulary assessment, Read (2000; 2007)

proposes two opposing viewpoints: one perspective suggests that lexis can be tested as a semantic field regardless of context (e.g., Yes/No and word associates formats). The second view assumes that words must not be tested out of their linguistic context but rather in embedded, comprehensive and context-dependent tests that measure vocabulary knowledge and use via different types of listening, reading (contextual understanding measure), speaking and writing activities (contextual use measure). In this context, performance tasks and short constructed response formats can be proposed to assess how well learners can use target words to solve communicative situational problems. i.e., in a controlled vocabulary production, for instance, learners can be asked to use target words (e.g. adjectives, verbs, and nouns) to write a chart of ethics for managers of companies adhering to child labor in a well-structured paragraph.

In other contexts, Read (1993) argues that contemporary language proficiency testing is in favor of integrative and communicative test formats. It advocates that words should be contextualized either in sentence context or in situational context (a larger discourse unit) as it is crucial for learners to recognize words in connected written or spoken discourse other than as decontextualized separate items. He adds that word contextualization recommends a reading or listening comprehension passage within the vocabulary test and this itself decreases the amount of words that can be tested. Contextualization, therefore, must be one aspect of vocabulary measurement. For Schmitt (2014) various word knowledge aspects such as collocations, register constraints, and frequency are extremely tied to vocabulary in contextualized use as they may carry distinct usage contexts.

Context dependence measure is favored in communicative teaching paradigm where language is taught and assessed in meaningful contextualized and situationalized settings. A measure on how well or how deeply words are learned in a communicative context can be established. Initiative insights of this approach can be traced back up to the 1980s with the introduction of Dell Hymes' communicative competence. The notion that gained a great monopoly over education and language teaching and learning led researchers to question the issue of test reliability and investigate learners' lexical competence, as part and parcel of communicative competence, and vocabulary knowledge and use henceforth.

To the best knowledge of the writers, the literature that have sought to investigate productive vocabulary use in context, as an aspect of depth, is lacking except for measures of lexical richness that tend to assess productive vocabulary use of texts produced by learners for lexical size estimation aims. One type of these methods is the lexical density test that refers to the proportion of content words as opposed to function words in a text. (Lindqvist, Gudmundson, & Bardel, 2013). The lexical diversity, or variation measure (McCarthy & Jarvis, 2007) is often designed to calculate and analyze types of vocabulary, high or low frequency vocabulary, produced by students in print and verbal scripts.

Measures of lexical sophistication (Fritz & Ruegg, 2013; Lindqvist, Gudmundson & Bardel, 2013; Kyle & Crossley, 2014) count the proportion of advanced words in a text. The same does hold true for the proportion of errors in a text. Other measures of lexical richness include the lexical frequency profile (Cobb, 2006) which is a program produced to estimate the proportion of high and low frequency words or simply to measure free productive vocabulary of written compositions or oral texts produced by learners.

As stated by Nation (2014), the application of lexical variation and the lexical frequency profile methods proved to be effectual in that they are reliable estimates of written productive knowledge and lexical diversity. Commenting on the lexical diversity assessed in Engber's (1995) research, Baba (2009) ascertains that it is a breadth of lexical proficiency, and in the meanwhile, it is productive. Overall, these studies conceptualize vocabulary depth as fluency because of the fact that they examine the effect of lexical proficiency in L2 writing. That is the extent to which lexical proficiency determines writing quality.

They take word use and correctness (spelling) as a measure of lexical sophistication, which is part of fluency or lexical proficiency but not as an aspect of depth since they aim at analyzing written and oral texts to provide estimates of learners total vocabulary threshold, and not to test how deeply certain lexical aspects are being grasped.

In a nutshell, provided that the productive vocabulary depth tests discussed so far do not consider word use in larger communicative contexts, this article stresses that measuring vocabulary use in written and spoken discourse contexts is a neglected aspect in vocabulary assessment. Productive vocabulary use in context is, therefore, further recommended as a measure of vocabulary depth whereby researchers can approach various components of lexical knowledge (e.g. word meaning, word parts, and word use) in a single contextualized integrative setting within the context of performance-based assessment and task-based paradigms. In these mainstreams, the main objective of assessment is to assess learners' ability to use previously acquired vocabulary to solve new real word problems, and the above formats used for diagnostic purposes do not fit the communicative endeavor.

5. CONCLUSION

In this article an-in depth vocabulary knowledge construct have been surveyed with its significant existing measures particularly used for its assessment. Eventhough these lexical measures have been criticized for their pitfalls in assessing how well lexical items are mastered in contextualized word use settings, they are still valid and influential instruments as discrete tests whereby learners' lexical quality and lexical organization can be elicited in various ways.

Each test is designed to measure a given aspect of vocabulary knowledge, and thus it is subject to the purpose it is designed to measure and cannot be adopted to assess word use in context because they represent target words discretely out of context. Most of tests designed for assessing depth

treat lexical items as decontextualized isolated points, and the literature on assessing lexical use in context is lacking.

When considering the need to develop more accurate tests for assessing word use in context, it is proposed to design such measures as authentic-based language guided writing tasks where learners can show their ability to use target words, as part of depth of knowledge, with their corresponding word forms (word parts) and meaning appropriateness. It is worth mentioning here that by context is meant not sentence context but larger contexts like paragraphs and discourses where learners can effectively use target words in more communicative authentic tasks.

In more traditional approaches, assessing language aspects in decontextualized test items sounds obvious whereas in more recent times when following the current approaches to language teaching and learning mainstream, it becomes a requirement to assess vocabulary in the context of productive use (performance contextualized tasks). In this context, it is recommended that lexis, being part and parcel of communicative competence, must be taught (and be assessed henceforth) in context to assess learners' lexical competence from the perspective of their ability to use target words in new situations corresponding to the stimulating authentic contextual information provided.

5. Bibliography List:

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. Dans J. T. Guthrie, *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Anthony, L. (2020). Resources for researching vocabulary. Dans S. Webb, *The Routledge handbook of vocabulary studies* (pp. 561-591). Routledge.
- Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, 18, 191-208.
- Bardakçı, M. (2016). Breadth and depth of vocabulary knowledge and their effects on L2 vocabulary profiles. *English Language Teaching*, 9(4), 239-250.
- Cobb, T. (2006). *VocabProfiler*. Consulté le 3 15, 2020, sur <http://www.lex tutor.ca/vp/>
- Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research*, 36(3), 206-217.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, 42, 895-901.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Ebrahimi, A. B. (2017). Measuring productive depth of vocabulary knowledge of the most frequent words. (T. U. Western, Éd.) Ontario, Canada. Consulté le 8 2021, sur Measuring productive depth of vocabulary knowledge of the most frequent words: <https://ir.lib.uwo.ca/etd/4894/>
- Edmonds, A., Clenton, J., & Elmetaher, H. (2022). Exploring the construct validity of tests used to assess L2 productive vocabulary knowledge. *System*, 1-38.
- Fountain, R. L., & Nation, I. S. (2000). A vocabulary-based graded dictation test. *RELC Journal*, 32(2), 29-44.
- Fritz, E., & Ruegg, R. (2013). Rater sensitivity to lexical accuracy, sophistication and range when assessing writing. *Assessing Writing*, 18, 173-181.
- Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective: Challenges and potential solutions. Dans C. Bardel, C. Lindquist, & L. Laufer, *L2 vocabulary acquisition, knowledge and use* (pp. 11-28). European Second Language Association: Eurosla Monographs Series.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in second language acquisition*, 21(2), 303-317.
- Kyle, K., & Crossley, S. A. (2014). Automatically assessing lexical sophistication: Indices, tools, findings, and application. , 1-30. *TESOL QUARTERLY*, 49(4), 757-786.
- Laufer, B., & Nation, I. S. (1999). A vocabulary size test of controlled productive vocabulary. *Language Testing*, 16(1), 33-51.
- Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). A new approach to measuring lexical sophistication in L2 oral production. *Eurosla Monographs Series*, 2, 109-126.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459-488.
- Meara. (1996). The dimensions of lexical competence. Dans G. Brown, K. Malmkjaer, & J. Williams, *Performance and competence in second language acquisition* (pp. 35-53). Cambridge University Press.
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins.

- Meara, P. M., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28, 19-30.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 142(4), 142-154.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.
- Milton, J., & Hopkins, N. (2005). *AuralLex*. Swansea, Wales: Swansea University.
- Miralpeix, I. (2020). L1 and L2 vocabulary size and growth. Dans S. Webb, *The Routledge handbook of vocabulary studies* (pp. 189-206). Routledge.
- Mukarto, F. X. (2005). Assessing the depth of second language vocabulary knowledge. *Physics*, 8(3), 151-169.
- Nation, I. S. (1990). *Teaching and learning vocabulary*. Boston: Heinle and Heinle.
- Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. (2013). *Learning vocabulary in another language* (éd. 2). Cambridge: Cambridge University Press.
- Nation, I. S. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), 1-16.
- Ozturk, G. (2012). The effect of context in achievement vocabulary tests. *Journal of Educational and Instructional Studies in the World*, 2(4), 126-134.
- Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 9-29.
- Qian, D. D. (1998). Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language. *Unpublished Doctoral Thesis*. University of Toronto. Récupéré sur <https://tspace.library.utoronto.ca/handle/1807/12079>
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56, 282-308.
- Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21, 28-52.
- Read, J. (1989). *Towards a deeper assessment of vocabulary knowledge*. Washington, DC: ERIC Clearing House on Languages and Linguistics.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? . Dans P. Bogaards, & B. Laufer, *Vocabulary in a second language* (pp. 209-227). Amsterdam: Benjamins.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7(2), 105-125.
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1-32.
- Read, J., & Dang, T. (2022). Measuring depth of academic vocabulary knowledge. *Language Teaching Research*, 1362-1688 .
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.

- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48, 281-317.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913-951.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *SSLA*, 20, 17-36.
- Sharakhimov, S., & Nurmukhamedov, U. (2021). Assessing learners' productive vocabulary knowledge: Formats and considerations. *English Teaching Forum*, 59(4), 16-25.
- Stahl, K. A., & Bravo, M. A. (2010). Contemporary classroom vocabulary assessment for content areas. *The Reading Teacher*, 63(7), 566-578.
- Thékes, I. (2018). *Studies in the assessment of foreign language vocabulary*. Gál Ferenc College: Gerhardus Publishing.
- Webb, S. (2013). *Depth of vocabulary knowledge*. Consulté le 4 12, 2021, sur Wiley Online Library: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405198431.wbeal1325>
- Webb, S. (2020). *The Routledge handbook of vocabulary studies*. Routledge.
- Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, 44(3), 263–277.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. Dans S. Webb, *The Routledge handbook of vocabulary studies* (pp. 371-386). Routledge.