Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta Polelo - Buka ya Thuto ya Puo - Jenale ya Thuto ya Dipuo - Ijenali Yekufundzisa Lulwimi - Jena?a ya u Gudisa Nyambo - Jenala yo Dyondzisa Ririmi - - Tydskrif vir Taalonderrig - Journal for Language Teaching - Ijenali yokuFundisa iLimi - IJenali yokuFundisa iiLwimi - Ibhuku Lokufundisa Ulimi - Tšenale ya tša Go ruta

Déogratias Nizonkiza

and

Themba Ngwenya

North-West University (Potchefstroom Campus)
and
North-West University (Mafikeng Campus)

Challenges of testing deep word knowledge of vocabulary: Which path to follow?

Abstract

The present research explores the challenges of testing deep word knowledge of the vocabulary of students of English as a Foreign/ Second Language (EFL/ESL) at higher education. A productive test modelled on the Lex30 test developed by Meara and Fitzpatrick (2000) was presented to the participants. Results indicate that (i) ESL students outperform their EFL counterparts of comparable class level, (ii) aspects of deep word knowledge among both higher education EFL and ESL students develop in the order of analytic relations, paradigmatic relations, and collocations; and (iii) aspects of deep word knowledge among both higher education EFL and ESL students grow alongside one another and correlate significantly with overall deep word knowledge; the strength of which may reflect the extent to which they contribute to it. On the basis of these findings, we conclude that Lex30 may constitute a good measure of deep word knowledge at the productive level, which is the path we suggest should be followed in order to test deep word knowledge of vocabulary at higher education. Furthermore, teaching implications aimed to foster deep word knowledge growth are discussed.

Key words: vocabulary deep word knowledge, testing deep word knowledge productively, vocabulary dimensions, English as a Foreign/Second Language (EFL/ESL)

1.   Introduction

Over the past few years, increased attention has been paid to research regarding vocabulary, the aim being to inform the learning and teaching of vocabulary (Hirsh, 2010). The available literature indicates that vocabulary knowledge can be classified into three dimensions, that is vocabulary size, also known in the literature as vocabulary breadth; deep word knowledge, also known as vocabulary depth; and receptive-productive dimensions. Vocabulary size refers to the number of words known (understood) by someone (cf. Henriksen, 1999; Meara, 1996; Read, 1993). Deep word knowledge refers to how well words are known (Gairns and Redman, 1986; Laufer and Paribakht, 1998; Van de Poel and Swanepoel, 2003; Zareva, Schwanenflugel and Nikolova, 2005). The receptive-productive dimension makes a distinction between receptive and productive knowledge. This distinction implies that words are understood receptively (receptive knowledge) before being productively used (productive knowledge) (Gairns and Redman, 1986; Laufer and Paribakht, 1998; Van de Poel and Swanepoel, 2003; Zareva et al., 2005).

As the available literature indicates, more research has been conducted with regard to vocabulary size than deep word knowledge (Ishii and Schmitt, 2009; Read, 2007; Schmitt, Ng and Garras, 2011). Extensive research into vocabulary size has resulted in the establishment of two principal pedagogical practices, one of which is determining how much vocabulary is needed at each learning stage (Nizonkiza and Van den Berg, 2014). Research evidence has indeed indicated that a minimal threshold of about 5,000 word families and an optimal one of approximately 8,000 word families are needed for understanding lectures at the undergraduate level (Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011). Employing vocabulary size tests for placement purposes is the other pedagogical practice that has developed from research into vocabulary size. Nation's Vocabulary Levels Test is the most widely used test to this end (see Read, 2007 among others).

However, as pointed out by Ishii and Schmitt (2009), Read (1993), Schmitt et al. (2011), Wesche and Paribakht (1996), among others, vocabulary knowledge should not be conceptualised only as the number of words a person knows (vocabulary size). It should also be regarded as the extent to which these words are understood in terms of depth, that is, how well the words are known. According to Read (1993: 359), how well a word is known entails knowing three aspects associated with it, namely paradigmatic, analytic, and syntagmatic relations. A paradigmatic relation implies that the words are related because they are synonyms or close in meaning with one being more general than the other; for example, *edit – revise; team – group*. Analytic relation means that the two words are related because one is a key word of the dictionary definition of the other; for example, *edit – publishing; team – together*. As far as syntagmatic relation is concerned, it refers to collocations, which means that the words are related because they often co-occur in a sentence; for example, *edit – film; team – scientists*.

Vocabulary depth has been measured mainly through Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) and Read's (1993) Word Associates Test (WAT) (see section 2.4). Initially the VKS was intended to test both receptive and productive knowledge, but it seems that the test did not succeed in this attempt. The WAT measures receptive knowledge by means of association tasks. Both the VKS and the WAT have been validated as measures of depth knowledge and have been proven to correlate with overall linguistic proficiency. So far, however, none of these tests has been standardised. Therefore, Schmitt et al. (2011) are right in their observation that the research and pedagogical consequences associated with depth tests do not appear to result in wider applications.

The difficulties associated with testing depth could be attributed to the lack of a clear definition and construct of what depth entails (Batty, 2012; Milton, 2009; Read, 2000, 2004). For instance, while the traditional approach to depth did not specifically state that productive knowledge of vocabulary could form part of depth; a more recent approach suggests that productive knowledge could be an integral component of depth knowledge. According to Schmitt et al. (2011: 107), depth not only indicates the manner in which words are understood receptively, but also the extent to which they can be used "productively, and how appropriately". While we have a number of productive tests frequently used today (see section 2.3), interpreting scores from these tests hardly makes reference to the aspects of depth, productive or otherwise. Therefore, vocabulary depth is tested solely receptively. Consequently, we do not know the way the aspects of deep word knowledge (paradigmatic, analytic, collations) develop productively, whether or not they grow alongside one another, and the extent to which they contribute to overall depth. This is where the focus of the present study lies. Furthermore, EFL or ESL students from different backgrounds may be different. We believe that considering each situation may be insightful with regard to growth of aspects of depth and the challenges they may pose, which may benefit teaching practices. Studies we have so far have focused on just one group, but not on two ESL or EFL groups let alone comparing ESL and EFL groups[1].

2. Related literature

2.1. Vocabulary dimensions

Increasing interest in vocabulary over the past few decades has resulted in vocabulary knowledge being conceptualised in terms of dimensions. Two main dimensions, that is, size and depth (Henriksen, 1999; Meara, 1996) and a third, which certain scholars consider to be a bridge dimension (Van de Poel and Swanepoel, 2003; Zareva et al., 2005), have been identified. Vocabulary size is the first dimension of vocabulary knowledge, which consists of how many words someone knows, irrespective of how well the words are known (cf. Henriksen, 1999; Meara, 1996; Read, 1993, 2000).

---

1   We thank one of the anonymous reviewers for this suggestion.

225

Deep word knowledge is the second dimension and is concerned with aspects of deep word knowledge, that is, how well a word is known (Greidanus and Nienhuis, 2001; Henriksen, 1999; Meara, 1996; Qian and Schedl, 2004; Read, 1993, 2000; Vermeer, 2001; Wesche and Paribakht, 1996). Deep word knowledge consists of associates of a word at the paradigmatic (synonym or close in meaning), syntagmatic (collocation), and analytic (key word of dictionary definition) levels (as defined in the previous section) (Greidanus, Bogaards, Van der Linden, Nienhuis, and Dewolf, 2004; Henriksen, 1999; Meara, 1996; Read, 1993, 2000).

The third dimension, as proposed by Henriksen (1999), is the receptive-productive dimension, which is viewed as a continuum between word comprehension and word use. ESL and EFL researchers define receptive vocabulary as the vocabulary used for comprehension, while productive vocabulary is that used for production; definitions on which they tend to agree (cf. Zareva et al., 2005). The receptive–productive distinction is based on the premise that word comprehension does not necessarily imply its correct use (Gairns and Redman, 1986; Laufer and Paribakht, 1998; Van de Poel and Swanepoel, 2003; Zareva et al., 2005). The receptive–productive distinction should not, however, be viewed as a dichotomous one. It should rather be considered as a continuum where a word passes a threshold from being receptive to being productive; which is the most widely accepted conceptualisation of the development of the receptive–productive vocabulary (Aitchison, 1989; Melka, 1997).

2.2. The importance of vocabulary in ESL/EFL contexts

The importance of vocabulary as a component of language has been demonstrated since the 1990s, a period in which the topic gained popularity among ESL/EFL researchers (Read, 2000). Among other things, a relationship between knowing words and linguistic proficiency has been established. Research has indeed shown that the number of words students know could reliably predict their linguistic proficiency level (cf. Beglar, 2010; Meara and Buxton, 1987; Meara and Jones, 1988; Nation and Beglar, 2007; Schmitt, Schmitt and Clapham, 2001). A firm relationship has also been established between vocabulary knowledge and reading comprehension (Hu and Nation, 2000; Nation, 2006). This predictive power of vocabulary knowledge can also be extended to other language skills, such as listening, speaking, and writing. According to Milton and Treffers-Daller (2013), research evidence seems to suggest that vocabulary is a good predictor of performance on the four language skills.

We agree with Nizonkiza and Van den Berg (2014), Ishii and Schmitt (2009), Milton (2009), Read (2007), Schmitt (2008), Schmitt et al. (2011) that major pedagogical consequences arose from extensive research into the topic of vocabulary knowledge. Among other things, researchers have been able to determine the amount of vocabulary needed (for comprehension) at different learning stages, which allowed developing course materials aimed to foster vocabulary growth among ESL/EFL students. Vocabulary tests can also be used for placement purposes. However, as already pointed out (cf. section 1), this progress made is based on research findings from one of the

226

vocabulary dimensions (vocabulary size) which does not suffice to describe vocabulary knowledge. Indeed, "…it is not enough to determine the extent of a learner's vocabulary size; we also have to evaluate its depth" (Nizonkiza and Van den Berg, 2014: 46). While depth has been associated with overall proficiency among EFL/ESL students (see section 2.4), results remain inconclusive; and the productive component seems to have been neglected. It thus makes sense to examine the question of measuring productive vocabulary knowledge from a deep word knowledge perspective with the hope that major pedagogical implications comparable to those realised at the vocabulary size level could arise from the results (Nizonkiza and Van den Berg, 2014).

2.3. Testing productive knowledge of vocabulary

Productive knowledge of vocabulary consists of two aspects, namely free productive and controlled productive knowledge. The available literature indicates that free productive knowledge is measured mainly through lexical richness and association tasks. According to Nizonkiza and Van den Berg (2014), while the Lexical Frequency Profile (LFP) developed by Laufer and Nation (1995) has proven to be the most widely used test to measure lexical richness, the Lex30 developed by Meara and Fitzpatrick (2000) seems to be the most popular test using association tasks. The LFP measures the proportion of words used in an essay in terms of the word frequency bands[2] from where the words used come. As described in Laufer and Nation (1995), test-takers are required to write an essay within a certain word limit (300 words for instance). A computerised system then counts and weighs the number of words in each frequency band against the total number of words in the essay. For example, in an essay of a total of 200 words, of which 150 words belong to the 1,000-word band, 20 to the 2,000-word band, 20 to the University Word List, and 10 to none of the bands considered, the LFP is 75%, 10%, 10%, and 5% respectively (Laufer and Nation, 1995). Scores are interpreted in terms of frequency bands and the more words from infrequent word bands are used the more proficient the learners are.

Being computerised constitutes the main advantage of the LFP. The test is also reported to be valid and reliable and to discriminate between students from different learning stages (linguistic proficiency levels) (Laufer and Nation, 1995). However, the LFP does not allow distinguishing words that are well known from those which are partially known. In particular, the test does not seem to make any reference to the way words combine in lexical phrases –collocations. We consider this as a limitation because collocations are among the determining factors which characterise good writing (Gledhill, 2000; Paquot, 2008; Schmitt, 1994) and should thus be subject to evaluation in a writing task. Furthermore, a valid measure of free productive vocabulary should require a huge amount of text, which is even difficult to get with native speakers whose texts may also consist of a small set of highly frequent words (Meara and Fitzpatrick, 2000).

---

2   Words in English are classified in frequency bands which consist of 1,000 words each. Nation (2006) first came up with a list consisting of 14 bands, which has been up-dated and consists of 25 frequency bands now.

The Lex30 consists of 30 stimulus words (henceforth the name Lex30 is used) all of which selected from Nation's (1984) first 1,000 most frequent words. According to the authors, most frequently used words are deliberately considered in this test in order to make the test usable even at lower levels of proficiency. Participants are presented with the stimulus words, one at a time, and instructed to write as many associates as they can (at least three). The associates are lemmatised using Range (Heatly and Nation, 1994), a computer programme that automatically recognises words and classifies them into word families according to their frequency bands[3].

Like the LFP, credit is accorded by considering the frequency bands, with only words from low frequency bands given credit. Words from the 1000-word band for instance, proper nouns, and numbers are not given any credit. The main reason for not giving any credit for words from the 1000-word band is that these are highly frequent words. For the test designers, words from this band may not distinguish between linguistic proficiency levels. They thus suggest giving credit for words from infrequent word bands because more proficient students produce more words that belong to infrequent bands; and in this case, the test helps distinguish between learners from different linguistic proficiency levels. Validation studies proved the test to be valid and reliable as well as able to discriminate between linguistic proficiency levels (Fitzpatrick and Clenton, 2010; Meara and Fitzpatrick, 2000; Walters, 2012).

Controlled productive knowledge is measured by means of the Productive Vocabulary Levels Test (PVLT). The test was developed by Laufer and Nation (1999) and measures controlled productive ability, which refers to "the ability to use a word in an unconstrained task such as writing a sentence or a constrained one such as supplying the missing words where a sentential context is provided" (Laufer and Nation, 1999: 37). It is a productive variant of Nation's Vocabulary Levels Test (1990). It retains words from the same word frequency bands, namely the 2000-word, 3000-word, 5000-word, 10000-word, and the Academic Word list. Each tested item is embedded in a sentential context with the first two letters provided. The test-takers' role is to fill in the missing letters (Laufer and Nation, 1999).

Given that some test items require more word knowledge and more use of contextual information, the test has been criticised with regard to what it measures as a whole (Read, 2000; Schmitt, 2010). Furthermore, "......this kind of test item can easily identify what the testees do not know, but it is rather less successful at identifying the full extent of what they do know" (Meara and Fitzpatrick, 2000: 21). The test has also been criticised for it does not make it clear if the test-taker can use the word or recall it if not prompted (Schmitt, 2010). Laufer and Nation (1999) used it for evidence of the validity of the blank-

---

3   Range has been updated several times, with its latest version accessible from: http://www.lextutor.ca/range/.

filling and argued that learners at higher proficiency levels obtained significantly better scores than lower proficiency learners, which is its merit. It is also a very practical test, which is, easy to administer, easy to mark, easy to interpret, and can be completed in a short time (Laufer and Nation, 1999).

2.4   Testing deep word knowledge of vocabulary practices

The different aspects of depth knowledge have been measured mainly through what Read (2000) refers to as a developmental approach tested through Wesche and Paribakht's (1996) VKS and the dimensions approach measured through Read's (1993) WAT. The developmental approach considers word knowledge as a gradual process from zero knowledge to full knowledge of the word while the dimensions approach suggests assessing Read's (1993) aspects of deep word knowledge, that is, paradigmatic, analytic, and syntagmatic relations (Read, 2000).

The VKS is a self-report test that combines self-assessment and the production of verifiable evidence, and therefore aims to test both receptive and productive knowledge in the same test battery. It presents decontextualised words to participants who have to report their knowledge of each item on a five point scale, that is:

"I.   I don't remember having seen this word before.

II.   I have seen this word before, but I don't know what it means.

III.   I have seen this word before, and I think it means ———. (synonym or translation)

IV.   I know this word. It means ———. (synonym or translation)

V.   I can use this word in a sentence: ———. (Write a sentence.) (If you do this section, please also do Section IV.)" (Paribakht and Wesche, 1997: 180).

The VKS exists in different versions with Wesche and Paribakht (1996) and Paribakht and Wesche (1997) being referred to most often in the literature and the most widely used (Bruton, 2009; Read, 2007; Schmitt, 2010; Schmitt et al., 2011). The test has been used as a research tool and studies such as Zareva et al. (2005) indicate that the test scores correlate with overall linguistic proficiency, meaning that the test can be used as a linguistic proficiency indicator.

However, not all scholars agree that the VKS can be used as an indicator of (linguistic) proficiency, and many have noted the serious limitations of the test. For example, the test does not seem to gauge lexical development as claimed by the test designers, but rather simply shows ability – inability of learners to label the target words in terms of their meaning and use (Bruton, 2009; Henriksen, 1999; Meara, 1996; Read, 2000).

229

The manner in which scores are obtained has also sparked criticism among scholars (Bruton, 2009; Read, 1998), mainly because the test relies on an unverified self-report (Read, 1993, 1998). Furthermore, in order to assess productive knowledge, the test requires learners to use a word in a sentence if they claim that they know it. However, they may use the word in a more general sense, which makes it difficult to assess the extent to which they know it (Bruton, 2009; Read, 1993, 1998). If a student for instance gives a sentence such as "I love mountains" when asked to use the word "mountain" in a sentence, it may be rather difficult to tell exactly how much of the word "mountain" is known by the student. Another criticism addressed to the VKS and a more fundamental one is that, the test does not seem to measure both receptive and productive knowledge as claimed by the test developers (cf. Bruton, 2009; Henriksen, 1999; Meara, 1996; Read, 2000).

The WAT was developed by Read (1993) and consists of a list of stimulus words that are presented to participants with a set of words, half of which are associates and the other half of which are non-associates. Associates match Read's (1993) aspects of deep word knowledge, that is, they are paradigmatic (synonyms, or close in meaning), analytic (key words of definition), and syntagmatic (collocations) relations of the stimulus word. In Read's (1993: 359) example, the word "edit" is presented with "arithmetic", "revise", "film", "risk", "pole", "surface", "publishing", and "text". Participants are instructed to identify the associates. Revise is related to edit, publishing, and text and film respectively through paradigmatic, analytic, and syntagmatic relations. The other words –arithmetic, risk, pole, surface– are distractors. The original version consisted of four associates and four non-associates. However, its modified versions use three associates and three non-associates instead, the main argument being that it is not always possible to find four words semantically related to a given stimulus word (Greidanus et al., 2004).

Validation studies indicate that the WAT is a good measure of deep word knowledge (Read, 1998, 2000; Schmitt et al., 2011; Schoonen and Verhallen, 2008). The same validation studies reveal that the WAT has also been widely used as a research tool (Greidanus et al., 2004; Read, 2007; Schmitt et al., 2011). Furthermore, the test distinguishes between (linguistic) proficiency levels (Nizonkiza, 2011; Schoonen and Verhallen, 2008). Another interesting observation is that, as opposed to the VKS, the test covers both meaning and collocations (Batty, 2012; Schmitt, 2010).

However, willingness to guess, which may lead to overestimating participants' knowledge, constitutes a threat to its reliability (Greidanus et al., 2004; Read, 1993, 1998; Schmitt et al., 2011). The aspects of depth (paradigmatic, analytic, and collocations) as measured by the WAT, may also consist of disconnected elements (Batty, 2012; Milton, 2009; Read 2000, 2004); calling into question the traditional definition of depth. Batty (2012) and Milton (2009) further argue that testing depth knowledge is challenging because of this absence of a clear definition and construct of what depth really entails. Batty (2012: 75), for instance, questions clustering together all the associates on the grounds that "knowledge of synonyms and collocates are distinct

subskills of vocabulary knowledge". In keeping with the above, Nizonkiza and Van den Berg (2014) support the call for redefining the concept of depth as a whole, which could contribute towards knowing which aspects "should remain together and which ones should be set apart". We agree with Nizonkiza and Van den Berg (2014: 55) that "identifying a component or several components which may best represent the others is what matters more, and testing only this/these then seems warranted". We believe that the best way to proceed for such identification is by testing the components of depth in comparison to one another. We also agree with the new conceptualisation of depth knowledge, which suggests considering productive knowledge as a fundamental component of depth knowledge (cf. Schmitt et al., 2011).

The present study adheres to this new approach and tests the aspects of deep word knowledge (paradigmatic, analytic, and collocations) as defined by Read (1993) by means of a test modelled on Lex30 (see section 2.3) and attempts to answer the following questions:

1. Do higher education EFL and ESL students of comparable class level perform differently on productive depth knowledge?

2. What is the order in which productive knowledge of aspects of depth grow in EFL and ESL students? In other words, the question is to know in which order components of depth knowledge grow and whether or not the same order is observed in both EFL and ESL students.

3. Do depth components grow alongside one another in EFL and ESL students and to what extent do they contribute to overall depth?

3. The present study

3.1. Population investigated

The present study was conducted on two groups of participants, that is, EFL and ESL students at higher education. The EFL participants were Burundians majoring in English at the beginning of their third year of the Bachelor of Arts (BA) degree in English Language and Literature at a University. These participants were enrolled in a degree to be completed in four years. Their average age was twenty four years. They were recruited from a writing course and were informed that participating in the study was meant for research purposes and they were thus encouraged to do so. All the students were informed a day before the actual data collection took place and everyone (n = 157) who was present that day participated in the study. They spoke four languages, that is, Kirundi, their mother tongue; French, an official language in Burundi; Kiswahili, a lingua

franca of the region; and English, the language they were majoring in and which holds the status of being a foreign language in Burundi.

The ESL participants were South Africans (n = 53) whose average age was twenty years. They were recruited from a University from where they were taking a second-year English course, which integrates literature and language as one of their courses in the Human and Social Sciences. For some of these students, English was a major subject and for some an additional subject only from the first to the second year levels. They participated in the study at the end of the second semester. The students mostly came from the townships, villages, and informal settlements near the campus. Given the status of English in South Africa and the additional exposure to English through radio, TV, newspapers, the Internet, etcetera; these learners are entitled to being ESL users of English. Participants were informed about the research objectives and they accepted the invitation to participate. Most of them spoke Setswana, their mother tongue, and English, their second language. Given that the ESL students participated in the study at the end of the first semester of their second year while the EFL students participated in the study at the beginning of their third year, the two groups could be assumed to be of comparable educational level: both groups of participants were more or less at the level of two years' experience in higher education.

3.2. Instruments

For the purpose of this study, a test modelled on Lex30 was developed and presented to the participants. It should be noted that even though the present study adopted the Lex30 test, the scoring pattern was adapted. We refer to the test used in the present study as a modified version of Lex30 (hence ModLex30). As opposed to the original version, credit is given by considering the number of associates produced irrespective of their frequency bands. The underlying reason for this decision is that associates, particularly collocations, are not necessarily infrequent. While the meaning of frequent collocations' constituents (individual words) does not seem to cause problems for comprehension, using them in correct combinations remains problematic for EFL/ESL students (Paquot, 2008). We believe that giving credit for associates, irrespective of their frequency bands, would offer some important information which is otherwise omitted in the original version of Lex30.

This study retains twenty nouns used in Nizonkiza's (2012) controlled productive test of collocations drawn from the 2000-word and 3000-word bands respectively (see Appendix A). The controlled productive test of collocations test consists of items selected from the 2000-word, 3000-word, and 5000-word bands (Nation, 2006) and the Academic Word List (Coxhead, 2000); but for the purpose of this study, items from the 2000-word and 3000-word bands were retained because they are considered as the most frequent words and thus the most widely used (cf. Schmitt, 2010). They are also likely to be known as individual words by learners from even low linguistic proficiency levels (cf. Meara and Fitzpatrick, 2000).

3.3. Data analysis procedure

The data collected by means of ModLex30 were analysed through different steps. First of all, the data were screened using an online dictionary[4] that provides definitions and synonyms as well as a hard copy dictionary, the *Oxford WORDPOWER Dictionary* (2006), which helped to decide on paradigmatic and analytic associates. The following is a screenshot of the online dictionary for the word 'bomb'. It gives up to five definitions; the first one being 'a hollow projectile containing an explosive, incendiary, or other destructive substance, esp. one carried by aircraft'. It also provides a list of synonyms such as 'explosive, bombard, destroy, charge, attack, device, blitz' and more synonyms can be obtained by clicking on the 'more' button, but only nouns were retained for the purpose of this study. More definitions are also provided under the heading 'World English Dictionary'.
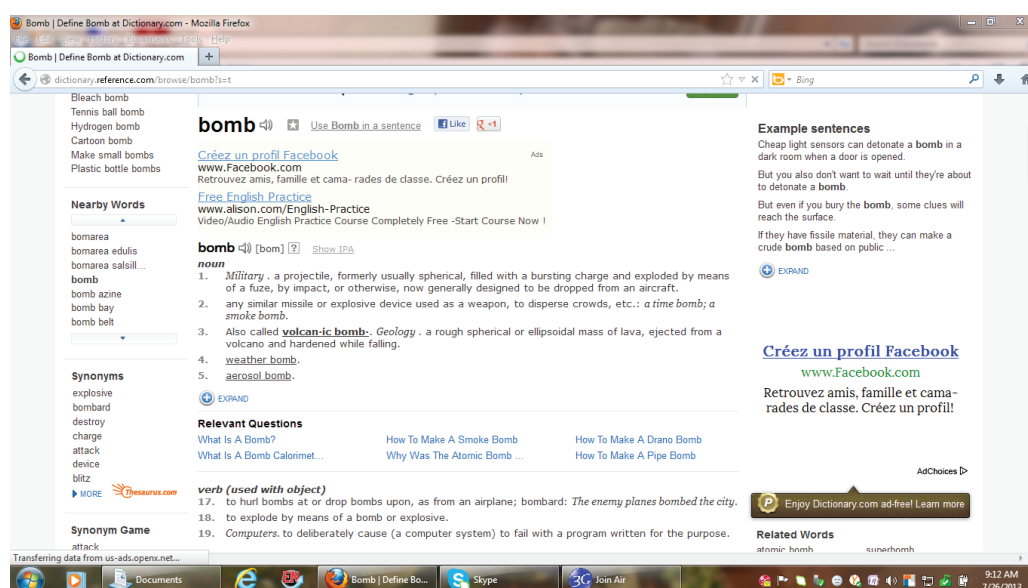
The *Oxford Collocations Dictionary for Students of English* (2002), the hard copy as well as an online version[5] of the same dictionary, which is in essence its new edition, were used in order to decide on whether or not an associate was a collocation. As the screenshot below indicates, once a word is typed in the search window, collocations are presented following the syntactic categories of the combinations. For instance, the word 'bomb' is a noun and collocates with 'big', 'huge', 'large', etcetera in the adjective-noun combination.

---

4   The online dictionary was retrieved from: http://dictionary.reference.com/; accessed from 20th to 27th June 2013.
5   The online collocation dictionary is accessible at: http://oxforddictionary.so8848.com/; accessed from 20th to 27th June 2013.

It collocates with 'place', 'plant', 'put', 'drop', 'release', etcetera in the verb-noun combination while it collocates with 'fall', 'rain', 'hit something', 'detonate', 'explode' and so forth, in the noun-verb combination. It also collocates with other nouns in noun-noun combinations such as 'bomb explosion'.
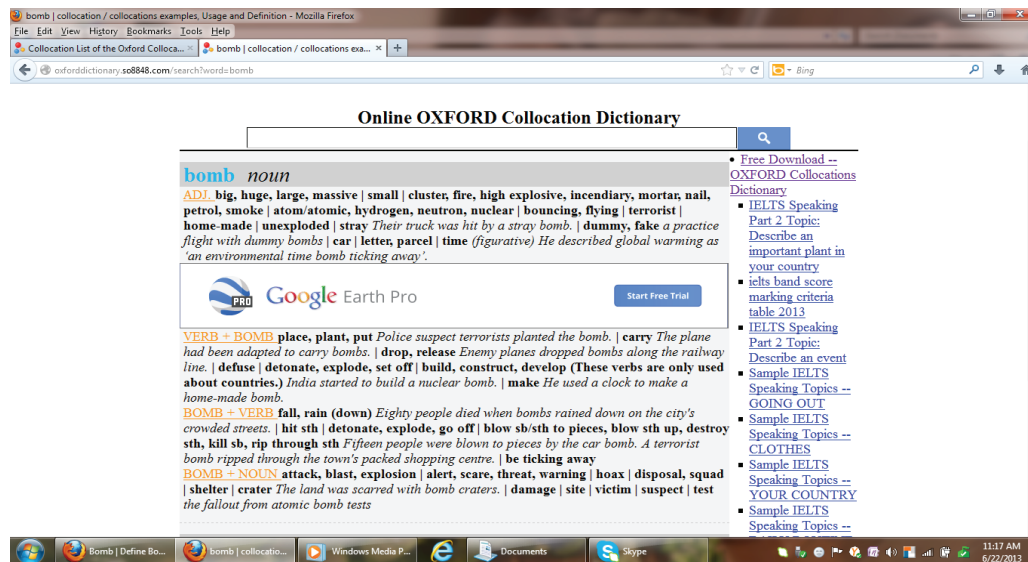
Secondly, the non-associates were eliminated and for comparative purposes, the rest of the associates were classified according to their types, namely paradigmatic and analytic relations and collocations. Students were awarded one point per correct associate and unlike the Lex30, the frequency bands were not considered. To be exact, students were awarded one point per synonym or any other word judged as being close in meaning under the component of paradigmatic relation, one point per word judged as a key element of the definition for the analytic relation component, and one point per correct collocate. For instance, a student who produced the following associates of 'bomb': 'weapon', 'explosion', 'explosive', was awarded one point for the paradigmatic relation because 'explosive' is a synonym of 'bomb'; s/he was awarded two points for the analytic relation because both 'explosive' and 'weapon' are key elements of the dictionary definition, and one point under the category of collocations because 'bomb explosion' is a noun-noun combination.

Scores from the associates (paradigmatic, analytic, and collocations) were averaged for each group of participants (EFL and ESL: the independent variable) and the overall depth scores were obtained by adding up scores from the associates. The mean scores (both overall and for each associate, constituting the dependent variable) were compared afterwards by means of an Independent-Samples T Test, which allowed testing the first research question about which group (EFL vs. ESL) performs better than the other. The means achieved at each of the components (associates) of deep word knowledge

234

(paradigmatic, analytic, and collocations) for both EFL and ESL put together and for each group were compared by computing a one-way repeated ANOVA and the Bonferroni post-hoc test. The latter allowed us to determine the order in which associates grow and how significantly they grow differently; which helped in answering the second research question. Finally, the third research question was answered by correlating associates with one another (paradigmatic, analytic, and collocations) on the one hand and associates to overall deep word knowledge on the other hand. This process was meant to explore the extent to which associates grow alongside one another and the extent to which they contribute to overall depth.

In brief, the first research question was answered by comparing EFL and ESL students' performance (both overall and for each associate) using an Independent-Samples T Test. The second research question was answered by computing a one-way repeated ANOVA involving depth associates and the Bonferroni post-hoc test, while the third research question was answered by correlating associates to one another on the one hand and associates to overall deep word knowledge on the other hand.

4. Results

Performance on deep word knowledge among EFL and ESL students

The first aim pursued in this study is to test productive depth knowledge of vocabulary among both higher education EFL and ESL students of comparable class level and find out which of the two groups performs better than the other. Scores from the administered test were averaged for each group of participants and then compared by means of an Independent-Samples T Test. Means of each depth component and overall depth as well as standard deviations for both groups are presented in Table 1.

Table 1:   Means of overall and components of depth produced by EFL and ESL students

	Group of students	N	Mean	Std. Deviation
Overall Depth	EFL	153	36.92	9.99
	ESL	32	52.00	9.71
Paradigmatic relations	EFL	153	10.85	4.98
	ESL	32	17.00	3.73
Analytic relations	EFL	153	18.77	5.99
	ESL	32	30.12	5.72
Collocations	EFL	153	7.30	5.35
	ESL	32	5.03	2.49

As can be seen from Table 1, results indicate that ESL students scored better than their EFL counterparts on overall depth with the following means: 52.00 and 36.92 respectively. The same holds for each of the components of depth where the two groups achieved mean scores of 17.00 and 10.85 on paradigmatic relations; 30.12 and 18.77 on analytic relations; respectively for ESL and EFL students. However, ESL students did not outperform their EFL counterparts on collocations with 5.03 and 7.03 of mean scores respectively. The observed differences were found to be statistically significant as the T-test for equality of means indicates (with a Sig. of .000 2-tailed; details can be found in Appendix B). These findings answer the first research question about which group (between EFL and ESL students) performs better than the other on productive deep word knowlege.

4.2. Order in which depth components grow among EFL and ESL students

The second aim pursued in the present study is to determine the order in which productive knowledge of depth components (paradigmatic, meaning synonyms, or words that are close in meaning, analytic or key words of definition, and syntagmatic or collocations) develop among EFL and ESL students. A one-way repeated ANOVA, involving scores at each of the components was performed (all participants put together and then each group, EFL and ESL separately). The mean scores are presented in Table 2 and clearly indicate that the number of associates produced by EFL and ESL students (put together or separately) vary greatly. In particular, the same descending order (analytic, or key words of definition, paradigmatic or synonyms, or words close in meaning, and syntagmatic, or collocations) is found in both groups of participants; implying that this might be the order in which aspects of deep word knowledge develop.

Table 2: **Mean scores on depth components**

	Overall: Groups together			ESL			EFL		
	Mean	Std Deviation	N	Mean	Std Deviation	N	Mean	Std Deviation	N
Analytic	20.78	7.37	185	30.37	5.62	32	18.77	5.99	153
Paradigmatic	11.95	5.34	185	17.19	3.63	32	10.85	4.98	153
Collocations	6.91	5.04	185	5.03	2.49	32	7.30	5.35	153

As can be seen from Table 2, for both groups put together, the mean is 20.78 for analytic relations; it drops to 11.95 for paradigmatic relations and to 6.91 for collocations. The exact same descending order is observed for ESL students with a mean of 30.37 for analytic relations, 17.19 for paradigmatic relations, and 5.03 for collocations. The same order is also observed in the EFL group with a mean score of 18.77 for analytic relations, 10.85 for paradigmatic relations, and 7.30 for collocations, respectively.

The mean differences between depth components are statistically significant as shown by the Sphericity Assumed Correction test with its related Sig., which is 0.000 (cf. Appendix C). This means that the mean differences between depth components in terms of scores are statistically significant, both for overall depth and for each group separately. However, the overall significance as shown by the Sphericity Assumed Correction test does not seem to point to where significant differences occur. To this end, the data were analysed further and the Bonferroni post-hoc test was performed. The latter is a multiple-comparison test which indicates between which two components significant differences arise. This was done for both groups put together (cf. Appendix D) and for EFL (cf. Appendix E) and ESL (cf. Appendix F) separately. The observed differences in terms of performance on each two components compared (cf. third column entitled "Mean Difference I-J" for differences) and their associated significance (cf. last column) are statistically significant. These findings answer the second research question about the order in which components of deep word knowledge develop and whether or not the same order is found in both EFL and ESL students.

4.3. Aspects of depth growth and their contribution to overall depth

The third question addresses the possible parallel growth among components of depth and their contribution to its development. It was answered by running a Pearson correlation between the different components of depth on the one hand and between each component and overall depth on the other. It is worth noting that for the latter, in order not to overestimate the correlations, each component was compared to overall depth minus the component in question. Details of the correlations are presented in Appendix G and Appendix H, for the EFL and ESL students respectively.

Among the EFL students, we observe a significant correlation between paradigmatic and analytic relations (r = .635**, p = 0.01). However, there is a negative correlation between paradigmatic relations and collocations, which is significant (r = -.241**, p = 0.01), while the correlation between analytic relations and collocations is positive, but not significant (r = .012). Components of depth correlate with overall depth in the following descending order: analytic relations (r = 515**, p = 0.01), which is large and significant; paradigmatic relations (r = .337**, p = 0.01), which is moderate and significant; and collocations (r = -.110), which is negative and not significant.

Among the ESL students, however, the correlations are somewhat different and all the components of depth correlate significantly. We observe a large and significant correlation between paradigmatic and analytic relations (r = .662**, p = 0.01). There are also large and significant correlations between paradigmatic relations and collocations (r = .402*; p = 0.05) and between analytic relations and collocations (r = .384*, p =0.05). The order of the strength of the correlations between the components and overall depth is similar to that of the EFL students, but with much stronger correlations (r = .792**, p= 0.01, for analytic relations; r = .746**, p = 0.01 for paradigmatic relations; and r = .460**, p = 0.01 for collocations). These findings answer the third question addressed in the present study about the growth of depth components in comparison with one another and their contribution to overall depth.

5.   Discussion

The present study explores whether vocabulary depth knowledge can be measured at the productive level. This was achieved by administering a modified version of Lex30 to EFL and ESL students of English at higher education. Results suggest that ESL students outperform their EFL counterparts of comparable higher education level. An in-depth analysis at the level of aspects of depth knowledge confirms this observation by showing that ESL students perform better than their EFL counterparts, except on collocations.

With regard to the order in which aspects of depth knowledge develop, results indicate that aspects of depth significantly grow in the following descending order among both EFL and ESL students: analytic relations, paradigmatic relations, and collocations. Production of collocations is much lower in both groups, which confirms previous findings that collocations may be more problematic than other aspects of depth (cf. Nizonkiza, 2011) and that both EFL and ESL students underuse collocations (Altenberg and Granger, 2001; Howarth, 1998). In other words, these findings confirm the observation that collocations remain problematic for both EFL and ESL students even at an advanced level (cf. Laufer and Waldman, 2011; Nesselhauf, 2005). Furthermore, and contrary to our expectations, the EFL students produced more collocations than their ESL counterparts. Even though we have no explanation of this phenomenon, which is supported with empirical evidence, we posit that this could be attributed to the fact that EFL students sat the test during a course on academic writing, one aim of which is to raise students' awareness of collocations[6].

Furthermore, the three aspects of depth measured in this study (i.e. analytic relations, paradigmatic relations, and collocations) were found to grow alongside one another in the group of ESL students, but contribute to overall depth to varying degrees. Indeed, analytic relations correlate more strongly than paradigmatic relations do with overall depth. The latter aspect, in turn, correlates more strongly than collocations correlate with overall depth. We believe that this strength of correlation may reflect the extent to which they contribute to the development of overall depth. The situation seems to be, however, less predictable among EFL students where only analytic and paradigmatic relations positively and significantly correlate. It is difficult to account for the absence of a significant correlation between analytic relations and collocations on the one hand, and the correlation between paradigmatic relations and collocations that is negative and significant on the other hand. This may be the result of some inconsistency in the growth of collocations among EFL students or the instrument used does not allow the testing of the three aspects in the same test battery (see concluding remarks) or the aspects tested might belong to different dimensions as assumed by Batty (2012) that depth may consist of disconnected elements. Correlations between the three different aspects of depth knowledge and overall depth are similar to those of the ESL students and are statistically significant (with the exception of collocations). We believe that they may contribute to overall depth in the same order of importance as in the ESL students.

---

6   It should be noted, however, that this was the first time collocations were introduced to students.

We are aware of the fact that not administering a standardised vocabulary size test to our participants or a proficiency measure, and therefore comparing their size or their proficiency to depth, is the main limitation of the present study. We plan to eliminate this limitation in a follow-up study (see section 7).

6. Teaching implications

In light of the discussion of the results in the above section, considering and discussing possible pedagogical implications arising from these findings seem to be warranted. This study is indeed an attempt to measure productive knowledge of deep word knowledge and the order in which its aspects grow among EFL and ESL students. Even though, the ESL students were found to do better than their EFL counterparts, which could actually be expected given the exposure to the language which is different for the two groups, the same order of difficulty of the aspects of depth is found in both groups: collocations, paradigmatic relations, and analytic relations. Furthermore, these aspects of depth seem to contribute to overall depth with more or less the same weight of importance in both groups. The same order of difficulty which is found in both groups irrespective of exposure to the language is very indicative here and could be interpreted in two ways.

First of all, this could reflect the order in which the aspects of depth actually develop. However, it could also be the result of the teaching practices adopted. While the teaching practices in use in Burundi and in South Africa are not exactly the same when it comes to teaching English, what is common is that a newly introduced word is defined (sometimes translated) and used in an example sentence. Once in a while a synonym or a related word in terms of meaning is provided. The collocation component seems to be overlooked and it is left to the student to acquire or learn collocations via exposure to the language. Exposure to the language could, without doubt, be the best learning condition for vocabulary to develop. However, it is not enough in ESL and EFL contexts where students need more explicit reference to the target word even for receptive purposes (cf. Schmitt, 2008). Explicit teaching of vocabulary is needed especially for productive purposes; collocations in particular, the use of which requires pedagogical treatment (cf. Granger and Paquot, 2008; Laufer and Waldman, 2011; Nesselhauf, 2005; Nizonkiza, 2012). For the above reasons, we advise adopting explicit teaching of aspects of depth.

Our suggestion is in line with our second interpretation of the same order of difficulty of aspects of depth observed in both groups. The latter could imply that the same teaching approaches could be adopted to varying degrees in ESL and EFL contexts. We suggest introducing the aspects of depth concomitantly. We are of the opinion that analytic relations seem to be the only readily acquired aspect possibly because this is the only aspect which is mainly focused on and introduced to learners each time learners come across a new word.

With regard to introducing all the aspects of depth as early as possible, we urge caution of course and advise ESL and EFL practitioners and material designers to consider among other things the frequency of the words involved and the learning stages of the learners as Nizonkiza and Van de Poel (2014) suggest[7]. This could also give a platform for testing the order in which these aspects grow. With the results from this study, we cannot convincingly argue that either aspect of depth is more challenging or grows faster than others because they do not seem to be attended to in the same way in the classroom. We posit that even though collocations are found to be very difficult for both ESL and EFL students, this might be related to the teaching approaches adopted in ESL and EFL contexts where less attention is paid to collocations. We believe that the teaching of collocations as early as possible could result in their growth, but only results from empirical evidence can tell us more about the extent to which this could be the case. The comparison of aspects could give us groundwork for remedial strategies to adopt for a productive oriented teaching of vocabulary.

7.   Conclusion

This study has examined the measurability of deep word knowledge at the productive level by means of ModLex30, a modified version of Lex30 administered to higher education EFL and ESL students. The results show that: (i) ESL students outperform their ESL counterparts on both overall depth and aspects of depth, except on collocations, (ii) aspects of depth grow in the order of analytic relations, paradigmatic relations, and collocations in both EFL and ESL students; (iii) and depth aspects grow alongside one another, but less predictably among EFL students where correlations are not as strong as those among ESL students. Aspects of depth correlate significantly with one another and with overall depth with much stronger correlations in ESL students than in EFL students where correlations are not significant between all the aspects of depth. These strengths of correlations between aspects of depth and overall depth may actually reflect the order of importance with which they contribute to overall depth. On the basis of these findings, we conclude that Lex30 may be a good measure of depth at the productive level, which is the path we suggest following in order to test deep word knowledge.

These findings answer the questions raised initially, but also give rise to new questions worth exploring in follow-up studies.

• First of all, even though we did not perform any qualitative analysis of individual items of the aspects of depth produced by participants, we observed that some students could produce only collocates or synonyms or key words of a dictionary definition for a particular item. The answers that were provided show the

---

7   We suggest adopting the collocation web model proposed by Nizonkiza and Van de Poel (2014) where among other things, frequency of words and learning stages of students are considered as determining factors in words' learnability.

words participants think of upon seeing a particular stimulus word, but this could overestimate or underestimate their knowledge of the different associates. We believe that the instrument used, as it stands, does not seem to allow for producing the aspects of depth to the full potential and to a more or less same extent. We therefore recommend replicating the present study through three sub studies, specifically asking participants to:

° Give paradigmatic associates (sub study one);

° Give analytic associates (sub study two); and

° Give collocations (sub study three).

Such a study could offer a broader view of the student's knowledge in terms of the different aspects of depth.

- Secondly, comparing deep word knowledge as measured by ModLex30 and the original version of Lex30 and vocabulary size, for instance the Vocabulary Levels Test (VLT) on the one hand, and overall proficiency on the other hand, could help gain greater insights into the development of depth knowledge in comparison with productive knowledge as well as proficiency; which we strongly recommend. This echoes Schmitt's (2010) suggestion that the best way to test depth is to combine approaches, which he expresses in the following terms:

"Perhaps the best solution is to combine approaches, with some measures estimating the 'quantity' realm (e.g. size of lexicon), and others tapping in the 'quality' of the lexical knowledge within that realm. These combined measures could be contained within the same study, or if time is a constraint, then within consecutive studies, whose results can be linked for greater understanding" (Schmitt, 2010: 241).

- Thirdly, extending this study to native speakers may contribute to determining the exact order of difficulty of depth aspects. In other words, as it stands now, we are not in a position to tell exactly if this is the real order in which aspects of depth develop or if it is dictated by the teaching approaches adopted or both. Alternatively, introducing the aspects of depth at the same time as suggested in the previous section and assessing the growth of each aspect in comparison with others may also contribute towards gaining insights into the growth of depth components in relation to one another.

In a nutshell, our study has achieved the set objectives and has demonstrated that vocabulary depth can be measured productively and that the order in which the various aspects of depth develop can be determined. We believe that this is insightful and that we have introduced a useful test that could be used in this research area. At least, this study could help to encourage debate around the topic of testing different aspects of vocabulary knowledge (or deep word knowledge) productively.

References

Aitchison, J. 1989. *Working with Words: An Introduction to Psycholinguistics*. London: Unwin Hyman.

Altenberg, B. and Granger, S. 2001. The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing. *Applied Linguistics* 22(2): 173-195.

Batty, A. 2012. Identifying Dimensions of Vocabulary Knowledge in the Word Associates Test. *Vocabulary Learning and Instruction* 1(1): 70-77.

Beglar, D. 2010. A Rasch-based Validation of the Vocabulary Size Test. *Language Testing* 27(1): 101-118.

Bruton, A. 2009. The Vocabulary Knowledge Scale: A Critical Analysis. *Language Assessment Quarterly* 6(4): 288-297.

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-239.

Fitzpatrick, T. and Clenton, J. 2010. The Challenge of Validation: Assessing the Performance of a Test of Productive Ability. *Language Testing* 27(4): 538-555.

Gairns, R. and Redman, S. 1986. *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge: Cambridge University Press.

Gledhill, C.J. 2000. *Collocations in Science Writing*. Tubingen: Gunter Narr Verlag.

Granger, S. and Paquot, M. 2008. Disentangling the Phraseological Web. In: Granger, S. and Meunier, F. (Eds.) 2008. *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company. pp. 27-49.

Greidanus, J. and Nienhuis, L. 2001. Testing the Quality of Word Knowledge in a Second Language by Means of Word Associations: Types of Distractors and Types of Associations. *Modern Language Journal* 85(4): 567-577.

Greidanus, T., Bogaards, P., Vander, L.E., Nienhuis, L. and Dewolf, T. 2004. The Construction and Validation of a Deep Word Knowledge Test for Advanced Learners of French. In: Boogards, P. and Laufer, B. (Eds.) 2004. *Vocabulary in Second Language*. Amsterdam: John Benjamins Publishing Company. pp. 191-208.

Heatly, A. and Nation, I.S.P. 1994. Range [Computer Program. Available at: http://vuw. ac.nz/lals/]. Victoria University of New Zealand.

Henriksen, B. 1999. Three Dimensions of Vocabulary Development. *Studies in Second Language Acquisition* 21(2): 303-317.

Hirsh, D. 2010. Researching Vocabulary. In: Paltridge, B. and Phakiti, A. (Eds.) 2010. *Continuum Companion to Research Methods in Applied Linguistics*. London: Continuum. pp. 222–239.

Howarth, P. 1998. Phraseology of Second Language Proficiency. *Applied Linguistics* 19(1): 24-44.

Hu, M. and Nation, I.S.P. 2000. Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language* 13(1): 403-429.

Ishii, T. and Schmitt, N. 2009. Developing an Integrated Diagnostic Test of Vocabulary Size and Depth. RELC Journal 40(1): 5-22.

Laufer, B. and Nation, I.S.P. 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16(3): 307-322.

Laufer, B. and Nation I.S.P. 1999. A Vocabulary Size Test of Controlled Productive Ability. *Language Testing* 16(1): 33-51.

Laufer, B. and Paribakht, S. 1998. The Relationship between Passive and Active Vocabularies: Effects of Language Learning Contexts. *Language Learning* 48(3): 365-391.

Laufer, B. and Ravenhorst-Kalovski, G.C. 2010. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language* 22(1): 15-30.

Laufer, B. and Waldman, T. 2011. Verb-noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61(2): 647–672.

Meara, P. 1996. The Dimensions of Lexical Competence. In: Brown, G., Malmkjaer, K. and Williams, J. (Eds.) 1996. *Competence and Performance in Language Learning*. Cambridge: Cambridge University Press. pp. 35–53.

Meara, P. and Buxton, B. 1987. An Alternative to Multiple Choice Vocabulary Tests. *Language Testing* 4(2): 142-154.

Meara, P. and Fitzpatrick, T. 2000. Lex30: An Improved Method of Assessing Productive Vocabulary in an L2. *System* 28(1): 19-30.

Meara, P. and Jones, G. 1988. Vocabulary Size as a Placement Indicator. In: Grunwell, P. (Ed.) 1988. *Applied Linguistics in Society*. London: Centre for Information and Language Teaching and Research. pp. 80-87.

Melka, F. 1997. Receptive versus Productive Vocabulary. In: Schmitt, N. and McCarthy, M. (Eds.) 1997. *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press. pp. 84-102.

Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Milton, J. and Treffers-Daller, J. 2013. Vocabulary Size Revisited: The Link between Vocabulary Size and Academic Achievement. *Applied Linguistics Review* 4(3): 151-172.

Nation, I.S.P. (Ed.) 1984. *Vocabulary Lists: Words, Affixes and Stems*: Victoria University of Wellington, Wellington, New Zealand: English Language Institute.

Nation, I.S.P. 2006. How Large a Vocabulary Is Needed for Reading and Listening? *Canadian Modern Language Review* 63(1): 59-82.

Nation, I.S.P. and Beglar, D. 2007. A Vocabulary Size Test. *The Language Teacher* 31(7): 9-13.

Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Company.

Nizonkiza, D. 2011. Relationship between Lexical Competence, Collocational Competence, and L2 Proficiency. *English Text Construction 4(1)*: 113-146.

Nizonkiza, D. 2012. Quantifying Controlled Productive Knowledge of Collocations across Proficiency and Word Frequency Levels. *Studies in Second Language Learning and Teaching 2(1)*: 67-92.

Nizonkiza, D. and Van de Poel, K. 2014. Teachability of Collocations: The Role of Word Frequency Counts. *Southern African Linguistics and Applied Language Studies 32(1)*: 301-316.

Nizonkiza, D. and Van den Berg, K. 2014. Dimensional Approach to Vocabulary Testing: What Can We Learn from Past and Present Practices? *Stellenbosch Papers in Linguistics Plus 43*: 1-14.

Oxford Collocations Dictionary for Students of English. 2002.

Oxford WORDPOWER Dictionary, 2006.

Paquot, M.   2008.   Exemplification in Learner Writing: A Cross-Linguistic Perspective. In: Granger, S. and Meunier, F. (Eds.)   2008.   *Phraseology in Foreign Language Learning and Teaching*.   Amsterdam: John Benjamins Publishing Company. pp. 101-119.

Paribakht, T.S. and Wesche, M. 1997. Vocabulary Enhancement Activities and Reading for Meaning in Second Language Vocabulary Acquisition. In: Coady, J. and Huckin, T. (Eds.) 1997.   *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press.   pp. 174–200.

Qian, D. and Schedl, M.   2004.   Evaluation of an In-depth Vocabulary Knowledge Measure for Assessing Reading Performance.   *Language Testing* 21(1): 28-52.

Read, J.   1993.   The Development of a New Measure of L2 Vocabulary Knowledge. *Language Testing* 10(3): 355-371.

Read, J.   1998.   Validating a Test to Measure Depth of Vocabulary Knowledge.   In: Kunnan, A. (Ed.)   1998. *Validation in Language Assessment*. Mahwah, NJ: Lawrence Erlbaum.   pp. 41-57.

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

Read, J.   2004.   Plumbing the Depths: How Should the Construct of Vocabulary Knowledge Be Defined? In: Bogaards, P. and Laufer, B. (Eds.) 2004. *Vocabulary in Second Language*.   Amsterdam: John Benjamins Publishing Company. pp. 209-227.

Read, J.   2007.   Second Language Vocabulary Assessment: Current Issues and New Directions.   *International Journal of English Studies* 7(2): 105-125.

Schmitt, N. 1994. Vocabulary Testing: Questions for Test Development with Six Examples of Tests of Vocabulary Size and Depth.   *Thai TESOL Bulletin* 6(2): 9-16.

Schmitt, N.   2008.   Review Article: Instructed Second Language Vocabulary Learning. *Language Teaching Research* 12(3): 329–363.

Schmitt, N.   2010.   *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., Ng, C.J.W. and Garras, J.   2011.   The Word Associates Format: Validation Evidence.   *Language Testing* 28(1): 105-126.

Schmitt, N., Schmitt, D. and Clapham, C.   2001.   Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test.   *Language Testing* 18(1): 55-88.

Schoonen, R. and Verhallen, M.   2008.   The Assessment of Deep Word Knowledge in Young First and Second Language Learners.   *Language Testing* 25(2): 211-236.

Van de Poel, K. and Swanepoel, P.   2003.   Theoretical and Methodological Pluralism in Designing Effective Lexical Support for CALL.   *CALL Journal* 16(2): 173-221.

Vermeer, A.   2001.   Breadth and Depth of Vocabulary in Relation to L1 and L2 Acquisition Frequency Input.   *Applied Psycholinguistics* 22(2): 217-234.

Walters, J.D.   2012.   Aspects of Validity of a Test of Productive Vocabulary: Lex30. *Language Assessment Quarterly* 9(2): 172-185.

Wesche, M. and Paribakht, S.   1996.   Assessing Second Language Vocabulary Knowledge: Depth versus Breadth.   *Canadian Modern Language Review* 53(1): 13-40.

Zareva, A., Schwanenflugel, P. and Nikolova, Y.   2005.   Relationship between Lexical Competence and Language Proficiency.   *Studies in Second Language* 27(4): 567-595.

Appendices

Appendix A: ModLex30

Productive Vocabulary Test

Name: Surname:

Instruction: Below is a list of nouns, could you give at least three words that you think of when you see them?

1. intention
2. bombs
3. commission
4. definition
5. energy
6. secret
7. accuracy
8. mess
9. permit
10. gap
11. anniversary
12. bow
13. congratulations
14. echo
15. glory
16. jewellery
17. nest
18. rank
19. soul
20. whistle

Appendix B: Independent-Samples T Test Results

		Levene's Test for Equality of Variances	T-test for Equality of Means	
		Mean difference	Sig. (2-tailed)	
Overall depth	Equal variance assumed	.568	-15.078	.000
	Equal variance not assumed			.000
Paradigmatic	Equal variance not assumed	.231	-6.150	.000
	Equal variance assumed			.000
Analytic	Equal variance assumed	.642	-11.350	.000
	Equal variance not assumed			.000
Collocations	Equal variance assumed	.000	2.269	.000
	Equal variance not assumed			.000

Appendix C: Sphericity Assumed Correction

Group	Sig.
EFL-ESL overall	.000
EFL students	.000
ESL students	.000

Appendix D: Pairwise Comparisons all Groups Put Together

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[b]
Paradigmatic	Analytic	-8.83*	.393	.000
	Collocations	5.038	.624	.000
Analytic	Paradigmatic	8.83*	.393	.000
	Collocations	13.87*	.707	.000
Collocations	Paradigmatic	-5.03*	.624	.000
	Analytic	-13.870	.707	.000
Based on estimated marginal means				
*. The mean difference is significant at the .05 level.				
b. Adjustment for multiple comparisons: Bonferroni.				

Appendix E: Pairwise Comparisons among EFL Students

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[b]
Paradigmatic	Analytic	-7.92*	.415	.000
	Collocations	3.54	.684	.000
Analytic	Paradigmatic	7.92*	.415	.000
	Collocations	11.47*	.680	.000
Collocations	Paradigmatic	-3.54	.684	.000
	Analytic	-11.47	.680	.000
Based on estimated marginal means				
*. The mean difference is significant at the .05 level.				
b. Adjustment for multiple comparisons: Bonferroni.				

Appendix F: Pairwise Comparisons for ESL Students

(I) factor1	(J) factor1	Mean Difference (I-J)	Std. Error	Sig.[b]
Paradigmatic	Analytic	-13.188*	.732	.000
	Collocations	12.156*	.646	.000
Analytic	Paradigmatic	13.188*	.732	.000
	Collocation	25.344*	1.080	.000
Collocations	Paradigmatic	-12.156*	.646	.000
	Analytic	-25.344*	1.080	.000
Based on estimated marginal means				
*. The mean difference is significant at the .05 level.				
b. Adjustment for multiple comparisons: Bonferroni.				

Appendix G: Correlations of Associates among EFL

		Paradigmatic	Analytic	Collocations	Paradigmatic-Analytic	Analytic-Collocations	Paradigmatic-Collocations	Overall Depth
Paradigmatic	Pearson Correlation	1	.635**	-.241**	.879**	.337**	.597**	.708**
	Sig. (2-tailed)		.000	.002	.000	.000	.000	.000
	N	157	157	157	157	157	157	157
Analytic	Pearson Correlation	.635**	1	.012	.927**	.777**	.515**	.872**
	Sig. (2-tailed)	.000		.884	.000	.000	.000	.000
	N	157	157	157	157	157	157	157
Collocations	Pearson Correlation	-.241**	.012	1	-.110	.638**	.634**	.369**
	Sig. (2-tailed)	.002	.884		.169	.000	.000	.000
	N	157	157	157	157	157	157	157
Paradigmatic-Analytic	Pearson Correlation	.879**	.927**	-.110	1	.644**	.609**	.883**
	Sig. (2-tailed)	.000	.000	.169		.000	.000	.000
	N	157	157	157	157	157	157	157
Analytic-Collocations	Pearson Correlation	.337**	.777**	.638**	.644**	1	.796**	.904**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	157	157	157	157	157	157	157
Paradigmatic-Collocations	Pearson Correlation	.597**	.515**	.634**	.609**	.796**	1	.869**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	157	157	157	157	157	157	157
Overall Depth	Pearson Correlation	.708**	.872**	.369**	.883**	.904**	.869**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	157	157	157	157	157	157	157

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix H: Correlations of Associates among ESL

		Paradigmatic	Analytic	Collocations	Paradigmatic-Analytic	Analytic-Collocations	Paradigmatic–Collocations	Overall Depth
Paradigmatic	Pearson Correlation	1	.662**	.402*	.869**	.746**	.886**	.892**
	Sig. (2-tailed)		.000	.023	.000	.000	.000	.000
	N	32	32	32	32	32	32	32
Analytic	Pearson Correlation	.662**	1	.384*	.982**	.969**	.792**	.957**
	Sig. (2-tailed)	.000		.023	.000	.000	.000	.000
	N	32	35	35	35	35	35	35
Collocations	Pearson Correlation	.402*	.384*	1	.460**	.600**	.787**	.596**
	Sig. (2-tailed)	.023	.023		.005	.000	.000	.000
	N	32	35	35	35	35	35	35
Paradigmatic-Analytic	Pearson Correlation	.869**	.982**	.460**	1	.974**	.877**	.987**
	Sig. (2-tailed)	.000	.000	.005		.000	.000	.000
	N	32	35	35	35	35	35	35
Analytic-Collocations	Pearson Correlation	.746**	.969**	.600**	.974**	1	.896**	.989**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	32	35	35	35	35	35	35
Tot-Paradigmatic-Collocations	Pearson Correlation	.886**	.792**	.787**	.877**	.896**	1	.965**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	32	35	35	35	35	39	39
Overall Depth	Pearson Correlation	.892**	.957**	.596**	.987**	.989**	.965**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	32	35	35	35	35	39	39

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

About the author

Déogratias Nizonkiza

North-West University (Potchefstroom Campus)
School of Languages
Private Bag X6001
2520 POTCHEFSTROOM

E-mail: deo.nizonkiza@nwu.ac.za

Déogratias Nizonkiza is a Postdoctoral researcher at North-West University (Potchefstroom Campus). He is also affiliated with the University of Burundi, where he teaches at the Department of English Language and Literature. His research interests include the relationship between vocabulary knowledge and foreign/second language (L2) proficiency; aspects of deep word knowledge; collocations growth, testing, and teaching; the role of collocations in academic texts/literacies.

About the author

Themba Ngwenya

Associate Professor
North-West University (Mafikeng Campus)
School of Human Sciences
Department of English and Academic Literacy
Pvt. Bag X2046
2735 Mmabatho

Email address: themba.ngwenya@nwu.ac.za

Themba Ngwenya is professor at the Department of English and Academic Literacy, North-West University (Mafikeng Campus). He has taught English second language at both high school and higher education level. His language areas of interest include, among other things, language planning and policy; language varieties, especially South African Black English; and academic literacy.