Vocabulary Threshold for Reading Comprehension during the Transition to Secondary School among Tanzanian Public School Students

Frida Alfred Mbwafu

Department of Linguistics and Literary Studies, faculty of Arts and Social Science, the Open University of Tanzania

E-mail: Frida.alfred@out.ac.tz/mbwafufrida@gmail.com

Abstract

This study aimed to examine the vocabulary size of the Form One students in Tanzanian Public secondary schools to determine whether their vocabulary size was sufficient for them to comprehend their textbooks. Scholarly, the link between vocabulary size and reading comprehension is well-reported in second language learning studies. Whereas, different studies report that a second/foreign language learner needs to understand between 95% and 98% of the running words of a text to have adequate comprehension. This suggests for the vocabulary size between 4,000 and 9,000 words to enable an ESL/EFL learner to understand 95% - 98% respectively of the running words for authentic texts, including academic ones. The study adopts a quantitative research approach and a case study design. Data were collected through the Vocabulary Size Test (VST) by Nation and Beglar (2007) and analyzed by SPSS version 23. The study was guided by the threshold hypothesis of Cummins (1976) to establish the factual evidence of the study. Consequently, the examined vocabulary size of the Tanzanian Public Secondary Schools students reveals an inadequacy of Vocabulary Size to support text comprehension. Form One students in Tanzanian Public Secondary schools were found with 3600 vocabulary family which could not enable the learners to read text books neither at minimal nor at optimal level of comprehension. Lastly, the paper recommends for the curriculum design based on learners' vocabulary family levels and academic language to be planned coherently across language supportive practices.

Keywords: Reading Comprehension, Vocabulary Size, and Vocabulary Threshold

1. Introduction

Across many countries in the global south, researchers have shown an increased interest in the study of language competence of the learners with academic excellence (Biseko, Ndabakurane, Sane & Mbwafu, 2024; Mbwafu & Biseko, 2023;2021; Ludewing & Schroder, 2022; Laufer & Ravenhorski, 2010). Whereas, many education systems in the Global South, including Tanzania, expect children to use a language for learning in school that is rarely used outside of schools, which is an international language, English. Using an unfamiliar language for learning harms learning (World Bank Group, 2021) and excludes vulnerable learners. Recognizing this, the government of Tanzania does support the use of Kiswahili language in early years education and English in lower secondary schools, high schools, and higher learning institutions. This means that a significant number of children experience a transition in the language of learning and

teaching (LoLT) in lower secondary school. As in Tanzania, transition to English occurs relatively late at the beginning of lower secondary education, and late language transition coincides with the curriculum splitting into different subjects, with their own set of academic language practices and taught by subject specialist teachers.

Different studies in applied linguistics have reported the significant role played by vocabulary knowledge in language competence. McCarthy (1990) asserts that no matter how well one might learn the grammar of ESL/EFL, no matter how successfully the sounds are mastered, without words (vocabulary) to express a wide range of meanings, communication in a second language cannot happen in any meaningful way. According to Mbwafu and Biseko (2023), vocabulary knowledge is of great significance amongst ESL/EFL learners for effective communicative competence. Moreover, vocabulary is identified askey knowledge for communication among ESL/ EFL learners (Beglar & Hunt, 2005; Coady, 1997; McCarthy, 1990; Schmitt, 2000). The more competent an ESL/EFL learner is in vocabulary knowledge, the more precisely that learner can interact with others and different materials. In light of the above arguments, Kitajima (2001) laments that an individual can only communicate intended meaning if, at all, he/she has words for objects, actions, and concepts that he/she wants to communicate about. Thus, it is apparent that competency in vocabulary knowledge is an indispensable tool that any ESL/EFL learner requires for successful language competence.

Based on the significant role played by vocabulary knowledge in communicative competence, different scholars have attempted to explain what the term vocabulary knowledge means. According to Schmitt (2014), vocabulary knowledge refers to awareness of word organisation, receptive and productive mastery, and fluency. Thus, vocabulary knowledge can be either productive or receptive. Generally, vocabulary knowledge is categorised into depth and breadth knowledge. Breadth of vocabulary knowledge is the quantitative knowledge or number of words that a learner knows at least at a level of meaning, and depth of vocabulary is a learner's qualitative knowledge about words.

While the two categories of vocabulary knowledge are equally important, the focus of the present study was on the vocabulary size. Since, several studies including (Farvardin & Koosha, 2011; Neemeh & Behzad, 2015; Stæhr, 2008) have established that vocabulary size is a good predictor of general proficiency in a foreign language, Laufer and Nation (1995) establish a

significant relationship between vocabulary size and writing skill; Koizumi (2005) establishes a significant relationship between vocabulary size and speaking; Stæhr (2008) reports about the significant relationship between vocabulary size and listening; and Stæhr (2008), Farvardin and Koosha (2011), Neemeh and Behzad (2015) and Masrai (2019) report a significant relationship between vocabulary size and text comprehension.

As far as the relationship between vocabulary size and text comprehension is concerned, it is worth noting that students with a large vocabulary size score higher in text comprehension than other students with a low vocabulary size. However, in Tanzania, different studies report on inadequacies of the learners' VS, which results in poor performances in their text comprehension (Mbwafu & Biseko, 2023; 2021). Given the importance of VS in text comprehension, a question arises: How many words (vocabulary) are needed for the ESL/EFL learner to comprehend academic texts? This study therefore answered the arouse question by examining the VS of the form one public secondary school students to determine whether their VS is enough for them to comprehend their textbooks. The findings are expected to inform education stakeholders and curricula developers about the quantitative vocabulary level (vocabulary size) of the form one students in Tanzania, whereas this knowledge is important to identify the knowledge gap between the actual students' quantitative vocabulary level versus the course books' vocabulary demand for comprehension. This will then motivate educators and curriculum designers to come up with not only appropriate strategies that will promote learners' vocabulary size but also may influence the development of language level policy and teaching and learning strategies that would help learners to comply with the language challenges based on their language competency.

2. Theoretical Framework

This study adopted the threshold hypothesis by Cummins (1976) as the theoretical foundation of the study. Cummins' threshold hypothesis states that there is a minimum level of competency that bilingual ESL/EFL learners should achieve to benefit from the use of that language for educational purposes. About vocabulary-reading comprehension studies, researchers have reported that there is a vocabulary threshold required for ESL/EFL readers to achieve meaningful reading comprehension of a text written in English (Biseko, 2023; Hu & Nation, 2000; Coxhead, 2000; Laufer, 1989; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Masrai, 2019; Nation,

2006; Sen & Kuleli, 2015; Schmitt, Jiang, & Grabe, 2011). These studies report that a reader is supposed to know 98% of running words to achieve optimal comprehension without a need for a dictionary (independent reader), and knowing 95% of running words in a text enables a reader to comprehend texts, although a dictionary will sometimes be needed (dependent reader). These studies have therefore established a vocabulary threshold of 8000-9000 families for ESL readers to achieve 98% of running words in authentic materials and 4000-5000 vocabulary families for 95% of running words. Thus, based on the available previous studies, the threshold hypothesis provided a base to set threshold levels at which reading comprehension is possible. These threshold levels were important for the interpretation of the results.

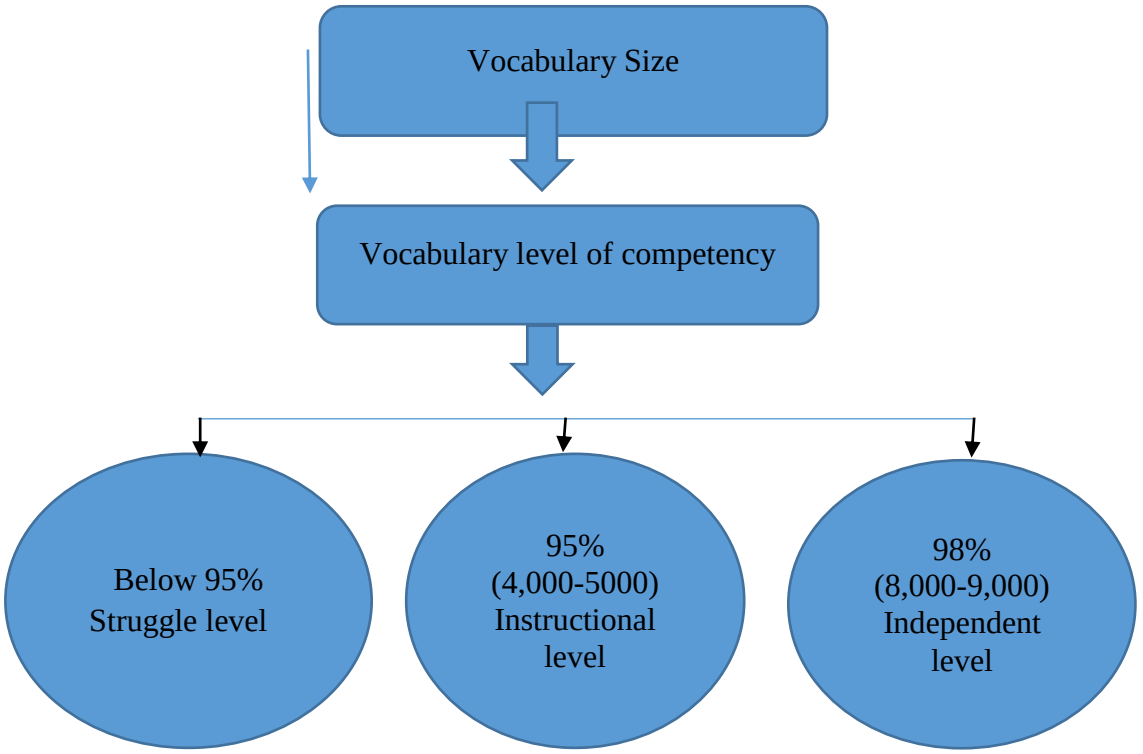

Figure 1: A sketched relationship of variables in the theoretical framework

Source: The researcher (2025)

Vocabulary Threshold

The vocabulary threshold is the minimum number of words a learner needs to know to understand a text or participate in a conversation with a certain level of comprehension. Different scholars define the term vocabulary threshold as the minimal vocabulary size necessary for

"adequate" reading comprehension (Laufer, 2016; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). Vocabulary size is thus said to be a point at which a learner's vocabulary size allows them to grasp the meaning of most written or spoken language effectively. According Qian (2002), vocabulary size is one of the dimensions of vocabulary knowledge that refers to the number of approximated words that an individual knows, at least at the surface aspect of meaning. The term 'vocabulary size' denotes that every speaker of any language possesses a stock of words that are used when communicating. Moreover, researchers report that vocabulary size is a predictor of language competency and it affects an individual's language performance in all language skills (Schmitt, 2008).

Based on the above arguments, the vocabulary threshold for comprehension refers to the level of vocabulary knowledge required for students to effectively understand written or spoken texts. In the context of Tanzanian public secondary school students, this topic involves understanding language proficiency levels, curriculum standards, and the challenges faced in achieving sufficient vocabulary knowledge for academic and general comprehension. Studies on vocabulary comprehension among Tanzanian students indicate that many struggle with academic texts, which often require knowledge of high-frequency and subject-specific vocabulary

Given the importance of vocabulary size, several researchers have embarked on studying this aspect of language;for instance, scholars have made attempts to establish the vocabulary size of native English speakers. Goulden, Nation, and Read (1990) claim that, on average, educated native speakers have a vocabulary size between 15,000-20,000 word-families, Aitchison (2003) reports the vocabulary size of 60,000 word-families, while Treffers-Daller and Milton (2013) suggest that monolingual speakers of English have a vocabulary size of about 10,000 word-families.

Despite this overwhelming number of words reported for native speakers of English, non-native speakers need less than 10,000 word-families to use the language appropriately for different purposes (Nation, 2001). Thus, testing vocabulary knowledge is currently something norm in different education programmes. Schmitt (2008) provides several reasons as to why there is a need to provide vocabulary tests to English education learners. Among the reasons, Schmitt argues that there is a need to assess if a learner has learned the vocabulary he was supposed to

(achievement test). He also presents that sometimes teachers may want to establish the existing gap between learners' knowledge and the expected level (diagnostic test). This is important as it may help teachers to provide special attention to such learners. He also identifies that vocabulary tests can be used as a tool to help teachers place the learner in a proper class (placement test). Apart from these reasons, Schmitt also identifies that other vocabulary tests are provided to assess a general level of proficiency. Other reasons are providing tests for the sake of motivating students to study, and assessing learners' progress in learning new words.

Researchers, therefore, argue that determining learners' vocabulary size is important when planning courses and when placing learners according to proficiency levels (Laufer & Aviad-Levitzku, 2017). In this attempt, learners' vocabulary size is approximated based on scores obtained from some standardised tests. According to Wero, Machmud, and Husain (2021), a valid and reliable way to measure vocabulary size must consider the use of standardised tests that have been proven suitable in terms of validity and reliability. Some of the commonly used tests are the Vocabulary Size Test (VST) by Nation and Beglar (2007), the Vocabulary Levels Test (VLT) by Nation (1983) revised by Schmitt, Schmitt, and Clapham (2001), the Yes/No Test by Meara (1992) and Word Associate Test by Read (1993). Using these tests, researchers in English as a Second/Foreign language have established the vocabulary size of students in different educational programmes. For example, Putra (2009) reported an average of 5,388 vocabulary families among university students in Indonesia, Nizonkiza and Van Dyk (2015) reported 4,500 vocabulary families among first-year university students in South Africa, and Huang (2006) reported 2,838 vocabulary families among Taiwanese students.

Based on the above arguments, this study employed the Vocabulary Size Test (VST) developed by Nation and Beglar in 2007. This VST measures how many words one knows, which is the estimated vocabulary size of the test takers. According to Senturk (2016) vocabulary size test is intended to provide an estimate number of vocabulary items an individual has in his or her lexicon for specific language use. It is argued that vocabulary size is directly related to the ability to use the English language in different ways, based on the function of the test.

Vocabulary Threshold for Reading Comprehension

It is argued that comprehension of a text that someone reads can be achieved at an ideal level when a reader attains a reasonable vocabulary size, often termed the vocabulary threshold (Coxhead, Stevens & Tinkle, 2010; Hsu, 2014). However, the question that arises is 'What level of vocabulary or how much vocabulary is considered "adequate" or "reasonable' for a student to read effortlessly and comprehend a text?' To answer such a question, it can be very subjective, based on the vocabulary size requirements of different texts. For example, children's story books will require less vocabulary size for their vocabulary threshold compared to university academic course books. Thus, texts for different educational levels or disciplines may need different vocabulary sizes to achieve the desired reading comprehension.

Available studies have established that any reader needs to know between 95% and 98% of the running words of a text so that they can be in a position to comprehend the text. Knowing the running words of the text less than 95% inhibits comprehension (Hu & Nation, 2000; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). This denotes that, if a reader does not know between 2% and 5% or 2to 5 words in a paragraph of 100 words, s/he can still comprehend the text. This is possible because the difficult words can be inferred based on the context (Nation, 2006). However, if the number of known words increases in a text, the level of text difficulty decreases. Thus, Laufer and Ravenhorst-Kalovski (2010) suggest two thresholds regarding the percentages of running words in a text. That is 95% for minimal comprehension and dependent readers who will still need the assistance of a dictionary to comprehend some words, and 98% of running words for optimal comprehension of a text. This is for independent readers who can infer the meaning of difficult words they meet in a text.

In line to Laufer and Ravenhorst-Kalovski (2010) different researchers, including Biseko (2023), Laufer (1992), Masrai (2019), Sen and Kuleli (2015), and Schmitt, Jiang, and Grabe (2011) have established that second/foreign language learners need to know 8,000 vocabulary families to comprehend 98% of running words and become independent readers of academic materials. These studies also suggest that a knowledge of 4,000-5,000 vocabulary families is needed to ensure understanding of 95% of running words of academic materials; hence, enabling the reader to manage a reading task with some assistance from dictionaries.

Reading Comprehension

Reading is an important source of knowledge, as much of the knowledge is preserved in written form. According to McMaster et al. (2012), reading comprehension is a complex interaction among automatic and strategic cognitive processes that enables readers to create a mental representation of the text. Snow (2002) defines reading comprehension as the process of simultaneously extracting and constructing meaning through interacting with and involving oneself in a written text. Reading is thus a cognitive process on written discourse or text and active involvement of the reader to construct meaning from the read text.

Previous studies, including Brown (2015) and Nation (2013), affirm that several factors affect the reading and comprehension process. In this list of factors, vocabulary knowledge is among them. Grabe (1997) argues that reading requires sufficient knowledge of vocabulary because it is an interactive process between a reader and the words in a text. Thus, understanding most of the running words in a text assures a reader to comprehend the text as meaning is always carried by words.

The second factor in a list is prior knowledge. According to Brown (2015), prior knowledge helps a reader relate what they read in a text to what they already know. He further asserts that students normally use background knowledge to assist them in guessing the meaning of unknown words from the context. The more a text is familiar to a reader's experience, the better a reader understands what s/he reads. The opposite is also true, that if a text talks about something quite new to a reader, it becomes difficult to grasp the meaning well as the schema fails to connect the story to background experience (John, 2017).

Based on the above scholars' arguments, in summary, there is no doubt that no way readers could comprehend the text without knowing the words in the text. Now, the question that arises is 'Is the vocabulary size adequate or "reasonable" for a form one student to read effortlessly and comprehend their textbooks?'

Research Gap

Since there is a transition from Kiswahili Medium of Instruction to English Medium of Instruction among secondary schools students particularly Form One Students, the reviewed

literature has provided a lot of valuable insight about the existing relationship between vocabulary size and text comprehension among ESL/EFL learners. According to Cummins (1976) there is a minimum level of competency that bilingual ESL/EFL learners should achieve to benefit from the use of that language for educational purposes. Whereas, vocabulary size between 4,000 and 9,000 is recommended to enable an ESL/EFL learner to understand 95% - 98% of the running words of authentic texts, including academic ones. However, in Tanzanian public secondary schools little has been done to determine the vocabulary size of ESL/EFL learners, particularly the form one students. The available study investigates the vocabulary demands of primary and secondary school textbooks to assess their adequacy for learners transitioning from primary to secondary education (Biseko, Ndabakurane, Sane & Mbwafu, 2024). Another study by Barret et al. (2024) compares language skills in the Form I Biology curriculum with those in the Standard 7 English curriculum. It analyses the TIE Form I Biology textbook for readability and language support features to identify curriculum changes needed to align language competencies in teaching materials. Since the previous studies did not examine the vocabulary size of the form one students to prove their hypotheses, this study therefore examined the vocabulary size of the Tanzania Public Secondary School Students to prove the hypothesis by Biseko, Ndabakurane, Sane and Mbwafu (2024) that for form one students to read effortlessly and comprehend their textbooks, they need a vocabulary threshold of 3,000 words for 95% to comprehend English textbook, and 7,000 words for the Biology textbook. For optimal comprehension, 5,000 words are necessary for the English textbook, and 13,000 words for the Biology textbook.

3. Methodology

Research Approach

The study used a quantitative research approach to examine the Vocabulary Size of the form one public secondary school students in Tanzania and reading comprehension. Punch (2006) defines the quantitative research approach as empirical research that uses numbers to establish facts about the research objectives at hand. This study is based on objective reality and a deductive approach.It is worth mentioning that objectivity in this study means that the collection, analysis, and interpretation of the data were not based on individual points of view, personal feelings, or opinions, but through objective collection, analysis, and interpretation of numerical data.

Further, the study was deductive-based as it was designed to test the conclusion by Biseko, Ndabakurane, Sane, and Mbwafu (2024) that form one students in Tanzania have less vocabulary size compared to the required vocabulary reading demand of the textbooks. According to Brotherton (2013), deductive research is a research approach in which the researcher begins with a general statement or hypothesis and designs the study to test the hypothesis. Thus, quantitative methods are suitable for measuring learners' performance objectively using standardized instruments like the Vocabulary Size Test (VST) and comprehension tests.

Research Design

This study used the case study design to delineate the context of the study. According to Walker (2005), research design is defined as a plan that shows how, when, and where the data for a study are collected and analysed. In line with the above, as far as the present study is concerned, the phenomenon under study was the 'vocabulary thresholds for text comprehension' and the definite context was 'form one students in Tanzania public secondary schools year, 2025'.

Area of the Study

This study was conducted in public secondary schools. It comprises five mainland regions in Tanzania: Dodoma, Dar es salaam, Arusha, Morogoro, and Mwanza. The regions were selected inclusively with the schools found in regional centres (urban schools). The criteria for urban schools was English language exposure among the ESL/EFL public secondary school learners and students are assumed to be aware of the English language compared to those students in remote regions. In line to the above argument, Li et al. (2025) report that, in urban regions children often show stronger English language outcomes than rural children due to the socio-economic and environmental factors. Moreover, studies repeatedly find that differences in family socio-economic statuses, home-literacy, and early language stimulation account for much of the urban advantage second language awareness.

Target Population, Sample Size, and Sampling Technique

The target population for this study was all Form One students in Tanzania public secondary schools who were registered in the 2025 academic year. The researcher employed purposive sampling. According to Berg (2001), "purposive sampling sometimes is called judgmental sampling technique whereby researchers use their special knowledge, expertise or prior

experience about some groups to select subjects to represent the population". Thus, purposive sampling was useful for the sampling of form one students based on the fact that they are abruptly shifting from the Kiswahili medium of instruction to English when entering form one. Table 1 provides the registration status of the students (2025)

Table 1: Form One registered students

#	**Region**	**School**	**Stream**	**Registered students**
1	AD	A	1	47
			2	43
			3	45
		B	1	46
			2	44
			3	48
		C	1	40
			2	25
			3	-
2	BD	A	1	44
			2	47
			3	36
		B	1	49
			2	35
			3	40
		C	1	45
			2	45
			3	51
	CA	A	1	49
			2	45
			3	35
		B	1	49

3			2	54
			3	55
		C	1	39
			2	46
			3	47
4	DM	A	1	47
			2	43
			3	45
		B	1	46
			2	44
			3	48
		C	1	40
			2	20
			3	-
5	EM	A	1	44
			2	17
			3	36
		B	1	49
			2	35
			3	40
		C	1	45
			2	45
			3	51

Source: School registration books of 2025

To get the sample size of the participating students, the formula for determining the sample size by Yamane (1967) was used [$\mathbf{n = N/ (1 + N(e)^2}$]. The formula was used to calculate the sample size from each region, which was later summed up to get the total sample size of the study.

According to Tejada and Punzalan (2012), Yamane's formula is appropriate if the population of the study is definite and the confidence coefficient is 95%. Since the number of students was definite, this formula was therefore significantly appropriate to get the sample size of 282 form one students out of a population size of 1844. Moreover, students were selected purposive based on the year of study which was only form one students in selected schools.

Data Collection Techniques

This study employed Vocabulary Size Test (VST) by Nation and Beglar (2007) for data collection.

Vocabulary Size Test (VST)

VST is a vocabulary research tool designed by Nation and Beglar to measure learners' vocabulary size. It was developed to provide a reliable and comprehensive measure of learners' vocabulary size (Nation & Beglar, 2007). VST was considered a genuine tool for measuring vocabulary size of the participants in this study because it is the only tool that is currently recommended as a valid and reliable assessment for measuring vocabulary size (Beglar, 2010; Laufer & Aviad-Levitzky, 2017; Leeming, 2014; McLean, Hogg & Kramer, 2014).

VST is available online at http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx. Words for the test are based on a word list developed by Nation from the British National Corpus (BNC). Nation created word lists for the first 14,000 word families of English based on the BNC. Questions for the VST are multiple-choice formats. The test has 10 test items for each band of 1000 vocabulary families. This study employed only a part of VST to measure learners' vocabulary level of the first 1000 to the eighth 1000 word families. Thus, only 80 items of the original VST was used because each vocabulary family is represented in VST by only 10 questions. Items from these word-families were intentionally used because researchers (see Hacking, Rubio, Tschirner, 2018; Laufer & Ravenhorst-Kalovski, 2010; Nation 2006; Nation & Beglar, 2007; Nizonkiza, & Van Dyk, 2015; Schmitt, 2008) recommend that ESL/EFL learners who are competent in the first to eighth word families can independently and successfully carry activities like reading varieties of texts including textbooks, novels, scholarly articles, newspapers and reports of all kinds. Thus, it was logical to assess the participants of the present study in these word families rather than testing them in all 14^{th} word families.

Participants were required to attempt the 80 multiple choice questions for 40 minutes as designers recommend one minutes for two questions. To reduce the effect of guessing, participants were required to use dummy names and attempt only questions that they were aware of while leaving unattempted all items that they were not sure of. After the students had answered the test, the researcher marked their works manually. Since the VST had only 10 questions in each vocabulary family, the researcher scored each vocabulary family over 10 marks while the whole test were marked over 80 scores. Lastly, each student's scores were entered in SPSS page indicating scores for each level and the sum of scores for all 8 levels. Using the VST, participants' scores can be easily interpreted and compared with scores attained by other learners in different studies worldwide (Beglar, 2010)

Data Analysis

Like other quantitative linguistic studies, this study was intended to depict the complexity of language use based on the form one student's vocabulary knowledge. Therefore, all the data that were collected through the VST and analysed quantitatively. The data that were collected through VST were analysed using SPSS version 23. The suggestion put forward by Laufer and Nation (2007) on how to interpret VST results was used to find each student's vocabulary size in each vocabulary level. Laufer and Nation suggest that, "because there are ten items in each 1000-word level, each item in the test represents 100-word families;a test taker's score needs to be multiplied by 100 to get their total vocabulary size. Thus, after entering students' scores in the SPSS page, each student's scores in each vocabulary family were multiplied by 100 to get the vocabulary size of each student in the particular family. Lastly, descriptive statistics were computed to depict the students' mean, maximum, and minimum vocabulary knowledge.

4. Results And Discussions

Table 2 shows vocabulary size scores of the students, and Table 3, on the other hand, illustrates cumulative descriptive statistics portraying mean vocabulary thresholds.

Table 2: Students' VST scores in each vocabulary family (N = 1844)

Vocabulary families	Minimum score	Maximum score	Mean score	Rounded Mean score

The 1st 1000	1	9	7.34	7
The 2nd 1000	1	9	6.44	6
The 3rd 1000	1	8	5.36	5
The 4th 1000	1	8	5.02	5
The 5th 1000	0	8	4.42	4
The 6th 1000	0	7	3.00	3
The 7th 1000	0	7	3.01	3
The 8th 1000	0	6	3.09	3

Source: Researcher (2025)

Table 4.1 revealed that the maximum score was ten (9) marks in the first to second vocabulary families nine (8) marks in the 3rd to 5th vocabulary family, (7) marks in the 6th to 7th vocabulary family, and (6) marks in the 8th vocabulary family. The data show that no student got all the questions right in all the vocabulary family. In general, the data for mean scores show that students scored more than half of the total marks in the first to fourth vocabulary families. This implies that these students had developed a good command of the 1st to the 5th vocabulary families compared to the knowledge they had in the fifth to eighth families. This suggests that vocabulary acquisition/learning among the participants is affected by the frequency level of the words. Thus, high-frequency words are acquired/learned first before the mid/low-frequency words.

Table 3: Estimated vocabulary size of the participants (N = 1844)

Vocabulary Families	**Rounded mean score**	**Estimation**	**Vocabulary size**
The 1st 1000	7	7 x 100	**700**
The 2nd 1000	6	6 x 100	**600**
The 3rd 1000	5	5 x 100	**500**
The 4th 1000	5	5 x 100	**500**
The 5th 1000	4	4 x 100	**400**
The 6th 1000	3	3 x 100	**300**
The 7th 1000	3	3 x 100	**300**

The 8th 1000	3	3 x 100	**300**
Vocabulary size	36	36 x 100	**3,600**

Source: The researcher (2025)

Table 3 showed that, measured at the 8000-vocabulary family, the average participants' vocabulary size was 3,600. The data also show that the participants had a larger vocabulary size in the high-frequency vocabulary families (the first 2000 vocabulary family). The vocabulary size seems to decrease as the more one goes to the low-frequency vocabulary families. Table 3 also reveals that students had at least a vocabulary size of 500 hundred and above in the 1st to 4th vocabulary families. In other families, their average vocabulary size was below half of the total 1000 words of a single vocabulary family. This implies that the students in the sample had achieved competency in the first to the fourth vocabulary family compared to the competency they had in other levels. Even though the descriptive statistics in Tables 2 and 3 have shown the performance of participants at each level, they lack information on the frequency of performance.

Moreover, as the answer to the hypothesis by Biseko, Ndabakurane, Sane and Mbwafu (2024) who assert that a vocabulary threshold of 3,000 words is needed for 95% which is the minimal comprehension of the form one English textbook, and 5,000 words for 98% which is the optimal comprehension and vocabulary threshold of 7,000 words for 95% to comprehend the Biology textbook and 13,000 words for 98% is concerned, the findings of this research also support the above study by Biseko, Ndabakurane, Sane and Mbwafu (2024) that some students are entered in form one with a vocabulary threshold that cannot support them to read and comprehend the textbooks designed for their studies due to the abrupt increase in vocabulary demands during the transition. Thus, this situation inhibits such students from using their textbooks to acquire knowledge for sustainable development.

Moreover, findings revealed that majority of the participants had developed vocabulary competency to the level that would only help them to read and comprehend most of the textbooks below, even the minimal level of comprehension that is 3600 words out of 4000 words. However, diversification of language competency among learners might be due to the language background as some might come from primary English medium and others from

Primary Kiswahili medium of instructions. Additionally, a study by Biseko, Ndabakurane, Sane and Mbwafu (2024) examined vocabulary coverage of form one English language textbook in relationship to form one Biology textbook which might have resulted into great discrepancy in vocabulary coverage as Biology textbooks contain technical terms (Barret et al., 2024). Thus ranges in between 3000 - 5000 vocabulary family which reflect the theoretical framework (Biseko, 2023; Hu & Nation, 2000; Coxhead, 2000; Laufer, 1989; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Masrai, 2019; Nation, 2006; Sen & Kuleli, 2015; Schmitt, Jiang, & Grabe, 2011)

5. Conclusion

This study determined whether the vocabulary size of the form one students in Tanzanian public secondary schools was sufficient for them to comprehend the textbooks used for their course of study. The findings revealed that the vocabulary size of most of the form one students in the sample could not enable them to read and comprehend their textbooks, neither at a minimal nor an optimal level of comprehension. The findings revealed that Form one students were still at the struggle level of vocabulary size, which was below 4000 words. This implies that most of the students in the sample may need the assistance of a dictionary to comprehend the textbooks. From the findings, there is a wake-up call for the curriculum reforms and teacher education training strategies to promote vocabulary levels among form one students in Tanzania. Lastly, there is a need to plan for vocabulary learning from primary schools.

Declaration

I declare that I have no conflict of interest related to this study.

References

Barrett et al. (2024). Language learning across transition in the language of learning and teaching: An analysis of the Tanzanian curriculum. Bristol Working Papers in Education #03/2024, May 2024, School of Education, University of Bristol.

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

Berg, B. (2001). Qualitative Research Methods for the Social Sciences. *Journal of Teaching Sociology*, 1(1), 5-18.

Biseko, J. M. (2023). Vocabulary size and comprehension of academic texts by Tanzanian university students: An exploratory study. *Southern African Linguistics and Applied Language Studies*, *41*(4), 496-513. https://doi.org/10.2989/16073614.2022.2151480

Biseko, J.M, Ndabakurane, J, Macha, E.S. &Mbwafu, F (2024).Understanding Vocabulary Demands in Textbooks during the Primary to Secondary Transition: Implications for Textbook Writing and Pedagogy. The Rwandan Journal of Education, 7 (3), 53-69

Coady, J. (1997). L2 vocabulary acquisition: A synthesis of research. In J. Coady& Th. Huckin(Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp.273-91). Cambridge: Cambridge University Press

Coxhead, A., Nation, P., &Sim, D. (2015).The vocabulary size of native speakers of English in New Zealand secondary schools. *NewZealand Journal of Educational Studies*, *50*(1), 121–135.University Press.

Coxhead, A., Stevens, L.,& Tinkle, J. (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics*, 16 (2), 37-52.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, 9, l-43. DOI: 10.5281/zenodo.11233387

Farvardin, M. T., &Koosha, M. (2011). The role of vocabulary knowledge in Iranian EFL students' reading comprehension performance: Breadth or depth? *Journal of Theory and Practice in Language Studies*, (11), 1575-1580.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 7(11), 42-54. https://doi.org/10.1093/applin/11.4.341

John, A. (2017). The use of textbook illustrations in the teaching of reading comprehension: a study of Secondary Schools in Kathiani District-Kenya. *Journal of Arts Humanities and Social Sciences*, 2(8), 1-14.

Kitajima, R. (2001). The effect of instructional conditions on students' vocabulary retention.*Foreign Language Annals*, 34, 470–482.

Koizumi, R. (2005). Attainment levels of speaking ability of Japanese junior high schoolstudents. *JLTA (Japan Language Testing Association) Journal*, 7(11), 31-50.

LauferB.,&Aviad-Levitzky T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal*, 101, 729-741.

Laufer, B. (1989). What percentage of text lexis is necessary for comprehension? https://www.researchgate.net/publication/258884242

Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *TESOL Quarterly*, 47(4), 867–872.

Leeming P. (2014). Analysis of the vocabulary size test. *The Language Education Journal, Foreign Language Edition*, 2(5), 73-87.

Ludewig, U., Hübner, N., & Schroder, S. (2022). Vocabulary, text coverage and the lexical threshold in elementary school reading comprehension assessment.*Reading and Writing. Online ahead of print.*https://doi.org/10.1007/s11145-022-10385-0

Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid-, and low-Frequency vocabulary knowledge. *SAGE Open*, 9(2), 43-63. https://doi.org/10.1177/2158244019845182

Mbwafu, F.A. & Biseko, J.M. (2021). Vocabulary Size and Learners' Performance across Levels of Text Comprehension: A Case of First-Year Students at the University of Dodoma. *JIPE 13*(2): 51-84.

Mbwafu, F.A. & Biseko, J.M. (2023). Madesa or books! Using a Coverage Comprehension Model to Assess University Students' Ability to Comprehend Reference Books.*Huria Journal, 30*(1), 165-188.

Mclean, S. (2021). The coverage comprehension model, its importance topedagogy and research, and threats to the validity with which it isoperationalized.*Reading in a Foreign Language, 33*(1), 126–140.

McMaster K. L., van den Broek P., Espin C. A., White M. J., Rapp D. N., Kendeou P., &Carlson S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, 22, 100-111.

Meara, P. (1992). *EFL Vocabulary Tests*. Swansea: University College Swansea, Centre for Applied Language Studies.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/ La Revue canadienne des languesvivantes,* 63(1), 59-82.http://dx.doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2013).*Learning vocabulary in another language*. Cambridge: Cambridge

Nation, I.S.P. and Beglar, D. (2007).A vocabulary size test.*The Language Teacher*, 31(7), 9-13.

Nizonkiza, D., & Van Dyk.T. (2015). Academic literacy of South African higher education level students: Does vocabulary size matter? *Stellenbosch Papers in Linguistics,* 44: 147-174. doi:10.5774/44-0-159

Punch, K. (2006). *Developing Effective Research Proposals*.London:Sage. https://books.google.sm/books?id=PFHV8oajR7gC

Putra, Y. (2009). The relationship between vocabulary size and writing ability of the English language laboratory students at UniversitasPelitaHarapan.(Unpublished Doctoral Thesis).UniversitasPelitaHarapan, Karawaci.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective.*Language Learning,* 52 (3), 513-536.

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press

Schmitt, N. (2000). *Vocabulary in language teaching*.Cambridge University Press. Cambridge.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research,* 12(3), 329- 363.

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

Schmitt, N., & Schmitt, D. (2014).A reassessment of frequency and vocabulary size in L2 vocabulary teaching.*Language Teaching*, 47(4), 484-503.

Schmitt, N., Grabe, W., & Jiang, X. (2011).The Percentage of words known in a text and reading comprehension.*The Language Modern Journal*, 95(1), 26-43.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing,* 18(1), 55-88.

Şen, Y., &Kuleli, M. (2015*).*The effect of vocabulary size and vocabulary depth on reading in EFL context.Procedia- *Social and Behavioral Sciences, 199, 555-562.* bspro.2015.07.546.

Senturk, B. (2016). Self-regulation strategies and vocabulary size of EFL Turkish University Students. *Journal of Procedia-Social and Behavioral Sciences*, 232, 90-97. https://doi.org/Doi:10.1016/j.sbspro.2016.10.023.

Stæhr, L. (2008). Vocabulary size and the skills of listening, reading and writing.*Language Learning Journal*, 36(4), 139-152.

Tejada, J. J., & Punzalan, J. R. B. (2012). On the misuse of Slovin' s formula. *The Philippine Statistician*, *61*(1), 129-136.

Treffers-Daller, J., & Milton, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151-172.

Walker, W. (2005). The strengths and weaknesses of research designs involving quantitative measures.*Journal of Research in Nursing*, 10, 571-582. https://doi.org/10.1177/136140960501000505

Wero, Y., Machmund, K.,& Husain, N. (2021). The Study on Students' vocabulary Size.*Jambura Journal of English Teaching and Literature,* 2(4), 22-34.

World Bank Group. (2021). *Loud and Clear: Effective Language of Instruction Policies for Learning, A World Bank Policy Approach Paper*.

Yamane, T. (1967). Statistics: An Introductory Analysis (2nd ed.). Harper and Row

Yin, R.K., (1984). *Case study research: Design and methods*. Beverly Hills, Calif: Sage Publications.