Understanding Vocabulary Demands in Textbooks during the Primary to Secondary Transition: Implications for Textbook Writing and Pedagogy

John Misana Biseko[1], Jesse Ndabakurane[1], Eliakimu Sane Macha[1] & Frida Mbwafu[2]
[1]University of Dodoma-Department of Foreign Languages and Literature
[2]Open University of Tanzania-Department of Linguistics and Literary studies
Corresponding author's Email: jessetuisabe@gmail.com

Abstract

The relationship between vocabulary knowledge and reading comprehension has been extensively researched in the realm of English as a Foreign Language (EFL) learning. Such studies have been reporting a positive correlation between the two variables. The present study investigated the vocabulary demands of two textbooks at the edges of primary – secondary transition. The study sought to answer one research question: Does the vocabulary coverage in the English for Standard Seven textbook adequately prepare learners for comprehending secondary school textbooks? Using a corpus analysis as a research design, two textbooks - English for Primary Standard Seven and Biology for Secondary Form One students were analysed. The findings showed a significant gap in the vocabulary demands of the two sampled textbooks, suggesting a burden faced by learners transitioning from primary to secondary school. Following these findings, a call for designing coherent curriculum materials for primary–secondary transition learners is emphasised.

Keywords: Vocabulary coverage, vocabulary threshold, primary-secondary transition, reading comprehension, textbook accessibility

Introduction

Vocabulary knowledge refers to knowledge of words, which consists of knowing the word forms, meaning, and word use in different contexts (Nation & Hunston, 2013). Several studies have reported that vocabulary knowledge is pivotal for individual competency in various language skills. Among others, Horst (2013) believes that vocabulary knowledge underpins all language proficiency and it is the foundation upon which acquisition of syntax, pragmatics, and other aspects of language crucially depends. Nation (2001) views that, learners in either second or foreign language contexts should acquire much vocabulary as a prerequisite for successful target language learning. Lewis (2008) contends that acquiring sufficient vocabulary is the most important task language learners should strive for. Additionally, Schmitt (2008) presents that successful vocabulary learning is essential for mastering a target language, while Lewis (1993) equates vocabulary knowledge with the heart of language learning.

Given the significance of vocabulary in language acquisition, a question arises: "How much vocabulary does English as a second/foreign language (ESL/EFL) learner need?" Nation and Waring (1997) identify three factors to consider when addressing this question: the number of words present in English, the number of words native speakers know, and the number of words required for specific tasks that ESL/EFL users want to perform. This paper focuses on the third factor, as it examines the relationship between vocabulary knowledge and comprehension of school textbooks. Previous studies (Ehsanzadeh, 2020; Ludewig et al., 2022; Schmitt, 2010) have demonstrated that ESL/EFL users require a specific vocabulary size to engage in language-related tasks. In particular, Laufer and

Ravenhorst-Kalovski (2010), Nation (2006), and Stæhr (2008) justify that ESL/EFL learners need specific vocabulary sizes to comprehend texts and that without this requisite vocabulary, comprehensibility of academic materials becomes challenging. This finding is crucial to consider during the primary-secondary transition, particularly in Tanzania, where English is only a subject in primary schools but a language of learning and teaching in secondary schools. Thus, comparing the vocabulary demand of the primary school English textbook and the vocabulary demand of secondary school textbooks is extremely important for assessing the smooth transition of the learners.

The present study compared the vocabulary demand of English language textbooks for primary standard seven learners and that of Secondary school Biology Form one students' textbook. This comparison was crucial for three reasons. First, the findings were needed to reveal potential gaps in vocabulary demand that learners might encounter during their primary-secondary transition. Second, this comparison was important as it attracts curriculum developers and other educational stakeholders to note gaps revealed in vocabulary demands for learners at the two edges of the education transition. Such awareness is important for designing a coherent curriculum that aligns vocabulary demand throughout learners' education journey; thus, ensuring that the vocabulary instruction provided in primary education suits learners for a smooth transition to secondary education, where English serves not only as a subject but also the language for engaging with content across all subjects. Third, the findings on vocabulary demands in these textbooks could inform professional development for both pre-and in-service teachers as they would call for equipping them with strategies to teach vocabulary effectively across different subjects. Thus, addressing specific challenges that learners may encounter in understanding textbooks for secondary education is essential.

Literature Review

Word Frequency and Vocabulary Families

Vocabulary acquisition is central to second language learning and effective communication (Decarrico, 2001). Alqahtani (2015) asserts that ESL/EFL learners often struggle with communication if they lack sufficient vocabulary. This challenge is particularly pressing for learners in multilingual African countries, where English is the language of learning and teaching; thus, developing a robust vocabulary is essential for effective knowledge access. Nevertheless, a significant challenge for curriculum designers is identifying the types and amount of target vocabulary learners should acquire. This is more challenging given that English comprises about 88,500 to 114,000 words (Nagy & Anderson, 1984; Goulden et al., 1990). Despite this extensive lexicon, Nation (2001) observes that ESL/EFL learners do not require such a large vocabulary to communicate successfully; instead, they only need a few essential words to achieve the purpose. This finding has attracted interest in researching the nature and quantity of vocabulary necessary for effective communication, revealing that not all words are equally useful (Nation & Waring, 1997). Milton (2009) elaborates that some words are used more frequently than others, a phenomenon that Douglas (2016) refers to as the "lexical frequency principle".

Using the lexical frequency principle, researchers have developed word lists based on corpus data from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These lists arrange words in descending order based on the frequency use criteria; thus, identifying commonly and rarely used words in various contexts. Consequently, vocabulary is divided into bands of 1,000 words, known as vocabulary levels or families. The first vocabulary family consists of the most frequent 1,000 words, followed by the second (1,001 to 2,000) vocabulary family, the third (2,001 to 3,000) vocabulary family, and the following bands continue this way. According to Schmitt and Schmitt (2014), vocabulary families are categorised into high-frequency families (first to third 1,000), mid-frequency families (fourth to ninth 1,000), and low-frequency families (beyond the ninth). High-frequency families are the basic or everyday vocabulary, as they are commonly known and frequently encountered; however, familiarity decreases from mid to low families. Based on these categories, texts with a higher proportion of words from the high-frequency vocabulary families are generally easier to comprehend than texts with a higher proportion of words from the low-frequency vocabulary families (Chujo & Utiyama, 2005; Mbwafu & Biseko, 2023).

National and Beglar (2007) identify 14-word families of 1,000 words each, while the lexical profiles available on Nation's website (http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx) and Cobb's site (http://lextutor.ca) divide words into 20-word families based on (BNC). Additionally, Nation (2017) presents a list of words organised into 1-25k frequency levels, accessed at https://lextutor.ca/vp/comp/. Table 1 provides examples of words from different word families adapted from Nation (2006).

Table 1

Examples of words from different levels/vocabulary families from LOB Corpus (1,000–14,000)

Frequency Level	Example of Words
1,000	the, history
2,000	accommodate, prefer
3,000	digest, receipt
4,000	elastic, thread
5,000	locker, tranquil
6,000	diligent, undertake
7,000	fossil, jagged
8,000	abhor, obtrusive
9,000	remorse, wrench
10,000	barricade, pigment
11,000	glitzy, scam
12,000	epitome, resonate

13,000	outdo, tipsy
14,000	secede, yearbook

Source: Nation (2006)

It is worth noting that evidence from numerous studies indicates that any language corpus consists of several vocabulary families. However, the portion of words from the first vocabulary family is consistently greater than words from the other families. Nation (2006) depicts this idea using analysis results from various written works, shown in Table 2.

Tale 2

Average coverage and range of coverage of a series of word-level

Vocabulary family	Number of levels	Approximate written coverage
1st 1,000	1	78 – 81
2nd 1,000	1	8 – 9
3rd 1,000	1	5 – 3
4th – 5th 1,000	2	3
6th – 9th 1,000	4	2
10th – 14th 1,000	5	< 1
Proper nouns	1	2 – 4
Not in the list	1	1 – 3

Source: Nation (2006)

Vocabulary threshold, Vocabulary Coverage and Text Comprehension

Vocabulary knowledge is significantly associated with text comprehension among ESL/EFL learners (Staehr, 2008). Several studies have reported that vocabulary knowledge significantly influences learners' achievement levels in reading comprehension (Laufer & Ravenhorst-Kalovski, 2010; Şen & Kuleli, 2015). Research by Mbwafu and Biseko (2021), Qian (2002), Stæhr (2008), and Milton, Wade, and Hopkins (2010) demonstrate that the correlation between vocabulary knowledge and reading comprehension often exceeds 50%, suggesting that vocabulary knowledge plays a central role in text comprehension. Specifically, Qian (2002) reports a correlation of r = .74, p < .01 among 217 individuals for whom English was a second language. Similarly, Stæhr (2008) reports a correlation of r = .83, p < .01, whereby vocabulary knowledge explained about 72% of the variance in reading comprehension scores among 88 Danish learners of English as a foreign language. Milton, Wade, and Hopkins (2010) report a correlation of r = .70, p < .01, with vocabulary explaining about 50% of the scores of learners from various L1 backgrounds, including Arabic, Chinese, Japanese, and others from European countries. Mbwafu and Biseko (2021) report a correlation of r = .673 between vocabulary size and literal comprehension, r = .628 between vocabulary size and critical comprehension,

and r = .606 between vocabulary size and inferential comprehension among 230 first-year university students in Tanzania. Additionally, the r^2 results from this study indicated that approximately 45% of participants' scores in literal comprehension, 37% in inferential comprehension, and 39% in critical comprehension were explained by vocabulary size.

Studies investigating the relationship between vocabulary knowledge and reading comprehension have introduced three interrelated concepts: 'vocabulary threshold', 'vocabulary/text coverage', and 'text comprehension'. The vocabulary threshold denotes the minimum vocabulary size required for reasonable comprehension (Nation, 2001). Nation distinguishes two perspectives regarding vocabulary threshold. These are the strong and the weak views of the vocabulary threshold. On the one hand, the strong view considers vocabulary size a "deterministic boundary," implying that readers comprehend a text if their vocabulary meets or exceeds this level; otherwise, successful comprehension is unlikely. On the other hand, the weak view regards the vocabulary threshold as a "probabilistic boundary," suggesting that possessing vocabulary at this level enhances the likelihood of comprehension. Despite these differences, both perspectives advocate that an adequate vocabulary size is crucial for successful text comprehension.

Vocabulary coverage is "the percentage of running words in the text known by the reader" (Nation, 2006). It represents the intersection between the words in a given text and a reader's vocabulary knowledge (Ludewig, Hübner & Schroeder, 2022). This notion is closely linked to vocabulary threshold, as a reader's coverage can determine their ability to understand the necessary threshold for comprehension. Hsu (2011) argues that a lexical threshold is contingent upon predetermined coverage points, which indicate the vocabulary size required for text comprehension. High vocabulary coverage facilitates the achievement of this threshold, thereby enhancing the likelihood of successful reading comprehension. Subsequently, researchers such as Ng et al., (2020), Hsu (2014), and Laufer and Nation (1995) assert that vocabulary coverage serves as an effective metric for evaluating the likelihood of text comprehension among students with varying vocabulary levels.

Several studies have analysed vocabulary coverage in different reading materials to determine thresholds and evaluate text accessibility relative to learners' vocabulary knowledge. Research by Hu and Nation (2000), Laufer (1992), Laufer and Ravenhorst-Kalovski (2010), and Mbwafu & Biseko (2023) indicate that the percentage of known words in the text largely determines text comprehension. Thus, an increase in unknown words obstructs comprehension, while fewer unknowns allow readers to infer meanings from context. These studies suggest that readers should know 95% to 98% of the words in a text to achieve comprehension. This proposes that 2% to 5% of unknown words, nearly 2 to 5 in a 100-word paragraph, can be tolerated. Additionally, these studies report that a vocabulary threshold between 4,000 and 9,000-word families is sufficient for comprehending 95% to 98% of texts, which is essential for effective reading comprehension.

In her study, Laufer (1992) reported that a vocabulary threshold of 3,000-word families provided 95% coverage, allowing learners at Haifa University to get a pass mark in reading comprehension. Hirsh and Nation (1992) found that a threshold of 5,000-word families was required for comprehending difficult novels. Hu and Nation (2000) discovered that no learners could adequately comprehend fiction texts at 80% coverage; though a few could comprehend at 90% and 95% and most learners understood the text at 100% lexical coverage. Besides this finding, the regression model by Hu and Nation (2000) suggested that 98% coverage was a probabilistic threshold. Thus, Nation (2006) points out the 8,000 to 9,000-word families as the threshold required for unassisted reading comprehension. However, Laufer and Ravenhorst-Kalovski (2010) propose two lexical thresholds: the knowledge of 8,000-word families to achieve 98% lexical coverage, for optimal text comprehension and the knowledge of 4000-5000-word families to achieve 95% lexical coverage, for minimal text comprehension.

Following evidence from various studies, the current consensus is that readers should know 98% of the text's running words to achieve optimal text comprehension. In other words, readers can still have optimal comprehension if they encounter two unknown words in a paragraph of 100 words. The vocabulary knowledge of 8,000 to 9,000-word families can assure optimal comprehension achievement. In contrast, readers should know 95% of the text's running words to attain minimal text comprehension. In other words, readers can still comprehend the text at this level if they encounter five unknown words in a paragraph of 100 words. To achieve this comprehension level, knowledge of 4,000-to-5,000-word families is required. However, Laufer (2021) emphasises that simplified texts may require a vocabulary size lower than 5,000-word families; however, readers should know 95% to 98% of running words for text comprehension.

Although the reviewed studies provide invaluable insights into the relationship between vocabulary knowledge and reading comprehension (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Mbwafu & Biseko, 2021; Şen & Kuleli, 2015; Sun & Dang, 2020), they also reveal one significant gap in the literature, which is the lack of investigation of vocabulary demands of textbooks for students in successive grades, especially those transitioning from primary to secondary education in Multilingual communities like in Tanzania, where primary-secondary education transition also means a transition of the language of learning and teaching. This study was designed to fill this gap. It examined the vocabulary demands of two textbooks representing key learning materials during the primary-secondary transition. In particular, the study analysed the Primary English textbook for Standard Seven learners and the Secondary Biology textbook for Form One students in Tanzania. The selection of the English textbook was based on the understanding that it was a key learning material at the pinnacle of English instruction at the primary level, preparing learners for English medium education upon transition to secondary education. On the other hand, the selection of the Biology textbook was based on the research findings, that Biology presents the most linguistically challenging scientific content (Barrett et al., 2014). Specifically, the study answered one research question: Does the vocabulary coverage in the English for Standard Seven textbook adequately prepare learners for

comprehending secondary school materials?

Method

Research Design and Approach

The present study adopted a corpus-based design and quantitative approach to examine the vocabulary coverage of the two textbooks. A corpus approach to text analysis consists of the collection and analysis of *"pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research"* (Sinclair, 2004, p.19). This design was considered genuine for this study based on its capability to reveal the vocabulary demand of any text.

Corpus Compilation

Compilation of corpus data marked the first and essential task laying the foundation of this study. The task began with identification and collection of the textbooks. In Tanzania, textbooks from different publishers exist. Nevertheless, the textbooks seek approval from the government authority before they are exposed for use. In this study, the two textbooks were from the same publisher. The availability criterion was used to sample the textbooks from this publisher as it was mandatory to get the two textbooks from one publisher in the local market. Some publishers had one of the two textbooks in the local market and thus were not selected for analysis. Generally, this was a random selection of the materials. These textbooks were scanned and converted into a machine-readable format using Optical Character Recognition software called Fine Reader   https://pdf.abbyy.com/blog/focusing-on-pdf. This was followed by a document cleaning stage that involved a manual inspection of the two documents to ensure that all spelling errors that might have occurred when converting the documents into a machine-readable format were corrected. The document cleaning also involved the removal of text features that were not relevant to the present study. The features are acknowledgements, edition notice, table of contents, tables, pictures, numbers, symbols and references cited. Lastly, the two files were separately saved ready for analysis.

Data Analysis

The corpus data of the two textbooks were analysed quantitatively using the latest version of Web Vocabulary Profiler (VP-Compleat Lexical Tutor), available at https://www.lextutor.ca/vp/comp/. VP-Compleat Lexical Tutor is a well-established and widely used tool for assessing the vocabulary coverage of texts. It compares the input corpus with the words in the BNC/COCA. The VP-Compleat Lexical Tutor arranges the vocabulary in the input corpus into 25 distinct frequency bands or families, making comparative analyses of texts possible. The selection of VP-Compleat Lexical Tutor over alternative analysis tools, such as AntWordProfiler, was based on three factors. First, VP-Compleat Lexical Tutor provides a detailed output broken down into various vocabulary families, extending to the 25,000-word family. This broader range of word families depicts a comprehensive picture of the text corpus analysed, showing both high and low-frequency families in the text. This nature of output contrasts VP-Compleat Lexical Tutor with

other tools, which provide limited word families that obscure discussion. Second, VP-Compleat Lexical Tutor integrates BNC and COCA, which are well-known and researched word lists, making it possible for the study to be replicated and compared with other studies. Third, VP-Compleat Lexical Tutor has been researched and validated in numerous studies, including Cobb (2013).

However, it is essential to acknowledge the limitation that the VP-Compleat Lexical Tutor relies on pre-existing frequency lists, namely the BNC and COCA. This may fail to capture some words that are not common in the BNC and COCA; particularly specialised words, newly emerging words, and words specific to some English dialects. To address this limitation, the present study considers this as a methodological limitation of the present study.

Results

The results section presents two tables to showcase the output of the VP-Compleat Lexical Tutor analysis. Table 3 on the one hand depicts descriptive statistics for the vocabulary coverage of the sampled textbooks. Table 4 on the other hand, illustrates cumulative descriptive statistics portraying vocabulary thresholds in the sampled textbooks. Collectively, the two tables provide a basis for comparing both vocabulary coverage and thresholds of the two textbooks in the sample.

Table 3

Vocabulary coverage of the two textbooks

Vocabulary Family	Token Frequency (%)	
	Class Seven English Textbook	**Form One Biology Textbook**
1,000	84.6	66.3
2,000	8.8	12.0
3,000	2.8	8.7
4,000	1.1	4.2
5,000	0.8	2.2
6,000	0.4	1.2
7,000	0.2	0.6
8,000	0.3	0.6
9,000	0.1	0.6
10,000	0.0	0.3
11,000	0.1	0.4
12,000	0.0	0.6
13,000	0.0	0.4
14-25,000	0.6	0.4

Off-list	0.2	1.5

Table 3 shows that the first vocabulary family has the highest coverage in the two textbooks. The Table shows that 84.6% and 66.3% of the running words in the English and Biology textbooks respectively come from the first 1000 vocabulary family. This suggests that, in comparison, students using the English textbook enjoy encountering more common words than those using the Biology textbook. Moreover, Table 3 shows that the Biology textbook has more words from the low-frequency vocabulary families than the English textbook. This further illustrates that the students face more challenging words in the Biology textbook than in the English textbook. However, Table 4 depicts the threshold levels of the two textbooks in the sample for further comparison.

Table 4

Vocabulary thresholds for the two textbooks

Vocabulary Family	English Textbook		Biology Textbook	
	Token *f* (%)	Cumulative *f* (%)	Token *f* (%)	Cumulative *f* (%)
1,000	84.6	84.6	66.3	66.3
2,000	8.8	93.4	12.0	78.3
3,000	2.8	(96.2)	8.7	87
4,000	1.1	97.3	4.2	91.2
5,000	0.8	(98.1)	2.2	93.4
6,000	0.4	98.5	1.2	94.6
7,000	0.2	98.7	0.6	(95.2)
8,000	0.3	99	0.6	95.8
9,000	0.1	99.1	0.6	96.4
10,000	0.0	99.1	0.3	96.7
11,000	0.1	99.2	0.4	97.1
12,000	0.0	99.2	0.6	97.7
13,000	0.0	99.2	0.4	(98.1)
14-25,000	0.6	99.8	0.4	98.5
Off-list	0.2	100	1.5	100

Table 4 is an extension of Table 3, revealing two significant issues regarding the vocabulary coverage of the sampled textbooks. First, the Table depicts the percentage of running words in the textbooks known by a student with a certain

vocabulary level. For instance, a student with knowledge of a 1,000-word family can only comprehend 84.6% and 66.3% of the running words in the English and Biology textbooks respectively. This suggests that such a student encounters 85 known and 15 unknown words in a paragraph of 100 words in the English language textbook. In contrast, the same student encounters about 66 known and 34 unknown words in a 100-word paragraph in the Biology textbook. Additionally, a student with vocabulary knowledge of 5,000-word families comprehends 98.1% and 93.4% of running words in English and Biology textbooks respectively. This implies that such a student encounters about 98 known and 2 unknown words in a paragraph of 100 words in the English textbook. On the other hand, the same student encounters about 93 known and 7 unknown words in a 100-word paragraph in the Biology textbook. These findings highlight significant differences in vocabulary coverage of the two sampled textbooks, illustrating different vocabulary demands. Based on the differences portrayed, the results suggest the vocabulary learning burden to learners in the primary-secondary transition.

The second issue in Table 4 is that it depicts the vocabulary thresholds required for a student to achieve 95% of running words for minimal comprehension and 98% for optimal comprehension. According to the data in Table 4, The vocabulary thresholds of 3,000 and 7,000 words are needed for a student to understand 95% of running words and achieve minimal comprehension of English and Biology textbooks, respectively. In contrast, to understand 98% of running words and achieve optimal text comprehension of the sampled textbooks, the vocabulary thresholds of 5,000 and 13,000 words are needed for optimal comprehension of English and Biology textbooks, respectively. In connection with the research question, the findings reveal a substantial gap in the vocabulary demands of the two sampled textbooks. The findings suggest that students transitioning from primary to secondary education encounter considerable challenges in comprehending the Biology textbook as it requires higher vocabulary demands than the vocabulary knowledge previously learned in primary school.

Discussion

The present study investigated the vocabulary demands of two textbooks used at the edges of primary-secondary transition. In particular, the study sought to answer the question "*Does the vocabulary coverage in the English for Standard Seven textbook adequately prepare learners for comprehending secondary school textbooks?* The essence of this question was to examine whether the English primary school textbook provides sufficient vocabulary for learners transitioning to secondary school, particularly regarding comprehension of the sampled Biology textbook. The results show that the Biology textbook contains a significant share of words from low-frequency word families compared to the English textbook. This implies that students involved in a primary-secondary transition get puzzled by the vocabulary they find in the Biology textbook, which is more difficult than what was previously exposed to them through the English textbook. Therefore, in line with Chujo and Utiyama (2005) and Mbwafu and Biseko (2023), the Biology textbook is more difficult for these learners to comprehend as it contains more words from low-frequency families.

In comparison with the findings reported by Nation (2006), our findings reveal that the two textbooks diverge significantly from the recommended average vocabulary coverage. While Nation suggests a coverage between 78% to 81% for the first 1,000 vocabulary family, the English textbook in the present study had 84.6%, surpassing the suggested one, and the Biology textbook had only 66.3%, which is less than the suggested coverage. These results imply that the English textbook was too simple, exposing learners to more basic words; thus, denying them the opportunity to interact with more challenging words. On the other hand, the findings suggest that the Biology textbook was too difficult as it exposed learners to more non-basic words than the suggested average vocabulary coverage. Relative to the average vocabulary coverage by Nation (2006), none of the two textbooks had well-selected vocabulary coverage.

Another critical issue that needs discussion is the vocabulary threshold of the two textbooks. Studies by Hu and Nation (2000), Laufer (1992), Laufer and Ravenhorst-Kalovski (2010), and Nation (2006) showed that a readers' knowledge of 95% of running words is required for minimal text comprehension, while knowledge of 98% is necessary to achieve optimal text comprehension. The present study's findings, as illustrated in Table 4, uncover two significant concerns regarding the vocabulary thresholds of the sampled textbooks. The first concern is the substantial discrepancy in vocabulary thresholds between the two sampled textbooks as depicted in Table 4. In particular, knowledge of the 3,000-vocabulary family is needed to understand 95% of running words in the English textbook and 7,000 words to guarantee understanding of the same percentage of running words in the Biology textbook. This implies that a student comprehending English texts with 3,000 words in primary school should abruptly uplift the vocabulary knowledge to 7,000 words upon primary-secondary transition to comprehend Biology texts. Likewise, knowledge of 5,000 words is required to guarantee understanding 98% of running words in the English textbook. In contrast, knowledge of 13,000 words is needed for the same percentage of running words in the Biology textbook. Generally, the gap between the vocabulary demands required to achieve text comprehension in the two levels of education is too big. Thus, it is questionable whether learners can uplift their vocabulary knowledge very abruptly after the primary-secondary transition to enable them to comprehend secondary school learning materials. Otherwise, this disparity in vocabulary demands is certainly the source of frustrations and poor performance among the learners; particularly, during this pivotal phase of transition.

The second concern is whether a Class Seven pupil in Tanzania possesses the 5,000 vocabulary families necessary for optimal comprehension of the Standard Seven English textbook. Additionally, we question whether a Form I secondary school student knows 7,000 and 13,000 vocabulary families required for minimal and optimal comprehension of the Biology textbook. Previous studies among ESL/EFL learners, including Nizonkiza and van Dyk (2015) in South Africa, Sungprakul (2016) in Thailand, and Tsuraya, Atmowardoyo, and Salija (2018) in Indonesia, indicate that university students typically have an average vocabulary range of 4,000 to 6,000 families by the time they join first-year university education. In Tanzania, Biseko (2023) found that first-year university students had an

average of 4,300 vocabulary families, while Mbwafu and Biseko (2021) reported an average of 4,500. Given these findings, it is certain that the vocabulary demands of the two sampled textbooks are far from reality for most Standard Seven and Form I students.

Additionally, it is worth noting that while the present findings suggest a significant gap in vocabulary demands between the two textbooks, some may claim that the higher vocabulary demands of the Biology textbook are to be expected, as the subject is characterised by technical vocabulary (Chung & Nation, 2004). Technical vocabulary is subject-specific, arising within a specialist domain, and forms part of a system of subject knowledge. In contrast, the English textbook is characterised by more general vocabulary, often high-frequency vocabulary. This distinction calls for the need for curriculum designers to strike a balance between subject content and knowledge accessibility. It is important to ensure that despite the nature of the subject, the content can still be accessed if the language used considers learners' ability.

The present findings have several implications for both textbook writing and pedagogy. Firstly, curriculum designers should ensure that school textbooks are coherently designed without exhibiting large discrepancies in vocabulary thresholds between two consecutive classes or educational levels. This can be achieved if material designers can use a language within learners' reach or slightly above the current learners' proficiency level (Krashen, 1982). This approach will maximise learners' comprehension of the materials and ensure learners' smooth vocabulary building as they climb the leader. Furthermore, the finding that the vocabulary thresholds of the two textbooks seem unattainable for most Tanzanian students underscores the message that textbook authors should strive to match the vocabulary level of their textbooks and the targeted learners' language ability. Regarding pedagogical implications, the present findings suggest that the vocabulary demands of the Biology textbook may overwhelm Form I students. This shows additional support from teachers is needed for learners to comprehend the vocabulary in this textbook. Therefore, Biology subject teachers should be trained to deliver the subject content and help learners in the language aspects, fostering a language-across-the-curriculum approach.

Conclusion

The present study compared the vocabulary demand between the Standard Seven primary English language textbook and the secondary Form One Biology textbook. This comparison was critical to understanding whether the vocabulary exposed to primary school Standard Seven learners through the English textbook provides a strong foundation for them to comprehend secondary school Form One textbooks. Using vocabulary coverage and threshold factors, the findings revealed a large discrepancy between the vocabulary thresholds of the two textbooks. This suggests that during the primary-secondary transition, learners do not have enough vocabulary to comprehend some secondary school textbooks. Based on the present findings, this is a case because of the lack of coherence in the vocabulary taught in primary and secondary school textbooks. Based on this discrepancy, it is certain that the

Standard Seven English textbook does not adequately prepare students for the vocabulary and reading demands of the Form One Biology textbook. The findings therefore have some important implications: first, textbook authors should ensure that textbooks for consecutive classes do not exhibit large discrepancies in their vocabulary threshold to facilitate smooth transitions for learners. Second, Biology teachers should be prepared to adopt a language through the curriculum approach as they have to provide additional support for students to comprehend the vocabulary of the Biology textbook. Lastly, it is important to acknowledge the methodological limitation of this study, particularly the use of VP-Compleat Lexical Tutor, which relies on pre-existing frequency lists namely BNC and COCA. Therefore, it fails to capture new vocabulary. Thus, a suggestion is made that alternative vocabulary analysis software should be used in future studies.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest concerning the research, authorship, and publication of this article.

References

Alqahtani, M. (2015). The Importance of Vocabulary in Language Learning and How to Be Taught. *International Journal of Teaching and Education, III*, 21-34.

Barrett, A. M., Kajoro, P., & Mills, M. (2014). *Strengthening secondary education in practice: Language Supportive Teaching and Textbooks in Tanzania (LSTT),* Pilot Study Report. Language Supportive Teaching and Textbooks. https://lstttanzania.files.wordpress.com/2015/04/lstt-pilot-report.pdf

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

Chujo, K., & Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. *Reading in a Foreign Language, 17*(1), 1-22.

Chung, T.M. & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*(2), 251-263.

Cobb, T. (2013). Range for texts v.3 [computer program]. Accessed 15 Sept 2015 at https://www.lextutor.ca/cgi-

bin/range/texts/index.pl

Criado, R. (2009). The Distribution of the Lexical Component in ELT Coursebooks and its Suitability for Vocabulary Acquisition from a Cognitive Perspective. A Case Study. *International Journal of English Studies*, *9*(3), 39–60. https://doi.org/10.6018/ijes.1.1.99511

DeCarrico, J. S. (2001). Vocabulary learning and teaching. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 285-299). Boston, MA: Heinle & Heinle.

Denise Bergström, Cathrine Norberg & Marie Nordlund (2023) "The Text Comes First" – Principles Guiding EFL Materials Developers' Vocabulary Content Decisions. Scandinavian Journal of Educational Research, 67(1), 154-168. DOI: 10.1080/00313831.2021.1990122.

Douglas, S. R. (2016, Friday, March 4, 2016). Vocabulary Thresholds that Support Post-Secondary Reading and Writing, Languages Success. (Paper presentation). Canada 9thAnnual Conference.

Ehsanzadeh. S. J. (2020). Assessing Threshold Level of L2 Vocabulary Depth in Reading Comprehension and Incidental Vocabulary Learning. *Language Education & Assessment, 3* (1), 1-12.

Goulden, R., Nation, P. and Read, J. (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics, 11,* 341-363. https://doi.org/10.1093/applin/11.4.341.

Hirsh, D., & Nation, P. (1992). What Vocabulary Size Is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language, 8*, 689-696.

Horst, M. (2013). Mainstreaming Second Language Vocabulary Acquisition. *The Canadian Journal of Applied Linguistics, 16,* (1), 171-188

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes, 30,* 247–257. http://doi.org/10.1016/j.esp. 2011.04.005

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *ESP in Asia, 33*, 54–65. https://doi.org/10.1016/j.esp.2013.07.001

Hu, M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Krashen, S. (1982). Principles and Practice in Second Language Acquisition. Pergamon Press.

Laufer, B. (1992). How Much Lexis Is Necessary for Reading Comprehension? In P. J. L. Arnaud, & H. Bejoing (Eds.), *Vocabulary and Applied Linguistics (pp. 129-132)*. London: Macmillan.

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics, 16,* 307-322. http://dx.doi.org/10.1093/applin/16.3.307

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, leaners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15-30.

Lewis, M. (1993). The lexical approach: the state of ELT and a way forward. Hove: Language Teaching Publications.

Lewis. M. (2008). Teaching Collections, Further Developments in the Lexical Approach. New York: Heinle.

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

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.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural Word Recognition and Oral Competence in English as a Foreign Language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. M. Torreblanca-López (Eds.), *Insights into Non-Native Vocabulary Teaching and Learning (pp.83-98).* Multilingual Matters. https://doi.org/10.21832/9781847692900-007.

Nagy, W.E. and Anderson, R.C. (1984). How Many Words Are There in Printed School English? *Reading Research Quarterly, 19*, 304-330. http://dx.doi.org/10.2307/747823.

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nation, I. S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63* (1), 59-82.

Nation, I. S. P. (2017). The BNC/COCA Level 6-word family lists (Version 1.0.0) [Data file]. Available from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists

Nation, I.S. P & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt, & M. McCarthy (Eds). *Vocabulary: Description, acquisition, pedagogy (pp. 6-19).* New York: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teachers*, *31*(7), 9-17.

Nation, I. S. P., & Hunston, S. (2013). Learning vocabulary in another language. Cambridge: Cambridge University Press.

Milton, J. (2009). Measuring second language vocabulary acquisition, Clevedon, UK: Multilingual Matters.

Ng, Y.J. Chong, S.T, Thiruchelvam, S. Chow, M.F. & Karthikeyan, J. (2020). Vocabulary Threshold for the Comprehension of Malaysian Secondary Engineering Texts as Compared to the Non-Engineering Genres. *International Journal of Innovation, Creativity and Change. 14*(1): 488-504.

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.

Norberg, C., & Nordlund, M. (2018). A corpus-based study of lexis in L2 English textbooks. *Journal of Language Teaching and Research, 9*(3), 463–473. https://doi.org/10.17507/jltr.0903.03

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. New York: Palgrave Macmillan.

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

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*, 261-274.

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

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

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.*

Sinclair, J. (2004). Trust the Text: Language, Corpus, and Discourse. London: Routledge.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.

Sungprakul, S. (2016). Measuring vocabulary size of Thai university students. *International Journal of Social Science and Humanities Research, 4,* 608-624.

Tsuraya, A.S. Atmowardoyo, A, & Salija, K. (2018). A Cross-Sectional Study on Vocabulary Size among Different Levels of the University Students. *ELT Worldwide, 5*(1), 57-68.

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