AN INVESTIGATION INTO LEXICAL CONTENTS IN ENGLISH STANDARDIZED TESTS

Dinh Thi Mai Anh, Tran Ba Tien*

Vinh University, Nghe An, Vietnam

ABSTRACT

This paper aims to critically analyze lexical contents in national English standardized tests in an EFL context. Using the Compleat Lexical Tutor corpus tool, the lexical profiles of these exams are meticulously examined, revealing the distribution of word types, tokens, and word families across various frequency bands. The findings show that the lexical demands in the standardized tests are inconsistent across the examined years, with a predominant emphasis on high-frequency words and recognition-based knowledge. The study concludes by offering recommendations to improve the test design, advocating for a more balanced approach that better reflects the complexities of vocabulary knowledge.

Keywords: Vocabulary; lexical profile; word types; test; examination.

1. Introduction

The significance of vocabulary in second and foreign language learning is widely acknowledged in the literature (e.g., Nation, 2013; Webb & Nation, 2017). Substantial empirical evidence suggests that a rich vocabulary repertoire not only correlates positively with second language (L2) skills (Milton, 2013) but also demonstrates a strong association with language achievement (Read, 2000), academic performance (Laufer *et al.*, 2004), and the reduction of language errors. Thus, assessing vocabulary knowledge also holds great importance within a language course. An adequate vocabulary assessment can yield beneficial washback effects on learners' attitudes towards vocabulary learning, thereby positively influencing their overall language proficiency (Schmitt & Schmitt, 2020). However, conventional vocabulary assessments prioritize the breadth of learners' vocabulary, focusing on the form-meaning connection while neglecting to evaluate the depth of their word knowledge (Schmitt, 1994). In EFL contexts (English as a foreign language), concerns have been raised over the appropriateness of the high lexical standards enforced by high-stakes standardized tests for secondary school

students. Therefore, this paper aims to critically evaluate high-school graduation English exam papers in an EFL setting with a focus on lexical aspects and then to offer relevant recommendations for improvements.

2. Literature review

2.1. Key issues in vocabulary assessment

Vocabulary assessment involves various aspects, such as word counts, lexical coverage needed for text comprehension, and different frameworks to conceptualize word knowledge. The determination of word count methodologies holds considerable importance, particularly in assessment (Read, 2000). Various approaches can be employed, including token-based, type-based, lemma-based, flemma-based and word family-based counting methods (Nation, 2013). Token count refers to the total number of running words in a text, irrespective of their repetition, encompassing every instance of word occurrence. Conversely, type count involves tallying individual words while disregarding repeated forms. Lemma-based counting involves grouping closely related words into one single-word form, known as a lemma. A lemma contains a headword along with its inflected forms and contractions. All units categorized in a lemma typically share the same word class (Francis & Kučera, 1982). Flemmas exhibit considerable resemblance to lemmas as they include inflected forms but differ in that they cover multiple parts of speech. An additional method of word counting involves word families, which comprise headwords and their corresponding inflectional and derivational forms. The selection of an appropriate counting unit relies on the specific purpose of counting, and it is crucial to ensure that the chosen unit aligns with that purpose (Nation, 2013; Schmitt & Schmitt, 2020). This decision becomes particularly significant when determining the required vocabulary knowledge for our learners, especially regarding assessment.

Exploring the level of lexical coverage necessary for text comprehension provides valuable information to establish vocabulary learning objectives and facilitate the selection of target vocabulary for assessment. Lexical coverage refers to a text's percentage of known words (Nation, 2006). Different researchers have proposed varying thresholds for adequate reading comprehension based on lexical coverage. For instance, Laufer (1989) suggested that a lexical coverage of 95% predicts satisfactory comprehension, while others like Hu and Nation (2000) proposed a higher threshold of 98%. These numbers can be seen as points on a roughly linear scale, with comprehension improving as lexical coverage increases (Schmitt *et al.*, 2011). Thus, a lexical coverage of 95% could be considered minimal, whereas coverage above 98% would be optimal (Laufer & Ravenhorst-Kalovski, 2010). The number of word families necessary to achieve different coverage figures varies depending on the mode of comprehension. For reading comprehension, a coverage of 95% is associated with approximately 4000-5000 word families, while 98% requires knowledge of 8000-9000 word families (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006).

Regarding the frameworks for conceptualizing vocabulary knowledge, Anderson and Freebody (1981) put forth the descriptive framework of breadth/size and quality/depth of knowledge, which provides a valuable approach to conceptualizing word knowledge. Breadth/size pertains to the number of known words, while depth/quality focuses on the extent and proficiency with which those words are known (Schmitt & Schmitt, 2020). The

size-depth distinction, extensively discussed in the literature (e.g., Read 2004; Schmitt, 2014), is more complex than initially perceived. While measuring vocabulary size may appear to be a clear-cut task of tallying known words, the critical consideration lies in determining the criteria used to define those known ones. Therefore, every size test also serves as a measure of depth (Schmitt, 2014). Schmitt (2014, p. 922) outlined various approaches to define and measure depth, as follows:

1. Receptive versus productive mastery
2. Knowledge of multiple word-knowledge components
3. Knowledge of polysemous meaning senses
4. Knowledge of derivative forms
5. Knowledge of collocation
6. The ability to use lexical items fluently
7. The degree and kind of lexical organization (word association)

The existence of diverse and incomplete conceptualizations of depth mentioned above strongly implies that a single comprehensive framework encompassing all aspects is unattainable (Schmitt & Schmitt, 2020). Consequently, it might be prudent to move away from the general notion of depth and focus on more specific measures, depending on different purposes (Read, 2004; Milton, 2009).

An alternative approach to conceptualizing word knowledge is distinguishing between receptive and productive knowledge. The former refers to understanding words in reading or listening, while the latter involves using words in speaking and writing (Schmitt & Schmitt, 2020). The transition from receptive to productive knowledge is an area that requires further exploration, as research suggests that receptive knowledge typically precedes productive knowledge (e.g., González-Fernández & Schmitt, 2019).

2.2. Purposes of vocabulary assessment

Vocabulary tests, like language tests in general, cannot universally apply to all learners and contexts (Read, 2000). Therefore, an essential aspect of vocabulary assessment is to specify the intended purpose, target learners, and educational contexts in which the test is designed (Schmitt *et al.*, 2020). Vocabulary tests can have diverse purposes (Nation, 2013). They can function as diagnostic tests to identify areas of difficulty for targeted intervention, placement tests to determine appropriate class levels, short-term achievement tests to assess recently learned word groups, long-term achievement tests to evaluate course success in teaching specific words, and proficiency tests to gauge overall vocabulary knowledge (ibid). Specific test purposes require distinct design choices, highlighting the necessity of early specification to guide the test development process.

Different target lexical items are chosen based on the test's specific purposes. An achievement test, for instance, often consists of items covered in class or that the students are supposed to learn throughout the course, while in placement or diagnostic tests, drawing from a broader spectrum of words may be necessary. It is noted that the inclusion of words on a test is significantly limited compared to the vast multitude of possible words available. Thus, it becomes imperative to meticulously consider the sample rate when determining which words to include in the test, ensuring a valid and reliable representation of the lexical items. In addition, the counting methodology for words is another crucial factor that requires careful consideration. Once again, tackling this matter entails complexity and is contingent upon the specific purpose of the test.

2.3. Test formats

Read (2000) suggests three dimensions that can assist in evaluating the most suitable type of vocabulary test aligned with its specific objectives. When teachers aim to specifically assess vocabulary knowledge independently from other language skills, tests that lean towards the discrete end of the Discrete/Embedded continuum are deemed more suitable. On the other hand, if their objective is to evaluate the application of words within a particular language skill, a measurement closer to the embedded end of the continuum is required. For instance, an essay examination would be appropriate for assessing vocabulary usage in writing.

Discrete tests allow selective assessment of specific words, making it relatively easier to focus on desired vocabulary items. In contrast, embedded tests pose a more significant challenge in constraining tasks to elicit particular words. As a result, skills tests often adopt comprehensive approaches to measure vocabulary usage. Instead of concentrating on the correctness or appropriateness of single words, comprehensive tests for productive abilities frequently use rating scales with specified criteria to evaluate the overall efficacy of vocabulary production.

Whether vocabulary testing should be conducted in context is still up for dispute. Designing formats that effectively assess word knowledge within context and successfully separate it from other skills, such as guessing from context, is challenging, leading to the prevalent use of isolated or single-sentence presentations (Schmitt & Schmitt, 2020). When evaluating productive skills, it is also tricky to focus exclusively on vocabulary and overlook other language-use errors. Hence, the choice of test formats is greatly influenced by the intended objectives of the assessment.

2. Methodology

This study employs a qualitative method of investigation. The Compleat Lexical Tutor application is utilized to analyze the lexical profile of the HGE. This tool is an online application for multi-purpose vocabulary learning, research and assessment, which provides platforms for educators and researchers to teach different ranges of lexical items, to build vocabulary profiles, patterns and frequency of words, or to design vocabulary tests. The current study makes use of the Compleat Lexical Tutor corpus tool to meticulously examine the lexical profiles of national standardized High School English Graduation Examinations in an EFL educational setting from 2021 to 2022, revealing the distribution of word types, tokens, and word families across various frequency bands based on Paul Nation's BNC-COCA list (2012). It also explores the aspects of word knowledge assessed by these exams, including form-meaning recognition, idioms, and phrasal verbs, and evaluates the overall difficulty of the test items.

3. Findings and discussion

The National High School Graduation Examination (HGE) is essential in the general education system investigated. It is a high-stakes standardized examination that serves multiple purposes: (1) recognizing high school graduation, (2) assessing the quality of general education, and (3) providing data for university and college admissions. It is conducted annually and covers the high school curriculum, primarily focusing on the 12th-

8

grade syllabus. The English test spans 60 minutes and consists of 50 questions, with each correct answer earning 0.2 points. Scores are reported on a scale of 0-10. The test evaluates students' proficiency in phonetics, vocabulary, grammar, communication functions, writing skills, and reading skills.

3.1. Lexical profile of the HGE

The Compleat Lexical Tutor corpus tool is utilized to analyze the lexical profile of the HGE. This tool provides valuable insights into the distribution of types, tokens, and word families across different frequency levels, such as K1 (the first 1,000 most frequent words of English - words 1 to 1,000), K2 (the second 1,000 most frequent words of English - words 1,001 to 2,000), K3 (the third 1,000 most frequent words of English - words 2,001 to 3,000), based on Paul Nation's BNC-COCA list (2012). Additionally, it reveals the number of words required to achieve 95% and 98% coverage (Table 1&2).

Table 1: *Cumulative coverage for the 2021 English papers*

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1:	337 (70.6)	443 (74.08)	1813 (86.0)	**86**
K-2:	93 (19.5)	106 (17.73)	205 (9.7)	**95.7**
Coverage 95				
K-3:	29 (6.1)	31 (5.18)	57 (2.7)	**98.4**
Coverage 98				
K-4:	12 (2.5)	12 (2.01)	22 (1.0)	**99.4**
K-5:	4 (0.8)	4 (0.67)	7 (0.3)	**99.7**
K-6:				
K-7:	1 (0.2)	1 (0.17)	2 (0.1)	**99.8**
K-8:				
K-9:				
K-10:	1 (0.2)	1 (0.17)	1 (0.0)	
K-......:				
K-25:				
Off-List:	??	0 (0.00)	0 (0.00)	
Total (unrounded)	477+?	598 (100)	2107 (100)	≈100.00

Table 2: *Cumulative coverage for the 2022 English papers*

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1:	392 (68.2)	494 (71.28)	1894 (85.9)	**85.9**
K-2:	105 (18.3)	115 (16.59)	186 (8.4)	**94.3**
K-3:	47 (8.2)	48 (6.93)	75 (3.4)	**97.7**

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
Coverage 95				
K-4:	13 (2.3)	13 (1.88)	19 (0.9)	**98.6**
Coverage 98				
K-5:	10 (1.7)	10 (1.44)	18 (0.8)	**99.4**
K-6:	3 (0.5)	3 (0.43)	3 (0.1)	**99.5**
K-7:	3 (0.5)	3 (0.43)	3 (0.1)	**99.6**
K-8:	1 (0.2)	1 (0.14)	1 (0.0)	
K-9:	1 (0.2)	1 (0.14)	1 (0.0)	
K-………				
K-25:				
Off-List:	??	5 (0.72)	5 (0.23)	99.83
Total (unrounded)	575+?	693 (100)	2205 (100)	≈100.00

Regarding the 2021 exam paper, attaining 95% coverage demands familiarity with 2,000-word families, whereas achieving 98% coverage necessitates knowledge of 3,000-word families. Conversely, for the 2022 test, learners must be acquainted with 3,000-word families to reach 95% coverage and 4,000-word families to attain 98% coverage. There is an imbalance in the vocabulary requirements that learners must prepare for in the national exams between the two years, with a higher vocabulary demand in the 2022 test than in 2021. Compared with the recommended lexical coverage figures for reading comprehension in the literature (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006), the vocabulary knowledge required for the HGE in both years is smaller. This difference can be attributed to the fact that the vocabulary for the HGE has been adapted to align with the student's proficiency level (B1 level in the CEFR framework). In contrast, the corpus used in the mentioned research consists of authentic English texts that require much higher lexical demands.

The Ministry of Education and Training (MOET) administers the exams and has established the required vocabulary size of 2,500 words for the high school English program (MOET 2018). However, the term "words" lacks a precise definition, which makes it challenging to identify the sample rate used in the test and hinders the evaluation of whether the chosen items in the HGE accurately represent the words learned in high schools. Furthermore, this ambiguity also challenges teachers and students to establish vocabulary learning objectives that align with their assessment aims. Other stakeholders, such as universities and admission committees that rely on the test results to decide students' program admission, need clarification about their lexical knowledge. Assuming that the term "word" refers to "word families", the vocabulary size required for the HGE in 2021 seems reasonable, with 2,000- and 3,000-word families necessary to achieve 95% and 98% coverage, respectively, in comparison to the targeted size of 2,500- word families. This signifies an improvement in test design regarding lexical coverage, considering that previous test papers from 2002 to 2018 were criticized for their extensive lexical

requirements (Vu, 2019). Despite the improvement recorded in the 2021 test paper, the subsequent test in 2022 displayed a continuation of the previously observed trend of increased lexical demands, which shows a mismatch between policy and practice, casting doubts on the validity of the assessment. When examining the alignment between test items and the curriculum, it becomes apparent that there is limited constructive alignment regarding vocabulary in the HGE.

Based on the analysis, high-school students who take the exams should be aware of the increased lexical demands in the HGE. To enhance their vocabulary knowledge, students can consider utilizing word lists (Coxhead, 2000; West, 1993) or engaging in activities that provide meaningful input, such as extensive reading (Vu & Peters, 2019). High-school EFL teachers who teach the students should emphasize the significance of vocabulary acquisition to their students. Test designers should also give careful attention to the lexical composition of the exam papers, ensuring they align appropriately with the required vocabulary. In addition, the MOET needs to establish a clearly defined and standardized interpretation of "words" in the context of vocabulary size requirements. By offering explicit guidance on counting methods and ensuring representative sampling, test results can more accurately reflect the students' proficiency in vocabulary knowledge.

3.2. Aspects of word knowledge being targeted in the HGE

The HGE emphasizes lexical knowledge, with approximately 21 out of 50 paper-based multiple-choice questions in the tests dedicated to assessing vocabulary. While the distribution of vocabulary-related questions in the test varies yearly, a standardized format remains constant, featuring four questions to test stress and pronunciation, four questions focused on synonyms and antonyms, and five questions assessing vocabulary knowledge embedded within reading texts. The remaining ones delve into additional dimensions of lexical knowledge, such as idioms, phrasal verbs, collocations, and other pertinent aspects of vocabulary comprehension. The test items comprise a blend of isolated, context-independent and integrated questions that necessitate contextual cues comprehension.

Evidently, the HGE test primarily emphasizes the form-meaning link, particularly in terms of form and meaning recognition, which are considered the two least challenging levels of vocabulary assessment. Regarding meaning recognition test items, the target words are presented within sentences where students must identify synonyms or antonyms instead of literal meanings. Similarly, in reading texts, students are prompted to identify words that closely resemble the meanings of the target words (Table 3).

Table 3: *Form-meaning link questions assessing meaning recognition*

Mark the letter A, B, C, or D on your answer sheet to indicate the word(s) CLOSEST in meaning to the underlined word(s) in each of the following questions.   **Question 22:** Many students are **<u>anxious</u>** about the coming interview.   **A.** worried   **B.** crowded   **C.** kind   **D.** noisy
Mark the letter A, B, C, or D on your answer sheet to indicate the word(s) OPPOSITE in meaning to the underlined word(s) in each of the following questions.   **Question 24:** My parents always encourage me to participate in social activities to **<u>gain</u>** more life experience.   **A.** warn   **B.** lose   **C.** get   **D.** post

Read the following passage and mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the questions from 31 to 35.
Music also helps you think of new ideas and believe more in yourself. If you are learning the guitar, for example, it can be really exciting when you are able to start **composing** your own pieces of music. One of the biggest benefits, of course, is that listening to music can reduce your stress. And composing music can make you feel the same way. Maybe your musical knowledge will even open up a great career path for you in the future!
Question 34: The word **composing** in paragraph 3 is closest in meaning to _____.
A. writing   **B.** copying   **C.** learning   **D.** sharing
(The HGE, 2021)

Concerning test items testing form recognition, the targeted words are also embedded in contexts in a single sentence or reading text. The target words are taken out from the single sentences and reading texts. Students are asked to choose one of the four options of word forms and decide which form will be the most suitable for those sentences and gaps in reading texts (Table 4).

Table 4: *Form-meaning link questions assessing form recognition*

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.
Question 14: Linh has been trying to solve this problem all week, but she still hasn't been able to _____ it.
A. shatter   **B.** break   **C.** crash   **D.** crack
(The HGE, 2021)

Read the following passage and mark the letter A, B, C, or D on your answer sheet to indicate the correct word or phrase that best fits each of the numbered blanks from 26 to 30.
Look at Tara Taylor's case, a mother **(28)** lives in the USA: when she uploaded a photo of her daughter to Facebook, a facefriend **(29)** _____ a problem with one of the child's eyes, so Tara took her to the doctor. It turned out that the girl had a rare disease, but her sight was saved! The story of communication is, in many **(30)** _____, the story of the human race: we have always shared knowledge and built relationships, whatever means of communication we use.
Question 29:   **A.** developed   **B.** avoided   **C.** created   **D.** spotted
Question 30:   **A.** ways   **B.** paths   **C.** roads   **D.** routes
(The HGE, 2022)

Presenting target words within sentences or reading texts might benefit us by checking students' understanding of how word forms function within different linguistic settings. However, as previously discussed, separating students' knowledge of target words from other skills, such as guessing the meaning from context, is difficult. Additionally, the provided options of word forms may limit the range of possible answers, potentially leading to guesswork or reliance on the process of elimination rather than a genuine understanding of word forms. This could impact the validity of the assessment, as students may succeed based on test-taking strategies rather than true mastery of form and meaning

recognition. Another essential aspect neglected in the test is recall knowledge, attributed to its paper-and-pencil multiple-choice format. As highlighted above, recall knowledge is inherently more arduous and demanding than recognition knowledge (González-Fernández and Schmitt 2019, Laufer and Goldstein 2004). Thus, if the tests solely concentrate on assessing form and meaning recognition, they may fail to comprehensively measure learners' vocabulary knowledge. Besides, it seems like the test items only tap into a part of receptive knowledge (i.e., understanding words in reading), and they do not test whether students can recognize the words while listening. Productive knowledge that involves using words in speaking and writing is ignored.

The test items focusing on pronunciation and stress are presented in a discrete format where students must identify words that exhibit differences in pronunciation or stress placement (Table 5).

Table 5: *Test items targeting pronunciation and stress*

Mark the letter A, B, C, or D on your answer sheet to indicate the word that differs from the other three in the position of primary stress in each of the following questions.
Question 3: **A.** nervous **B.** polite **C.** careful **D.** happy
Question 4: **A.** festival **B.** attention **C.** customer **D.** capital
Mark the letter A, B, C, or D on your answer sheet to indicate the word whose underlined part differs from the other three in pronunciation in each of the following questions.
Question 5: **A.** start<u>ed</u> **B.** rain<u>ed</u> **C.** cri<u>ed</u> **D.** seem<u>ed</u>
Question 6: **A.** f<u>a</u>ce **B.** s<u>a</u>ve **C.** t<u>a</u>ke **D.** l<u>a</u>nd
(The HGE, 2021)
Mark the letter A, B, C, or D on your answer sheet to indicate the word whose underlined part differs from the other three in pronunciation in each of the following questions.
Question 1: **A.** point<u>ed</u> **B.** cycl<u>ed</u> **C.** play<u>ed</u> **D.** occurr<u>ed</u>
Question 2: **A.** t<u>a</u>ke **B.** w<u>a</u>ve **C.** bl<u>a</u>me **D.** wr<u>a</u>p
Mark the letter A, B, C, or D on your answer sheet to indicate the word that differs from the other three in the position of stress in each of the following questions
Question 3: **A.** village **B.** student **C.** worker **D.** hotel
Question 4: **A.** relative **B.** equipment **C.** customer **D.** document
(The HGE, 2022)

Concerning Nation's (2013) framework, these test items aim to assess the knowledge of spoken forms, assuming that students need to produce the words to perceive pronunciation or stress differences and choose the correct answers. However, there is scepticism regarding the validity of this approach, as students can rely on prescribed or learned rules to select the correct answers without accurately pronouncing the words. Therefore, it is hard to ensure that with these types of questions, the knowledge of spoken form is being tested, and we are wondering whether students can recognize the words if they hear or can produce these words correctly. Thus, they could be a better choice of test items that need to change to ensure the validity of the test.

The HGE also evaluates vocabulary knowledge in terms of multiword units or formulaic language, which is highly relevant as it constitutes a significant portion of overall language usage. Both tests emphasise assessing knowledge of idioms and phrasal verbs, as evident from the presence of multiple questions in the 2021 and 2022 test papers (three and five questions, respectively) (Table 6 & 6).

Table 6: *Questions assessing knowledge of idioms*

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 21:** When the visitor saw with her own eyes the beach covered with a mountain of plastic rubbish washed up from around the world, a chill went down her _____.   **A.** spine   **B.** mind   **C.** bone   **D.** brain   **(The HGE, 2021)**
Mark the letter A, B, C, or D on your answer sheet to indicate the word(s) OPPOSITE in meaning to the underlined word(s) in each of the following questions.   **Question 7:** The arguments Prof. Parker made were controversial and some scholars were **taking issue** with him.   **A.** showing agreement   **B.** expressing doubt   **C.** becoming aware   **D.** getting annoyed   **(The HGE, 2022)**

Table 7: *Questions assessing knowledge of phrasal verbs*

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 20:** Peter loves kitchen work. He always helps his wife _____ the dishes.   **A.** wash through   **B.** wash over   **C.** wash away   **D.** wash up   **(The HGE, 2021)**
Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 16:** You should_____the form and hand it in as soon as possible.   **A.** fill in.   **B.** go on   **C.** put up   **D.** give up   ***Read the following passage and mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the questions from 36 to 42.***   When I saw her name on the screen, it took me right back to the moment she had moved away with her family ten years before. We had been such close friends at primary school. To begin with, she and I had not **hit it off**, perhaps because we were both timid. We had quickly got over that, though, and as time went on, we had become really good friends, until you almost never saw one of us without the other.   **Question 37:** The phrase **hit it off** in paragraph 1 mostly means___.   **A.** progressing slowly   **B.** disliking each other   **C.** arguing with each other   **D.** becoming friends immediately   **(The HGE, 2022)**

It is worth noting that while English has many idioms, their usage frequency is low (Moon, 1998), with only a tiny proportion being relatively common in everyday texts (Vilkaite, 2016). Similarly, research indicates that phrasal verbs constitute a mere one percent of texts in the BNC Baby Corpus, whereas collocations account for approximately

7.5 percent (Vilkaite, 2016). These findings underscore the prevalence of collocations compared to phrasal verbs and idioms. However, the number of test items focusing on collocations is relatively limited, with three in the 2021 test and just one in the 2022 test (Table 8). This raises questions regarding the assessment balance across these distinct types of formulaic language.

Table 8: *Questions assessing knowledge of collocations*

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 9:** My grandfather often _____ to the radio to get the latest news.   **A.** hears   **B.** listens   **C.** watches.   **D.** talks   **Question 11:** The next SEA Games will be held _____ Viet Nam.   **A.** in   **B.** on   **C.** at   **D.** up   ***Read the following passage and mark the letter A, B, C, or D on your answer sheet to indicate the correct word or phrase that best fits each of the numbered blanks from 26 to 30.***   ... Others, on the other hand, say that exams help students study better. They will have to study throughout the year, and if they do well, they will become more **(28)** _____ in studying…   **Question 28: A.** nervous.   **B.** bored   **C.** friendly.   **D.** interested   **(The HGE, 2021)**
Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 21:** He lives in a small house_____the city.   **A.** on.   **B.** in   **C.** at.   **D.** from   **(The HGE, 2022)**

A closer analysis of the collocation questions in both tests reveals a narrow focus on grammatical collocations and the category involving arbitrary prepositions associated with the location (Table 8). This limited prioritization fails to capture the full breadth of test-takers' collocation knowledge. Hence, test designers should include additional questions assessing different lexical collocation forms.

Some questions in the HGE may not be well-constructed (Table 9). Some items intended to evaluate lexical knowledge may inadvertently assess other aspects, such as grammatical knowledge of word form. This issue can be observed in questions 10 (The HGE, 2021) and 13 (The HGE, 2022), where the focus appears to be on assessing word form grammar rather than purely lexical knowledge.

Table 8. *Examples of some poorly designed questions*

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 10:** I like eating in this restaurant because the waiters are very _____ and friendly.   **A.** help   **B.** helpfulness   **C.** helpfully   **D.** helpful   **(The HGE 2021)**

Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.

Question 13: In order to_____soft skills, my brother has decided to attend a training course.

A. sharp   **B.** sharply   **C.** sharpen   **D.** sharpness

Question 23: The_____in the theatre gave the singer a long applause for her excellent performance.

A. watchers   **B.** spectators   **C.** audiences   **D.**viewers

Mark the letter A, B, C, or D on your answer sheet to indicate the underlined part that needs correction in each of the following questions.

Question 45: The principle of <u>comprehensive</u> inputs states that the language <u>provided</u>

                                            A   B

for students should go only a little beyond their <u>current</u> experience and <u>understanding</u>.

                                                      C   D

(The HGE, 2022)

In the HGE 2022, two errors were identified, which is highly unacceptable in a national test with significant consequences for high school students, determining their graduation and university admissions. Regarding question 23, option C, "audiences", was mistakenly designated as the correct answer by the MOET, despite the sentence mentioning a single group of audience. The accurate choice should be option B, "spectators", which refers to a collective group of individuals who attend a live performance. Shifting the focus to question 45, the targeted term, "comprehensible input", is considered unsuitable for a high school English test, as it belongs to a specialized field of second language acquisition. Additionally, the word "inputs" used in the question is incorrect; it should be in the singular form, "input", as first introduced by Krashen (1982). This question may lead students to rely on elimination strategies rather than genuinely understand this technical term.

4.3. The difficulty of the HGE

Various variables influence the difficulty of tests. Due to the limited word count, this paper specifically focuses on assessing the impact of word frequency of target words, correct answers and distractors, distractor design, and targeted word knowledge. Upon closer examination of the frequency of chosen target words in both tests, it becomes evident that most of them are found within the first 1,000 word families, and this pattern is similarly observed for distractors and correct answers. A noteworthy observation is that words with higher frequencies are more likely to be familiar to candidates, enhancing comprehension and test performance. As a result, the HGE should consider the distribution of frequency levels among target words and distractors and correct answers to increase the difficulty of tests. For illustration, an emphasis should be placed on including more words at the K-2 level in questions while ensuring they remain within the vocabulary pool of 2,500 words introduced for high school programs by the MOET.

As previously mentioned, the HGE tests predominantly evaluate form and meaning recognition, considered the least challenging vocabulary assessment, while overlooking recall knowledge, which is inherently more demanding (Fernández & Schmitt, 2019; Laufer & Goldstein, 2004). To align with the principle of a well-designed test that includes a balanced mix of simple and challenging items (Brown & Abeywickrama, 2010), it is recommended that the HGE test incorporates questions that assess recall knowledge of form and meaning, thereby increasing its level of difficulty. In the HGE, some instances of poor choices of distractors might affect the difficulty of the tests, as they are not appealing and plausible enough (Table 10).

Table 10: *Instances of poor choices of distractors*

Read the following passage and mark the letter A, B, C, or D on your answer sheet to indicate the correct word or phrase that best fits each of the numbered blanks from 26 to 30.   There are **(29)** _____ students who prefer only final exams. They say that they have to work hard for two months a year and so they have more time for their leisure activities. They think that this is a better way of **(30)** _____ student' knowledge and ability in the subjects they are studying.   **Question 30: A.** assisting   **B.** assessing   **C.** accessing   **D.** accepting   **(The HGE, 2021)**
Mark the letter A, B, C, or D on your answer sheet to indicate the correct answer to each of the following questions.   **Question 19:** Most of the houses in the neighbourhood_____terrible damage in the fire.   **A.** perceived   **B.** sustained   **C.** conceived   **D.** persevered   **(The HGE, 2022)**

For example, in question 30 of the 2021 test paper, the distractors were designed based on their similar spelling to the correct answer, or the distractors provided in question 19 are unrelated both in terms of similar forms and meanings, indicating their inadequacy as effective distractors (Table 10).

5. Conclusions

The lexical demands in the national exams could be more consistent between 2021 and 2022, with a higher demand for vocabulary in the latter. The tests dominantly target the recognition of form and meaning links and the knowledge of idioms and phrasal verbs, raising concerns about their ability to evaluate students' vocabulary knowledge comprehensively. Moreover, when considering high-frequency words, a predominant focus on recognition knowledge, and insufficiently designed distractors and questions, doubts arise regarding the test's capacity to accurately assess students across different proficiency levels, particularly those with advanced proficiency. Therefore, test designers must revise the test and consider the recommended changes outlined in this paper to ensure a more accurate and fair assessment for students.

REFERENCES

Adolphs, S., & Schmitt, N. (2003). Lexical coverage and spoken discourse. *Applied Linguistics, 24*(4), 425-438. DOI: 10.1093/applin/24.4.425

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279. DOI: 10.1093/ijl/6.4.253

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Longman.

González-Fernández, B., & Schmitt, N. (2019). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics, 41*(4), 481-505. DOI: 10.1093/applin/amy057

Laufer, B., & 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.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *EuroSLA Monographs, 2*, 57-78.

Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Clarendon Press. DOI: 10.1093/oso/9780198236146.001.0001

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139858656

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161. DOI: 10.1017/S0267190504000078

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108569057

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479. DOI: 10.1093/applin/ams074

Vu, D. V. (2019). A corpus-based lexical analysis of Vietnam's high-stakes English exams. Paper presented at *The 20th English in Southeast Asia Conference*, National Institute of Education, Nanyang Technological University, Singapore, December 20, 2019.

TÓM TẮT

NGHIÊN CỨU NỘI DUNG KIỂM TRA TỪ VỰNG
TRONG CÁC BÀI THI TIẾNG ANH TIÊU CHUẨN HÓA QUỐC GIA

Đinh Thị Mai Anh, Trần Bá Tiến
Trường Đại học Vinh, Nghệ An, Việt Nam

Bài báo này phân tích nội dung từ vựng trong các bài thi tiếng Anh tiêu chuẩn hóa quốc gia trong bối cảnh tiếng Anh như một ngoại ngữ (EFL). Sử dụng công cụ phân tích khối liệu Compleat Lexical Tutor, hồ sơ từ vựng của các bài thi được phân tích từ nhiều bình diện nhằm thống kê và luận giải loại từ vựng, dạng thức biểu hiện và họ từ vựng xuất hiện ở các tần suất khác nhau. Kết quả nghiên cứu cho thấy các bài kiểm tra từ vựng được thiết kế không nhất quán ở các năm được khảo sát, chủ yếu tập trung vào nhóm từ có tần suất cao và dựa vào kiến thức nhận biết. Kết quả nghiên cứu đã đưa ra một số khuyến nghị nhằm cải tiến việc ra đề thi tiếng Anh tiêu chuẩn hóa cấp quốc gia, hướng tới một cách tiếp cận cân bằng, phản ánh tốt hơn sự phức tạp của kiến thức từ vựng.

Từ khóa: Từ vựng; hồ sơ từ vựng; loại từ; kiểm tra; bài thi.