

LEXICAL DIVERSITY AND CEFR VOCABULARY IN CRITICAL ACADEMIC READING PASSAGES: A CONCEPTUAL REPLICATION

¹Anealka Aziz, ²Tuan Sarifah Aini Syed Ahmad, ^{*3}Suryani Awang, ⁴Roslina Abdul Aziz,
¹Nurul Afifah Azlan & ¹Siti Nurshafezan Ahmad

¹ Akademi Pengajian Bahasa, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

² Akademi Pengajian Bahasa, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan, Malaysia.

³ Akademi Pengajian Bahasa, Universiti Teknologi MARA, Cawangan Kelantan, Kampus Machang, 18500 Machang, Kelantan, Malaysia.

⁴ Akademi Pengajian Bahasa, Universiti Teknologi MARA, Cawangan Pahang, Kampus Jengka, 26400 Jengka, Pahang, Malaysia.

*Corresponding author: suryani337@uitm.edu.my

Received: 05.03.2026

Accepted: 16.06.2026

ABSTRACT

Background and Purpose: Critical academic reading skills demand learners to engage in higher-order reading skills which require advanced lexical knowledge. The two dimensions of lexical knowledge, namely lexical diversity and vocabulary variety, however, are mostly assessed in isolation using unstable measurement instruments, hence giving misleading results. The purpose of this study was to examine these two dimensions in critical academic reading passages through an integrated approach using a more reliable measurement instrument.

Methodology: The corpus consisted of 32 selected reading passages. They were first cleaned to remove any interference, such as numbering that might affect the lexical calculation, before being converted to plain text format. The Text Inspector was then used to generate the Measure of Textual Lexical Diversity (MTLD) and CEFR vocabulary difficulty levels for all passages. The scores were then analysed descriptively to examine the patterns of lexical diversity and CEFR vocabulary distributions across the passages. Additionally, Spearman's rho correlational analysis was conducted to examine the relationship between MTLD scores and CEFR-based vocabulary levels.

Findings: All 32 examined passages generally exhibit high lexical diversity, with many common and basic words. Lexical diversity was found to increase when advanced vocabulary was added to the passage, indicating a significant positive correlation between the two dimensions.

Contributions: The integrated approach in lexical analysis adopted in this study results in a comprehensive picture of lexical complexity rather than using the single metric alone, hence offering a clearer and fairer approach for instructors in evaluating and selecting reading passages to be used with students.

Keywords: CEFR profiling, critical academic reading, lexical difficulty, lexical diversity, Measure of Textual Lexical Diversity (MTLD).

Cite as: Aziz, A., Syed Ahmad, T. S. A., Awang, S., Abdul Aziz, R., Azlan, N. A., & Ahmad, S. N. (2026). Lexical diversity and CEFR vocabulary in critical academic reading passages: A conceptual replication. *Journal of Nusantara Studies*, 11(2), 228-242. <https://dx.doi.org/10.24200/jonus.vol11iss2pp228-242>

1.0 INTRODUCTION

Critical academic reading is a fundamental aspect of higher education; it demands engagement beyond surface-level understanding. Learners are required to engage in higher-order reading skills in interpreting arguments, evaluating evidence, synthesising ideas, and reflecting critically on academic material (Liu et al., 2023). These processes necessitate advanced lexical knowledge, as learners are required to navigate both the breadth and difficulty of lexical items within a text (García-Ostbye & Martínez-Sáez, 2023; Hasan & Shabdin, 2016; Hijazi et al., 2024; Ng et al., 2023). While advanced lexical knowledge has been acknowledged as of paramount importance for comprehending and evaluating complex ideas when dealing with academic reading materials, Yasin and Ahamad Shah (2019) particularly pointed out that critical academic reading skills and lexical knowledge are essential competencies for all language learners.

Despite the crucial role of lexical knowledge in critical academic reading, existing assessment practices exhibit a few notable shortcomings. First, reading passage selection and levelling depend heavily on text length, topic familiarity, and expert judgement instead of systematic evidence-based lexical analysis. This gap leads to inconsistencies in passage selection, causing an increased risk of misalignment between the expected proficiency level and the actual lexical demand. The misalignment may affect the construct validity and fairness of critical academic reading assessments. Length-robust lexical measures with CEFR-based vocabulary profiles are required to address this issue.

Second, most vocabulary assessments are still reliant on simplistic indices of text difficulty, specifically the type–token ratio (TTR) (Zhang & Lu, 2025). Although widely used, TTR is highly sensitive to text length and has been shown to produce unstable and misleading estimates of lexical diversity; hence, it does not sufficiently capture the lexical complexity of texts targeted for higher-order reading skills (McCarthy & Jarvis, 2010). As asserted by McCarthy and Jarvis (2010), reliance on TTR may obscure the true lexical demands of reading passages and undermine the validity of reading assessments.

To address what is lacking in TTR, advances in lexical analysis have introduced more robust measures of lexical diversity, such as the Measure of Textual Lexical Diversity (MTLD), which is largely unaffected by text length (McCarthy & Jarvis, 2010; Solnyshkina et al., 2023). At the same time, Common European Framework of Reference (CEFR)-based vocabulary profiling provides a widely recognised framework for describing lexical difficulty in terms of learner proficiency levels (Bannò et al., 2025; Liu et al., 2023; Qiao et al., 2021). Despite the availability of these advanced lexical diversity measures such as MTLD, they are rarely applied to critical academic reading passages (Solnyshkina et al., 2023). In addition, even though CEFR-based vocabulary profiling is commonly used to index learner proficiency, it has not been integrated with lexical density analysis to systematically explore how lexical variety and difficulty shape reading passage complexity (Jiménez, 2023; Gizatulina et al., 2020; Owen et al., 2021; Westbrook, 2022).

Another shortcoming of lexical analysis in existing research is that much of the lexical diversity literature has focused on learner writing or large general corpora, rather than on critical academic reading passages used for assessment purposes. As a result, there remains limited empirical evidence on how lexical diversity and CEFR-based vocabulary difficulty jointly shape the lexical demands of texts designed to assess critical academic reading skills.

The present study addresses these gaps by conducting a conceptual replication of established lexical diversity research that applies MTLD together with CEFR-based vocabulary profiling to critical academic reading passages. Text Inspector was used to examine how vocabulary variety and vocabulary difficulty are distributed across selected passages. It was also used to assess how these dimensions relate to one another. By integrating length-robust lexical diversity measures with CEFR-aligned vocabulary profiles, the study aims to provide a

clearer and more interpretable framework for evaluating reading passages and to support fairer assessment design in critical academic reading contexts. The study aims to fulfil the following objectives:

1. Investigate lexical diversity patterns across selected critical academic reading passages using MTLT.
2. Examine lexical difficulty across selected critical academic reading passages using CEFR-based vocabulary profiling.
3. Determine the relationship between lexical diversity indices and CEFR-based lexical distributions within the critical academic reading passages.

Upon achieving the above research objectives, the present study contributes significantly to critical academic reading instruction since it addresses the common issue in reading passage selection, which is largely based on text length and general judgement. The integration of MTLT and CEFR vocabulary levels enables a more accurate and objective method of mapping reading passages to learners' proficiency level (McCarthy & Jarvis, 2010; Westbrook, 2022). The method facilitates instructors in distinguishing between vocabulary-rich passages that are accessible and those that are lexically demanding but repetitive, hence improving both fairness and construct validity of critical reading assessments. Since the findings foster critical text engagement, instructors are able to accurately determine whether students' reading difficulties come from complex vocabulary or from a limited lexical repertoire. As a result, instruction becomes more focused and better aligned with instructional goals.

The integration of MTLT and CEFR also provides a more robust methodological approach to lexical analysis (McCarthy & Jarvis, 2010; Westbrook, 2022). This approach makes research easier to interpret. It also ensures that findings can be replicated, which is particularly valuable in conceptual replication research. The study generally offers a practical analytical framework that is theoretically grounded. Thus, it strengthens the overall quality of research and assessment design, as well as providing meaningful support for instructional decision-making in the area of critical academic reading.

2.0 LITERATURE REVIEW

2.1 Critical Academic Reading and Its Characteristics

Critical academic reading (CAR) requires a deep understanding of textual features. García-Ostbye and Martínez-Sáez (2023) state that one of the key aspects of textual features is lexical richness, which includes both lexical diversity and lexical density. This distinction is crucial because lexical diversity reflects the range of vocabulary used, while lexical density indicates the proportion of content words relative to function words, both of which directly influence the cognitive load placed on readers.

Understanding textual features is essential to help students engage effectively with academic discourse, handle complex information, and ultimately produce their own academic work (Liu et al., 2023). Hence, higher education offers CAR courses to develop the students' CAR skills. These courses not only emphasise textual analysis but also foster metacognitive strategies to enable students to monitor their comprehension and critically evaluate sources. The mastery of the skills is assessed through standardised assessments using passages which evaluate students' ability to comprehend and analyse academic texts.

The main challenge in designing the assessments is linguistic accessibility. Passages should be strictly matched with students' vocabulary thresholds and CEFR classifications to ensure equitable evaluation (Zhang & Lu, 2025). Original passages used for assessments commonly undergo rigorous pedagogical alignment, which reduces their genuine complexity.

Over-simplified passages for lower proficiency levels lose basic lexical richness (García-Ostbye & Martínez-Sáez, 2023) that characterises authentic academic texts. Therefore, it is essential to evaluate the suitability of passages for academic assessments. It requires a meticulous examination to ensure that the texts align with students' vocabulary knowledge and their CEFR levels (Zhang & Lu, 2025). The alignment is critical to maintain assessment validity and ensures that the assessment accurately reflects the intended constructs of reading comprehension and academic proficiency (Liu et al., 2023).

2.2 Vocabulary Knowledge and Reading Comprehension

Vocabulary knowledge is multifaceted and multidimensional, described as collective information about words stored in the mental lexicon and the cognitive processes required to access it (Dujardin et al., 2021; Teng, 2025). It is commonly conceptualised through two dimensions, namely breadth or vocabulary size (the total number of words a learner knows) and depth (how well a word is understood, including its multiple meanings, collocations, and associative networks) (Allagui & Naqbi, 2024; Schmitt, 2014; Teng, 2025). Educational models tend to focus on the easily quantifiable measure of vocabulary size, but true academic involvement is rooted in lexical depth. This is because the meanings of words change from one discipline to another. Therefore, instruments to assess student readiness must capture these complex associative networks, rather than surface-level word recognition.

Vocabulary knowledge has a strong association with reading comprehension. Hijazi et al. (2024) state that good vocabulary is key to decoding, comprehending, and interpreting complex academic texts. Ng et al. (2023) emphasise that having sufficient vocabulary can significantly improve a learner's ability to achieve high levels of comprehension, especially for texts aligned with CEFR standards. Hasan and Shabdin (2016) argue that insufficient vocabulary knowledge impedes comprehension, whereas robust vocabulary enhances understanding and engagement with challenging academic content. It is clear that vocabulary knowledge is a fundamental determinant of reading comprehension. Even small gaps in vocabulary can quickly grow into bigger problems that hinder overall understanding. Therefore, the belief that general reading strategies alone can make up for the absence of specialised vocabulary is mistaken, especially when dealing with complex academic texts.

The vocabulary size of second language learners is often below the level needed to understand unadapted academic texts (Nation, 2006). Vocabulary barriers are often the causes of reading failures in dense texts, not a lack of critical thinking skills (Teng, 2025). This poses a challenge to adjusting the vocabulary difficulty of assessment passages, as overly difficult passages measure basic vocabulary, not critical reading. Thus, validity should be ensured through the match between the text accessibility and CEFR levels. Precise lexical adjustment is not a dilution of academic rigour; it is a methodological necessity to isolate particular reading skills, making the test a sharp diagnostic instrument. Therefore, it is important to examine the vocabulary difficulty of the texts used for testing CAR. This helps to ascertain their accessibility and suitability for specific proficiency levels according to the CEFR.

2.3 Methods to Measure Lexical Difficulties

Quantitative and qualitative methods are used to assess the linguistic complexity and readability of academic texts for informing their suitability for critical academic reading assessments (Liu et al., 2023). The methods often involve analysing lexical sophistication, syntactic complexity, and discourse coherence to determine how challenging a text might be for a particular group of learners (Aylin & Bekir, 2025). One prominent approach is the use of readability formulas such as Flesch-Kincaid, Gunning Fog, and Dale-Chall. The purpose of using the formulas is to quantify text difficulty based on factors such as sentence length and

word frequency (Zhang & Lu, 2025). However, they often lack the specificity to capture the intricate lexical and syntactic features which are crucial for evaluating academic reading passages, particularly lexical diversity and concerning CEFR-aligned vocabulary.

Considering the above, more sophisticated lexical analyses are required. One of them is the Measure of Textual Lexical Diversity (MLTD). Chen and Liu (2020) state that MLTD is widely used to explore how vocabulary affects reading comprehension. They highlight the importance of going beyond simple word counts in order to capture the richness and complexity of academic texts. According to Solnyshkina et al. (2023), the MLTD is a reliable computational tool to measure how varied the vocabulary is within a text. The MLTD is different from the traditional approaches which depend heavily on text length or word frequency, as it provides a more precise indicator of lexical diversity (Solnyshkina et al., 2023).

As in the current study, Text Inspector was used as a tool for analysing the CEFR levels of words in the chosen texts for the CAR assessments. The CEFR categorises vocabulary into six language proficiency levels, namely A1, A2, B1, B2, C1, and C2, with C1 and C2 corresponding to increasing linguistic complexity and sophistication, which represent the advanced levels expected in academic contexts (Parashchuk et al., 2021).

Being a powerful online platform, Text Inspector can be used to conduct lexical analysis. The vocabulary level of a text is evaluated by mapping individual items to CEFR proficiency levels (Gizatulina et al., 2020; Westbrook, 2022). The analysis relies on the English Vocabulary Profile built from the Cambridge Learner Corpus which shows how learners at different stages typically use words and expressions (Bannò et al., 2025; Jiménez, 2023). This approach differs from traditional readability formulas, which often depend on simpler measures like word length or frequency, as it provides a much deeper insight into the demands of academic vocabulary (Escobar-Acevedo et al., 2022).

2.4 Previous Studies Using The Measure of Textual Lexical Diversity (MTLD) to Determine Text Difficulties

Text difficulty is often measured by using MTLD. The MLTD stands out as a reliable tool because it is different from traditional measures such as the Type-Token Ratio. Apparently, it is less affected by text length. The following are studies that have applied the MTLD to examine linguistic complexity and text difficulty.

Zhang and Lu (2024) examined how different aspects of linguistic complexity influence second language (L2) learners' perceptions of text difficulty. They found that lexical diversity in terms of the range of vocabulary used in texts is crucial in distinguishing texts across difficulty levels. The findings demonstrated that learners consistently perceived texts with higher word variation as more challenging. The study also showed that diverse texts were a key driver of perceived difficulty for non-native readers. It is due to processing a wider range of unique words demands more cognitive effort.

Su et al. (2023) studied the lexical complexity of exemplar texts used for grades 1 through 12 in the China's English curriculum. The study investigated whether the texts were appropriate for each grade level. The texts were analysed by using a range of features related to lexical diversity and sophistication. The study was able to establish quantified standards for benchmarking the difficulty of teaching materials. Clear differences in lexical complexity are demonstrated in the findings across grade levels. Thus, it indicates the importance of using diversity and sophistication metrics when source texts are adapted to match the proficiency levels of the target students.

Churunina et al. (2023) investigated whether lexical diversity indices are the predictors for the complexity of Russian language textbooks for elementary school students. The research analysed a large corpus of nearly 440,000 words drawn from 17 textbooks by comparing diversity indices with traditional readability measures. However, the findings did not show a

consistent increase in lexical diversity across grade levels. In addition, the readability score of text and its lexical diversity did not demonstrate any statistically significant correlation. Thus, the findings suggest that the difficulty of the materials is influenced by factors beyond simple vocabulary variation.

Kurdi (2020) evaluated the text complexity in intelligent tutoring systems designed for English language learners by using the Measure of Textual Lexical Diversity (MTLD). In the study, the MTLD was found to be significantly less sensitive to text length than other metrics when the text consisted of at least 100 tokens. The study shows that the MTLD is a more stable tool for assessing non-native speaker writing. Furthermore, the ANOVA results demonstrate that the MTLD provides strong effect sizes when categorising texts by linguistic complexity. Hence, it is evident that the MTLD is particularly useful for improving the accuracy of automated tutoring and assessment systems.

A decade before Kurdi's study, McCarthy and Jarvis (2010) conducted foundational validation research across multiple corpora. The research aimed to evaluate the effectiveness of the MTLD in measuring lexical variation by comparing MTLD with other lexical diversity measures such as vocd-D and HD-D. The researchers found that MTLD was the only index that did not vary with text length, making it a highly reliable metric for assessing lexical diversity in academic research. Their significant findings suggested that MTLD, along with HD-D and Maas, captures unique lexical information that can effectively differentiate the complexity levels of various text samples.

The above studies indicate that MTLD has long been used worldwide to measure lexical variations mainly due to its ability to uniquely capture lexical information and hence, the complexity of reading passages.

2.5 Previous Studies Using The CEFR Vocabulary Levels to Determine Text Difficulties

Text Inspector is a prominent online lexical analysis tool that evaluates text difficulty by mapping vocabulary to CEFR proficiency levels using the English Vocabulary Profile. The tool identifies the percentage of words at each level (A1–C2) to help educators and researchers determine if a text is appropriate for a specific target audience. The following studies have used Text Inspector to determine text difficulty based on CEFR vocabulary levels.

Bannò et al. (2025) explored the use of the English Vocabulary Profile through Text Inspector to assess L2 word-level vocabulary proficiency, specifically noting the tool's ability to identify proficiency levels associated with different word meanings. The study found that Text Inspector's rule-based implementation of the CEFR framework is a valuable starting point for assessing writing proficiency, as it presents users with potential proficiency levels for ambiguous or polysemous words. Their findings emphasised that human oversight is still required to resolve linguistic ambiguities, although the tool is highly effective for mapping vocabulary to the CEFR.

Jiménez (2023) used Text Inspector to evaluate the vocabulary complexity of academic texts produced by EFL university students. The tool was used to compare their word choices against the English Vocabulary Profile. The study highlights how the tool triggers CEFR results by contrasting learner data with the Cambridge Learner Corpus, allowing a precise measurement of lexical sophistication. By comparing with using traditional diversity indices alone, the findings reveal that a more comprehensive perspective on a writer's lexical development can be provided by Text Inspector. The tool functions to track CEFR levels alongside academic vocabulary coverage.

Westbrook (2022) analysed the lexical richness and CEFR profiles of spoken and visual input in a listening-into-writing assessment by using Text Inspector. The findings reveal that the spoken input remained fairly consistent in the level of vocabulary throughout the lecture. However, the visual materials presented on the slides varied greatly in their CEFR distribution.

The results suggest that the visual components of an assessment can unintentionally introduce shifting levels of lexical demand which are not present in the auditory input.

Owen et al. (2021) conducted a large-scale study by using Text Inspector for analysing the lexical thresholds and profiles of more than 6,000 writing responses from the British Council's Aptis Writing test which were benchmarked against the CEFR. The purpose of the study was to identify clear lexical markers that distinguish performance across different CEFR levels. Strong empirical evidence is provided by the findings that specific vocabulary thresholds are needed to achieve higher proficiency ratings. The study also highlights that Text Inspector is a reliable tool for validating large-scale standardised assessments.

Gizatulina et al. (2020) examined fluctuations in text complexity within the Basic State Examination in English by using Text Inspector. The study focused on how the tool tags words with their corresponding CEFR levels. The results indicated that Text Inspector provided reliable metrics for sentence and word length that aligned with other advanced analysers such as Coh-Metrix. The significant findings indicated that the tool is effective for benchmarking the linguistic difficulty of standardised exam materials and that vocabulary is ensured to remain within the appropriate proficiency range for test-takers.

The review of the studies demonstrates that Text Inspector offers significant advantages for evaluating vocabulary variety and text complexity and for aligning them with CEFR-based difficulty levels. It is evident that Text Inspector is proven to be reliable in precisely evaluating and enhancing language learning and assessment since it is used for various aspects of analysis,

3.0 RESEARCH DESIGN

This study employed a quantitative corpus-based textual analysis with a conceptual replication design. The purpose of this study was to examine lexical diversity and lexical difficulty in critical academic reading passages. A conceptual replication approach was adopted to extend the use of MTLTD from the previous lexical profiling study, together with CEFR-based vocabulary profiling in a new corpus of critical academic reading passages. Lexical diversity of the passages was measured by the Measure of Textual Lexical Diversity (MTLD), while lexical difficulty was measured by the Common European Framework of Reference for Languages (CEFR) proficiency levels, ranging from A1 (beginner) to C2 (mastery).

The corpus consisted of 32 selected reading passages used in a higher education learning environment. Selected passages for higher-education reading assessments were purposively sampled based on academic tone, relevance to critical academic reading, and appropriate text length. Only reading passages were used, and other parts of the assessment sections were not included. Lexical analysis was conducted using Text Inspector to generate Measure of Textual Lexical Diversity (MTLD) scores and CEFR-based vocabulary profiles (A1–C2) for each passage. All texts were converted into plain text format and analysed individually using consistent tool settings. Prior to analysis, the passages were cleaned up to remove any possible interference such as instructions, numbering, and formatting symbols in the lexical calculation.

The main research instrument was Text Inspector, which was used to generate MTLD scores and CEFR vocabulary levels for all passages. All MTLD scores and CEFR vocabulary levels were analysed descriptively (minimum, maximum and standard deviation) to examine patterns of lexical diversity and CEFR vocabulary distributions across the passages. These analyses addressed Research Questions 1 and 2. For Research Question 3, Spearman's rho correlational analysis was conducted to examine the relationship between MTLD scores and CEFR-based vocabulary levels. This analysis helped to determine whether the passages with greater lexical diversity tended to have higher proportions of advanced-level vocabulary.

The reliability of the analysis was ensured by using well-researched automated software (Text Inspector) and MTLD, a common measure of lexical diversity that works well with texts

of different lengths. All passages went through the same preparation procedure and analysis to ensure consistency of results. As for the validity, the use of authentic academic reading passages and established measures of lexical diversity and difficulty helped ensure that the findings reflected the lexical characteristics of the selected passages.

4.0 ANALYSIS AND DISCUSSION

This section reports the results of the current study according to its research objectives.

Objective 1: Investigate lexical diversity patterns across selected critical academic reading passages using MTLD.

The first research question aims to identify the extent of vocabulary variation in the reading passages measured by MTLD. This analysis helps to determine whether the passages use a wide range of vocabulary frequently, which reflects the complexity of the passages.

Table 1: Descriptive statistics of the Measure of Textual Lexical Diversity (MTLD)

	N	Minimum	Maximum	Mean	Std. Deviation
MTLD	32	80	169	106.83	23.173

Table 1 presents descriptive statistics for lexical diversity for 32 critical academic reading passages. The mean MTLD score is 106.83, which indicates the passages generally have high lexical diversity. In other words, the passages generally contain a diverse range of words. The MTLD values range from 80 to 169, which indicates that some passages use moderately varied vocabulary, while others show higher lexical diversity. The dispersion around the mean demonstrates that the passages generally have a high lexical diversity. The standard deviation of 23.17 shows a moderate spread of lexical diversity across the passages. It shows that the lexical diversity is not uniformly distributed across passages. In other words, although most texts have good vocabulary variation, the level of diversity is not the same across all passages. Such differences, possibly caused by different thematic focus, rhetorical purpose and disciplinary conventions, suggest learners are exposed to different degrees of lexical challenge that will influence their cognitive processing and reading engagement.

Despite this, the MTLD score of 106.83 indicates lexical richness in the passages, hence providing a challenge for learners in dealing with critical academic reading tasks. While offering learners complex academic reading discourse, lexical richness in reading texts also helps learners' vocabulary growth and fosters their academic literacy development (Liu et al., 2023).

Overall, the findings indicate that the critical academic reading passages provide a linguistically rich input environment, characterised by generally high lexical diversity alongside meaningful variation in vocabulary distribution across texts. As lexical diversity correlates with increased cognitive processing demand, high vocabulary variety would result in learners applying deeper inferencing and decoding strategies when performing critical academic reading tasks (Zhang & Lu, 2024).

While MTLD provides an overall indicator of lexical diversity, a more detailed understanding of passage difficulty requires examining the distribution of vocabulary across different proficiency levels. This is addressed by the second research objective.

Objective 2: Investigate lexical difficulty across selected critical academic reading passages using CEFR-based vocabulary profiling.

This section presents the CEFR levels of the words used in the passages. The finding helps determine whether the passages are appropriate for learners' language abilities and whether the lexical difficulty is consistent across the passages.

Table 2: CEFR level distribution of vocabulary used in the passages

CEFR Levels	N	Minimum	Maximum	Mean	Std. Deviation
A1-Beginners	32	115	187	143.06	17.502
A2-Pre-Intermediate	32	58	111	84.84	12.721
B1-Intermediate	32	66	119	91.66	14.526
B2-Upper Intermediate	32	57	120	78.78	14.094
C1-Advanced	32	11	41	23.50	7.121
C2-Proficiency	32	6	30	14.13	5.735

Table 2 presents descriptive statistics of CEFR-based vocabulary distribution across the 32 critical academic reading passages. The results show that the passages primarily use common and basic words. This is evident from the highest mean vocabulary frequency at the A1 level ($M = 143.06$, $SD = 17.50$), indicating that the passages are consistently grounded in high-frequency words. The standard deviation is relatively high, indicating that although basic words are used frequently in the passages, the A1 vocabulary varies moderately across passages.

A considerable number of words are at the A2 level ($M = 84.84$, $SD = 12.72$), B1 level ($M = 91.66$, $SD = 14.53$), and B2 level ($M = 78.78$, $SD = 14.09$). It shows that the passages consistently include vocabulary from beginner to upper-intermediate levels. The occurrence of lower to relatively higher-level vocabulary is in line with Liu et al.'s (2023) argument that reading passages that are linguistically accessible yet challenging for learners would ensure construct validity of the assessment. Aligned with CEFR recommendations, accessibility is offered to learners when the reading passages contain high-frequency words, evident by the presence of A1-B2 level vocabulary (Zhang & Lu, 2025).

In contrast, words at the advanced and proficiency levels are used less frequently in the passages. The mean scores for C1 and C2 are low, with $M = 23.50$, $SD = 7.12$ and $M = 14.13$, $SD = 5.74$, respectively. These results show that advanced words are used only when necessary to explain academic ideas clearly, without making the passage too hard to understand. Past studies have shown that advanced-level vocabulary should be used sparingly to prevent learners' cognitive overload. Ng et al. (2023), for instance, contended that substantial exposure to lower-level vocabulary is required to ensure the learners can comprehend C1 – C2 items with ease. The moderate standard deviations across CEFR levels further suggest relative consistency in vocabulary distribution among passages, despite some variation in lexical difficulty.

Overall, the results show that the critical academic reading passages use a graded vocabulary approach, with many common words forming the main foundation and a careful use of more advanced words when needed. As highlighted by Dujardin et al. (2021), the inclusion of graded lexical components supports the hierarchical development of learners' vocabulary as they experience increased difficulty in dealing with the texts. By providing learners with increasing lexical difficulty that follows the CEFR recommendation on vocabulary teaching, the learning process will be more engaging for learners, hence facilitating their critical academic reading development.

While the descriptive statistics establish the overall composition of the passages, they do not explain how lexical diversity relates to vocabulary difficulty. To address this relationship, Table 3 presents the Spearman correlation analysis examining the associations between vocabulary distribution across the critical academic reading passages.

Objective 3: Evaluate the relationship between lexical diversity indices and CEFR-based vocabulary distributions within the critical academic reading passages.

Table 3 presents the Spearman's rho correlation analysis examining the relationship between lexical diversity, measured by the Measure of Textual Lexical Diversity (MTLD), and CEFR-based vocabulary distributions across the critical academic reading passages.

Table 3: Correlation analysis between lexical diversity and CEFR vocabulary distribution

			MTLD	A1	A2	B1	B2	C1	C2
Spearman's rho	MTLD	Correlation Coefficient	1.000	.338	.160	.318	.357*	.263	.409*
		Sig. (2-tailed)	.	.059	.383	.076	.045	.146	.020
*. Correlation is significant at the 0.05 level (2-tailed).									
**. Correlation is significant at the 0.01 level (2-tailed).									

The results in Table 3 indicate that lexical diversity demonstrates significant positive correlations with higher proficiency vocabulary levels, specifically B2 ($r_s = .357$, $p = .045$) and C2 ($r_s = .409$, $p = .020$). The findings suggest that passages that include a higher proportion of upper-intermediate and advanced lexical items have the tendency to contain greater lexical diversity. Thus, lexical variation increases as texts incorporate more sophisticated vocabulary resources.

There are no statistically significant correlations between MTLD and low-level vocabulary at A1, A2, and B1 levels. The condition aligns with the basic concept of lexical development, in which low-level vocabulary only has a small contribution to lexical diversity. This is due to its limited semantic range despite its high frequency in a text. However, advanced vocabulary usually carries more detailed semantic, morphological, and collocational features, which result in greater lexical richness (Schmitt, 2014; Teng, 2025). This explains that the presence of basic vocabulary supports accessibility rather than contributes substantially to lexical richness.

In contrast, significant positive correlations are shown between the MTLD and higher CEFR vocabulary levels, which are B2 and C2 levels. The finding demonstrates that advanced vocabulary used in the passage increases lexical diversity. This is because the MTLD is highly sensitive to high lexical items (Solnyshkina et al., 2023; McCarthy & Jarvis, 2010). The stronger association between the MTLD and advanced vocabulary levels shows that lexical diversity reflects qualitative lexical development rather than the quantity of simple vocabulary. Obviously, texts become more lexically diverse when higher-level vocabulary alongside foundational lexical items is selectively integrated.

From another perspective, the strong relationship observed between MTLD and advanced vocabulary (C2) indicates learners' exposure to dense academic reading passages. In relation to this, Bannò et al. (2025) and Parashchuk et al. (2021) assert that vocabulary at C1 and C2 levels functions as the main indicator of language proficiency and reading difficulty in academic texts.

Overall, the findings show that lexical diversity and lexical difficulty are closely related, but they affect text complexity in different ways. The MTLD shows how varied the vocabulary is while CEFR vocabulary levels indicate how difficult the vocabulary is in terms of proficiency levels. Hence, the findings suggest that passage difficulty increases when advanced vocabulary is carefully added, rather than simply increasing the number of basic words.

Spearman's rho correlation analysis was further used to analyse the data. Then, the relationships between CEFR-based vocabulary levels and the critical academic reading passages were examined. The findings in Table 4 show clear and systematic patterns of lexical organisation that reflect graded vocabulary progression.

Table 4: Correlation analysis among CEFR vocabulary levels

			A1	A2	B1	B2	C1	C2
Spearman's rho	A1	Correlation Coefficient	1.000	.584**	-.335	-.403*	-.345	-.128
		Sig. (2-tailed)	.	.000	.061	.022	.053	.486
	A2	Correlation Coefficient	.584**	1.000	.023	-.392*	-.259	-.179
		Sig. (2-tailed)	.000	.	.901	.027	.153	.328
	B1	Correlation Coefficient	-.335	.023	1.000	.673**	.585**	.304
		Sig. (2-tailed)	.061	.901	.	.000	.000	.090
	B2	Correlation Coefficient	-.403*	-.392*	.673**	1.000	.628**	.533**
		Sig. (2-tailed)	.022	.027	.000	.	.000	.002
	C1	Correlation Coefficient	-.345	-.259	.585**	.628**	1.000	.337
		Sig. (2-tailed)	.053	.153	.000	.000	.	.059
	C2	Correlation Coefficient	-.128	-.179	.304	.533**	.337	1.000
		Sig. (2-tailed)	.486	.328	.090	.002	.059	.
*. Correlation is significant at the 0.05 level (2-tailed).								
**. Correlation is significant at the 0.01 level (2-tailed).								

Table 4 reveals that A1 and A2 vocabulary levels had a strong positive correlation ($r_s = .584$, $p < .001$). Evidently, texts with higher frequencies of basic vocabulary tend to simultaneously include lower-intermediate lexical items. The pattern suggests that textual accessibility is collectively maintained by foundational vocabulary bands.

In contrast, beginner-level vocabulary demonstrates negative correlations with higher proficiency levels, particularly between A1 and B2 ($r_s = -.403$, $p = .022$) and between A2 and B2 ($r_s = -.392$, $p = .027$). These relationships demonstrate a distributional trade-off in lexical composition. Passages heavily grounded in foundational vocabulary generally contain fewer upper-intermediate lexical items. The findings reflect intentional lexical balancing within academic texts to control overall difficulty.

Intermediate and advanced vocabulary bands also had strong positive correlations. Significant associations between B1 and B2 ($r_s = .673$, $p < .001$), B1 and C1 ($r_s = .585$, $p < .001$), and B2 and C1 ($r_s = .628$, $p < .001$) demonstrate that lexical difficulty develops cumulatively across proficiency levels. Thus, linguistic sophistication increases when passages include intermediate vocabulary where advanced lexical resources are added. The significant relationship between B2 and C2 vocabulary ($r_s = .533$, $p = .002$) further supports this hierarchical progression, which suggests that advanced lexical items tend to co-occur rather than appear in isolation.

Overall, the findings demonstrate that CEFR vocabulary levels are organised in a clear progression of difficulty rather than appearing randomly. Basic vocabulary helps make texts easy to understand. In contrast, intermediate and advanced words gradually make the passages more complex. These correlations reaffirm that despite their distinctive concepts, lexical diversity and lexical difficulty are complementary constructs. With the MTLT capturing the distributional richness of vocabulary and CEFR levels quantifying vocabulary difficulty, a combination of both promotes critical and comprehensive understanding of a reading text. Hence, this supports the recommendation of the literature that an integration of multiple indices leads to a more effective evaluation of academic reading texts for L2 learners (Liu et al., 2023; Zhang & Lu, 2024).

While the above section mainly reports the research findings, the following is the discussion of results with some implications on reading assessments, the selection of reading passages and benefits offered for instructors. The findings indicate that the critical academic reading passages analysed are characterised by high lexical diversity. This reflects the lexical richness of authentic academic discourse. Variation in MTLT scores suggests that learners are likely to encounter different levels of lexical challenge, which may influence their cognitive processing and reading engagement. Such lexical richness can support vocabulary growth and

academic literacy development (Liu et al., 2023), while also encouraging deeper inferencing and meaning-making during critical reading tasks (Zhang & Lu, 2024). Meanwhile, the CEFR profile suggests that the passages achieve a balance between accessibility and challenge by relying primarily on high-frequency A1–B2 vocabulary while incorporating advanced lexical items. Such vocabulary distribution is consistent with the notion that academic text selection must be both comprehensible and challenging (Liu et al., 2023; Zhang & Lu, 2025). The selective use of C1–C2 vocabulary also suggests a gradual progression in lexical difficulty that can promote learning without overwhelming readers (Ng et al., 2023; Dujardin et al., 2021).

These findings carry pertinent implications for the design of reading assessments and the selection of reading passages. The integration of MTLTD and CEFR vocabulary profiling provides a more objective approach to text difficulty. Thus, this moves beyond the dependence on text length and subjective evaluation. By adopting the approach, developers can make a clear distinction between texts that are lexically diverse but accessible, from those that are repetitive but lexically demanding. Consequently, critical academic reading assessments can align more effectively among lexical demands, intended proficiency levels and assessment constructs, resulting in improved fairness and construct validity. In addition, the findings also benefit instructors. It provides them with practical guidance for the selection and design of critical academic reading material. Language instructors can incorporate lexical diversity together with CEFR vocabulary profiles in academic reading passage selection to ensure they align with learners' proficiency levels, while still promoting critical engagement with academic materials. Moreover, the findings can assist instructors in effectively determining whether reading challenges stem from excessive lexical demands or restricted lexical variety. This allows them to effectively determine the targeted vocabulary interventions and cultivate students' critical academic reading skills.

The findings also indicate positive associations between MTLTD and higher-level vocabulary bands (B2 and C2). This suggests that lexical diversity is driven more by lexical sophistication than by the frequency of basic vocabulary. Although basic vocabulary contributes to accessibility, higher-level lexical items broaden semantic and collocational variation, leading to greater lexical diversity and increased demand on academic reading (Schmitt, 2014; Solnyshkina et al., 2023; Teng, 2025). The CEFR correlations further reveal a clear progression of lexical difficulty across proficiency levels, highlighting that MTLTD and CEFR profiling capture complementary dimensions of text complexity and, when combined, provide a more comprehensive evaluation of academic reading materials for L2 learners (Liu et al., 2023; Zhang & Lu, 2024).

These findings have important theoretical implications. They emphasise the central role of vocabulary in critical academic reading. They also show how text complexity is influenced by lexical diversity and CEFR-based vocabulary levels. The notion that critical reading is purely a matter of cognitive processing is challenged by the findings of this study. Instead, vocabulary variety and proficiency levels are emphasised as essential support for higher-order processes such as making inferences, evaluating arguments, and synthesising information. The study demonstrates a richer way of conceptualising text difficulty which moves beyond conventional readability formulas and aligns more closely with current indices of critical reading.

Methodologically, the study highlights the advantages of incorporating the MTLTD into a well-established proficiency framework (CEFR) for analysing critical reading text selection. Integration can improve the interpretability and replicability of lexical analysis, specifically within conceptual replication research. More importantly, the study shows that automated text analysis tools can be effectively aligned with pedagogical and assessment frameworks, reinforcing the link between corpus-based research and classroom applications.

In general, the study provides a theoretical basis and practical framework for analysing lexical appropriateness in critical academic reading texts. It supports more informed decisions in research, assessment and pedagogy, hence enhancing the quality, fairness and transparency in academic reading settings.

5.0 CONCLUSION

The study investigated the lexical diversity in critical academic texts using MLTD, lexical difficulty using CEFR-based profiling and the relationship between diversity and CEFR levels. Three main conclusions were drawn from the study.

First, the study concludes that the academic texts analysed are both accessible and lexically rich. More importantly, the results indicate that MLTD provides a consistent and effective measure of lexical diversity, capturing both the breadth of vocabulary and its distribution across passages. Its sensitivity enables it to differentiate between different levels of lexical richness.

Second, the analysed passages recorded a balanced level of lexical difficulty based on the CEFR profiling. The passages mainly contain high-frequency vocabulary (A1–B2), ensuring ease of comprehension, with controlled use of C1-C2 vocabulary that enhances learning potential. The findings suggest that CEFR profiling is a stable and effective method in analysing and controlling lexical difficulty in critical academic reading contexts.

Third, the findings suggest integrating MLTD and CEFR-based profiling provides a more comprehensive picture of lexical complexity rather than using the single metric alone. MLTD captures vocabulary richness, while the CEFR-profiler provides information on text difficulty and accessibility. Together, these measures can be effectively applied by instructors in the selection and evaluation of critical academic reading texts.

In sum, this study has clearly shown that, by integrating MTLTD and CEFR vocabulary profiling to analyse lexical diversity and difficulty levels of academic texts, the approach not only ensures fairness in student evaluation but also helps them make proper progress in language learning.

ACKNOWLEDGEMENT

The article is funded by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS) FRGS/1/2022/SSI07/UITM/02/46.

REFERENCES

- Allagui, B., & Naqbi, S. A. (2024). The contribution of vocabulary knowledge to summary writing quality: Vocabulary size and lexical richness. *Teaching English as a Second or Foreign Language*, 28(1), 1-27. <https://doi.org/10.55593/ej.28109a5>
- Aylin, Ü., & Bekir, A. (2025). *Evaluating text suitability for English for academic purposes reading assessment: Teachers versus Lexile, Coh-Metrix and ChatGPT*. British Council.
- Bannò, S., Knill, K. M., & Gales, M. (2025). Exploiting the English vocabulary profile for L2 word-level vocabulary assessment with LLMs. *Journal of Latex Class Files*, 18(9), 1-24. <https://doi.org/10.48550/arxiv.2506.02758>
- Chen, C., & Liu, Y. (2020). The role of vocabulary breadth and depth in IELTS academic reading tests. *Reading in a Foreign Language*, 32(1), 1-27. <https://doi.org/10.17863/cam.51399>
- Churunina, A. A., Solnyshkina, M., & Yarmakeev, I. (2023). Lexical diversity as a predictor of complexity in textbooks on the Russian language. *Russian Language Studies*, 21(2), 212-227. <https://doi.org/10.22363/2618-8163-2023-21-2-212-227>

- Dujardin, É., Auphan, P., Bailloud, N., Écalte, J., & Magnan, A. (2021). Tools and teaching strategies for vocabulary assessment and instruction: A review. *Social Education Research*, 3(1), Article 34. <https://doi.org/10.37256/ser.3120221044>
- Escobar-Acevedo, A., Guerrero-García, J., & Guzmán-Cabrera, R. (2022). A model text recommendation system for engaging English language learners: Facilitating selections on CEFR. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 14(3), 1-8. <https://doi.org/10.21659/rupkatha.v14n3.17>
- García-Ostbye, I. C., & Martínez-Sáez, A. (2023). Reading challenges in higher education: How suitable are online genres in English for medical purposes? *ESP Today*, 11(1), 53-74. <https://doi.org/10.18485/esptoday.2023.11.1.3>
- Gizatulina, D., Ismaeva, F., Solnyshkina, M. I., Мартынова, Е. В., & Yarmakeev, I. (2020). Fluctuations of text complexity: The case of basic state examination in English. *SHS Web of Conferences*, 88(1), Article 02001. <https://doi.org/10.1051/shsconf/20208802001>
- Hasan, M. K., & Shabdin, A. A. (2016). Conceptualization of depth of vocabulary knowledge with academic reading comprehension. *PASAA*, 51(1), 235-268. <https://doi.org/10.58837/CHULA.PASAA.51.1.9>
- Hijazi, H., Gomes, M., Castelhana, J., Castelo-Branco, M., Praça, I., Carvalho, P., & Madeira, H. (2024). Dynamically predicting comprehension difficulties through physiological data and intelligent wearables. *Scientific Reports*, 14(1), 1-17. <https://doi.org/10.1038/s41598-024-63654-z>
- Jiménez, W. C. (2023). Vocabulary complexity in EFL university students' academic texts. *Revista de Linguas Modernas*, 37(1), 1-22. <https://doi.org/10.15517/rlm.v0i37.50826>
- Kurdi, M. Z. (2020). Text complexity classification based on linguistic information: Application to intelligent tutoring of ESL. *Journal of Data Mining & Digital Humanities*, 2020(1), 1-40. <https://doi.org/10.46298/jdmdh.6012>
- Liu, H., Shi, X., Qiu, J., Shi, Y., Hao, Y., Zhu, L., Yan, C., & Li, H. (2023). Academic word coverage and language difficulty of reading passages in college English test and test of English for Academic Purposes in China. *Frontiers in Psychology*, 14(1), 01-11. <https://doi.org/10.3389/fpsyg.2023.1171227>
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, VOCED-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381-392. <https://doi.org/10.3758/brm.42.2.381>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Ng, Y. J., Hussin, A. A., Roslim, N., & Abidin, D. Z. (2023). Vocabulary benchmarking for the comprehension of CEFR-aligned assessment reading texts. *Journal of Language and Communication*, 10(2), 241-256. <https://doi.org/10.47836/jlc.10.02.06>
- Owen, N., Shrestha, P., & Bax, S. (2021). *Researching lexical thresholds and lexical profiles across the common European framework of reference for languages (CEFR) levels assessed in the Aptis test*. British Council.
- Parashchuk, V., Yarova, L., & Parashchuk, S. (2021). Automated complexity assessment of English informational texts for EFL pre-service teachers and translators. *Arab World English Journal*, 7(1), 155-164. <https://doi.org/10.24093/awej/call7.11>
- Qiao, Y., Zhou, W., Kerz, E., & Schlüter, R. (2021). The impact of ASR on the automatic analysis of linguistic complexity and sophistication in spontaneous L2 speech. *arXiv*, 1(1), 1-7. <https://doi.org/10.48550/arXiv.2104.08529>
- 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>

- Solnyshkina, M., Gatiyatullina, G. M., Kupriyanov, R., & Ziganshina, C. (2023). Lexical density as a complexity predictor: The case of Science and Social Studies textbooks. *Research Result: Theoretical and Applied Linguistics*, 9(1), 11-26. <https://doi.org/10.18413/2313-8912-2023-9-1-0-2>
- Su, Y., Liu, K., Liu, F., Lee, J., & Jin, T. (2023). Lexical complexity in exemplar EFL texts: Towards text adaptation for 12 grades of basic English curriculum in China. *International Review of Applied Linguistics in Language Teaching*, 62(1), 137-165. <https://doi.org/10.1515/iral-2022-0236>
- Teng, M. F. (2025). Cross-lagged panel analysis of reciprocal effects of metacognitive knowledge and breadth of vocabulary knowledge in a foreign language context. *Studies in Second Language Learning and Teaching*, 16(1), 133-164. <https://doi.org/10.14746/ssllt.44595>
- Westbrook, C. (2022). The impact of input format on written performance in a listening-into-writing assessment. *Journal of English for Academic Purposes*, 61(1), Article 101190. <https://doi.org/10.1016/j.jeap.2022.101190>
- Yasin, Z. M., & Ahamad Shah, M. I. (2019). The interactive roles of lexical knowledge and reading strategies on reading comprehension performance. *Journal of Nusantara Studies*, 4(1), 273-299. <https://doi.org/10.24200/jonus.vol4iss1pp273-299>
- Zhang, X., & Lu, X. (2024). Testing the relationship of linguistic complexity to second language learners' comparative judgment on text difficulty. *Language Learning*, 74(3), 672-706. <https://doi.org/10.1111/lang.12633>
- Zhang, X., & Lu, X. (2025). Aligning linguistic complexity with the difficulty of English texts for L2 learners based on CEFR levels. *Studies in Second Language Acquisition*, 47(5), 1407-1434. <https://doi.org/10.1017/s0272263125101125>