انموذج الانقرائية الخاص بمؤشرات التعقيد المعجمي في نصوص القراءة لمتعلمي اللغة الانكليزية بوصفها لغة أجنبية في العراق

الباحث حسين عادل يوسف
قسم اللغة الانكليزية / كلية الآداب
جامعة البصرة

الاستاذ المساعد الدكتوره بهيجه جاسم محمد
قسم اللغة الانكليزية / كلية الآداب / جامعة البصرة

الاستاذ المساعد الدكتور حسين عبد الكريم يعقوب
قسم اللغة الانكليزية / كلية الآداب /
جامعة البصرة

البريد الإلكتروني Email : hussain.adil@uobasrah.edu.iq

الكلمات المفتاحية: التعقيد المعجمي، المقروئية، كتب اللغة الانكليزية كلغة اجنبيه، الانحدار ، اللسانيات الحاسوبية.

كيفية اقتباس البحث

يوسف ، حسين عادل ، بهيجه جاسم محمد ، حسين عبد الكريم يعقوب ، انموذج الانقرائية الخاص بمؤشرات التعقيد المعجمي في نصوص القراءة لمتعلمي اللغة الانكليزية بوصفها لغة أجنبية في العراق،مجلة مركز بابل للدراسات الانسانية، أيلول 2026،المجلد:16 ،العدد:9.

مفهرسة في Indexed
IASJ

مسجلة في Registered
ROAD

A readability model for lexical complexity indicators in reading texts for learners of English as a foreign language in Iraq

Res. Hussein Adil Yousif
Department of English Language / College of Arts, University of Basra

Asst. Prof. Dr. Baheeja Jasim Mohammed
Department of English Language / College of Arts, University of Basra

Asst. Prof. Dr. Husseın Abdulkareem Yaqoob
Department of English Language / College of Arts, University of Basra

Keywords : Lexical Complexity, EFL-specific Readability, EFL Textbooks, Regression, Computational Linguistics.

How To Cite This Article
Yousif, Hussein Adil , Baheeja Jasim Mohammed , Hussein Abdulkareem Yaqoob, A readability model for lexical complexity indicators in reading texts for learners of English as a foreign language in Iraq,Journal Of Babylon Center For Humanities Studies, September 2026,Volume:16, Issue 9.

المستخلص

في الأدبيات الحالية ، طور الباحثون مجموعة واسعة من مؤشرات التعقيد المعجمي بهدف التنبؤ بمقروئية نصوص اللغة الانكليزية بوصفها لغة أجنبية. على الرغم من ذلك ، لا تزال هناك مشكلة تكرار المؤشرات و استخدامها غير المتسق في سياق الجامعات العراقية. سعيا للاجابة على سؤال البحث المتعلق بتحديد مؤشرات التعقيد المعجمي الأكثر قدرة على التنبؤ بقابلية قراءة نصوص اللغة الانكليزيه بوصفها لغة أجنبية لدى طلبة الجامعات العراقيه. تهدف هذه الدراسة الى تطوير نموذج من مؤشرات التعقيد المعجمي يعكس مقروئية النصوص العراقية. تم اختيار

أربعة عشر نصا قرائيا من كتاب Select Readings: Intermediate، و تم حساب قيم ثمانية مؤشرات للتعقيد المعجمي و مؤشرين خاصين بالمقروئية مخصصة للغة الانكليزية بوصفها لغة أجنبية باستخدام البرمجيات الحاسوبية Coh–Metrix و TAALED و LCA. تم تطبيق أسلوب الانحدار الجزئي بالحد الأدنى للمربعات الصغرى (PLSR) و تحليل أهمية المتغيرات في الاسقاط (VIP) لكل مؤشر من مؤشرات التعقيد المعجمي بالنسبة لقيم المقروئية ، استنادا الى الترابط المتعدد بين المتغيرات. و جرى اعتماد معيارين لتحديد المؤشرات المضمنة في النموذج: دقة النموذج و قيم VIP في النماذج الناتجة. استنادا الى نتائج PLSR و VIP تم تضمين ثلاثة مؤشرات في النمذج النهائي ، و هي: lexical_density_types و P_Lex و MATTR. يعالج هذا النموذج الحساس للمقروئية مشكلة عدم الاتساق في الدراسات السابقه ، و يقدم نموذجا احصائيا أكثر دقة لحساب التعقيد المعجمي.

Abstract

In the existing literature, researchers have constructed a large number of lexical complexity (LC) indices in order to predict the readability of texts in English as a foreign language (EFL) contexts. However, in the Iraqi context, there still exists the problem of redundancy of indices and their use. Attempting to answer the research question of which lexical complexity indices are most predictive of the readability of Iraqi EFL university-level reading texts, the present study aims to develop a model of LC indices that can reflect the readability of Iraqi reading texts. A total of 14 reading texts from the Select Readings: Intermediate textbook were selected. The scores of eight LC and two EFL-specific readability indices were computed using the computational software Coh-Metrix, TAALED, and LCA. Based upon the multicollinearity of variables, partial least squares regression (PLSR) and variable importance for projection (VIP) were utilized for each LC index in accordance with readability scores. To determine the most predictive LC indices in the final model, two criteria were applied including model accuracy and VIP scores across the resulting models. The results of PLSR and VIP indicated three indices as the more readability sensitive: lexical_density_types, P_Lex, and MATTR. This readability-sensitive model addresses the issue of inconsistency in previous studies as it provides a statistically more accurate model for calculating LC.

1. Introduction

Lexical Complexity (LC) is "a multifaceted, multidimensional, and multilayered phenomenon that has cognitive, pedagogical, and linguistic

Journal of Babylon Center for Humanities Studies: 2026, Volume: 16, Issue: 9

dimensions, developmental and performance aspects, and can manifest itself at all levels of language structure, learning, and use" (Housen, 2014, p. 63). *Reading proficiency evaluation through LC has been widely used in the fields of computational linguistics and natural language processing (Chen, 2019). Within the frameworks of English as a foreign language (EFL) teaching and learning, the construct of LC plays an important role in determining foreign language proficiency (*Bozdağ & Kilimci, 2023). *Concerning the connection between LC and the quality of EFL reading, many researchers (e.g.,* Erarslan, 2021*;* Bakuuro, 2024*) have found a significant correlation between them.*

Theoretically, LC, also referred to as lexical richness, is a multidimensional construct that consists of three sub-dimensions, including lexical density (LD), lexical sophistication (LS), and lexical variation (LV) (Lu, 2012). Based on these dimensions, researchers have developed a large number of different quantitative measures primarily because they have attempted to approach LC from various perspectives (Lu, 2012). What is more, there exist different calculation methods for the same LC dimension (Yang & Zheng, 2024). For instance, lexical density is calculated as the proportion of content words relative to the total number of words (Ure, 1971), while it is also measured by dividing the number of lexical items by the number of ranking clauses (Halliday, 1985). This has resulted in considerable proliferation of LC indices within the literature. Within the scope of the LC studies conducted in Iraq, measures and indices are used inconsistently, particularly when it comes to the readability of Iraqi university-level reading texts. For example, Boskany and Ahmed (2025) studied LC in relation to readability from the LD perspective using Ure's index, excluding LV and LS. Nonetheless, other researchers, such as Wafaa and Rana (2018), investigated LC and reading comprehension, excluding LD and LV, but including LS. However, little is known about which LC indices best reflect readability in Iraqi EFL contexts. Therefore, it is necessary to organize the indices available in the literature and develop a new model of indices that is sensitive to the readability of texts found within the Iraqi context. This work reviews some of the LC and readability measures and indices in the literature, and tries to determine which indices best reflect the readability of Iraqi EFL university-level reading texts.

2. Research Questions

The present study seeks to answer the following research questions:

1-Which lexical density indices most predict the readability of Iraqi EFL university-level texts?

2- Which lexical variation indices most predict the readability of Iraqi EFL university-level texts?
3-Which lexical sophistication indices most predict the readability of Iraqi EFL university-level texts?

3. Literature Review

3.1 Indices for Lexical Complexity Measurement

In EFL research, LC has been proposed as a valid descriptor of EFL performance, an indicator of proficiency, and as an index of linguistic development (Bulté & Housen, 2014). Within the current available literature, a large body of research into the operationalization of LC exists (e.g., Read, 2000; Lu, 2012; Bulté & Housen, 2012). Several scholars have defined LC as lexical diversity or lexical richness (Bulté & Housen, 2012; Li & Zhang, 2021). However, Lu (2012) considers LC a multidimensional construct that consists of three quantitative dimensions, namely LD, LV, and LS. In contrast, Read (2000) includes "lexical errors" as a fourth dimension alongside LD, LV, and LS. Moreover, Bulté and Housen (2012) introduce "lexical compositionality" as another dimension of LC. The model of Lu (2012) intentionally omits lexical errors and lexical compositionality, focusing on dimensions that can be approached computationally. For instance, the Lexical Complexity Analyzer (LCA) (Lu, 2012) operationalizes the framework by solely providing measures of density, variation, and diversity. The inclusion of clearly defined and countable linguistic features makes the framework of Lu (2012) gain precision in EFL research. As indicated above, LC typically encompasses three dimensions: LD, LV, and LS. Therefore, the present study investigates the predictive power of these dimensions when it comes to the prediction of readability in Iraqi EFL texts.

3.1.1 Lexical Density

Among the dimensions of lexical complexity, LD is the one that has been least employed in vocabulary research (Kyle et al., 2010). The indices of LD compare the number of content words to the total number of words in a given text, providing a measure of the informational density in a text (Liu & Dou, 2023). Capitalizing on these two linguistic factors, texts encompassing a larger proportion of content words are often associated with increased information packaging, complexity, and cognitive effort (Johansson, 2009; Bloomfield et al, 2010). Earlier research (e.g., Ure, 1971) has indicated that LD indices help distinguish between spoken and written modes of language. Ure concludes that the large majority of spoken texts have an LD of under 40% while a large majority of written texts have an LD of 40% or higher. Since the coinage of the term "lexical

density," various measurement variants have been proposed. A popular variant that is wieldy applied in educational research is the Ure's index. The index is calculated by dividing the number of content words (i.e., lexical words) by the total number of words (Ure, 1971). This measurement solely focuses on content-bearing vocabulary while excluding functional or grammatical words.

3.1.2 Lexical Sophistication

Lexical sophistication, also referred to as lexical richness, is defined as the "number of low-frequency words that are appropriate to the topic and style of writing" (Read, 2000, p.200). According to Lu (2012), lexical sophistication is likely a multidimensional construct, as distinct lexical features (e.g., word frequency and word familiarity) may measure the same construct (e.g., word exposure). According to Kim et al. (2018), there are numerous lexical features that function as indicators of LS, including bigrams (i.e., association strength), content word properties, word specificity, and frequency. However, despite the existence of many measurement methods that employ a variety of lexical features, there has been little agreement beyond word frequency. Frequency, in particular, is a reliable indicator of adequate reading comprehension (Husna et al., 2025) and lexical proficiency (Kim et al., 2018) in EFL settings. Lexical sophistication indices targeting frequency employ frequency lists known as reference corpora (e.g., BNC, COCA, and CELEX) to benchmark words in a target text (target corpora) with words found in the reference corpora (Chen, 2019). Advancements in the fields of corpus linguistics Computational linguistics and corpus linguistics led to the introduction of numerous LS indices that employ various reference corpora in their calculation, including the Lexical Frequency Profile (LFP) (Laufer & Nation, 1995), LS1 and LS2 indices (Hyltenstam, 1988; Laufer & Nation, 1995), and P_Lex (Meara & Bell, 2001).

P_Lex is an EFL-specific measure that combines LV and LC insights in its calculation (Meara & Bell, 2001). P_Lex is based upon the idea that it is possible to take advantage of the fact that sophisticated words occur infrequently in texts (Meara & Bell, 2001). It adopts a distributional method that targets word frequencies in 10-word segments and returns an index that shows the likelihood of the occurrence of these words in a reference corpus (Meara & Bell, 2001). In addition to P_Lex, Linnarud's index is another method to calculate LS. The index is calculated by dividing the number of sophisticated lexical items (Nslex) by the total number of lexical words (Nlex) (Yang & Zheng, 2024). The index is available for automatic calculation in the LCA (Lu, 2012), and it is referred to as "Lexical Sophistication-I" (LS1). Laufer and Nation (1995)

introduced an additional method in Lexical Frequency Profile (LFP) that involves the division of sophisticated word types (Ts) by the total number of word types (T). This index for LS measurement is referred to as "Lexical Sophistication-II" (LS2) within the LCA.

3.1.3 Lexical Variation

Lexical variation (LV), also referred to as lexical diversity (e.g., Bestgen, 2023), is defined as "the variety of the different words used in a text" (Read, 2000, p. 200). A higher LV means that the text possesses a more varied and diverse vocabulary (Johansson, 2009). Within the literature, different indices have been developed to quantitatively measure LV. However, the most operationalized index for the calculation of LV is what is known as the simple type-token ratio (TTR) (Kyle et al., 2020). Lexical variation indices like TTR measure the variety of lexical items in a text, which is considered to be a reflection of the lexical knowledge of the writer or the reader of that text. Moreover, they are indicators of writing quality, speaker competence, hearing variation, and speaker socioeconomic status (McCarthy & Jarvis, 2010).

A well-documented issue of TTR is its sensitivity to variations in text length (McCarthy & Jarvis, 2010). The issue of text length insensitivity become more evident in longer texts where the values of TTR become less stable due to the repetitive nature of longer texts where the same words appear more frequently (Majumdar, 2025). Recent years have seen the development of more robust approaches that address this issue by employing various statistical corrections (e.g., segmentation, curve fitting) or more advanced computational methods. These approaches are aimed at normalizing this ratio and making it more stable across the texts of different lengths (Majumdar, 2025). The ever-evolving of advanced methods led to the creation of updated variants of TTR, such as Measure of Textual Lexical Diversity (MTLD) (McCarthy & Jarvis, 2010), Hypergeometric Distribution D (HD-D) (McCarthy & Jarvis, 2007), and Moving-Average Type-Token Ratio (MATTR) (Covington & McFall, 2010).

3.2 Readability

Readability refers to the ease of reading, particularly as it results from a writing style (Klare, 1963). In EFL contexts, the readability of texts plays a dual role. It helps language teachers in simplifying incomprehensible texts for the targeted students and supports educators in designing reading comprehension materials with an appropriate level of difficulty (Zainurrahman et al., 2024). An extensively utilized instrument for readability assessment, which has proven its worth in over 80 years of application, is the readability formulas (DuBay, 2004). According to

McLaughlin (1969), a readability formula is a mathematical equation that is derived through regression analysis. While traditional readability formulas were originally developed for native English speakers, their applicability in EFL contexts has been a subject of investigation (Lee & Lee, 2020). In this regard, Brown (1997) found that the first language readability indices were not accurate in predicting EFL difficulty. As a response to this need, researchers (e.g., Greenfield, 2003; Crossley et al., 2008) have developed indices that more accurately predicted EFL readability.

The Miyazaki EFL Readability index (MEFLRI) (Greenfield, 2003) adopts the same variables as the Flesch and Flesch-Kinkaid formulas, but adjusts them in a way that reflects the reading performance of EFL learners (Greenfield, 2003). The formula for calculating the MEFLRI uses letters per word and words per sentence to provide a score on a 100-point scale, where higher scores suggest easier texts (Greenfield, 2003).

The synthesis of findings in the fields of computational linguistics and corpus linguistics has paved the way for new measurement approaches that move beyond the conventional linguistic features. A representative of such advancements is the Coh-Metrix (Graesser et al., 2004). Coh-Metrix is a computer facility that "measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis" (Crossley et al., 2008, p. 480). Despite the availability of more than 100 measures in Coh-Metrix, the one that stands out is the Coh-Metrix L2 reading index (RDL2) (see Crossley et al., 2008 for the full description of RDL2). According to Kiselnikov et al. (2020), RDL2 is an index that reflects the cognitive and psycholinguistic processes involved in reading. RDL2 is derived from the regression of three banks of indices, including lexical index, syntactic index, and meaning construction index (Crossley et al., 2008). The lexical index is concerned with word frequency as measured through CELEX corpus (Crossley et al., 2008). It accounts for word familiarity, based upon the findings that frequent words are processed more quickly than the infrequent ones (Haberlandt & Graesser, 1985). The syntactic index captures the similarity of syntactic constructions between adjacent sentences, grounded in research showing that more predictable and consistent sentence structures ease processing demands (Perfetti et al., 2005). The meaning construction index calculates content word overlap across adjacent sentences (Crossley et al., 2008). Research has shown that overlapping vocabulary is an important aspect in text processing, leading to better text comprehension and increased reading speed (Rashotte & Torgesen, 1985).

4. Methodology

4.1 Research Design

This research adopted a quantitative predictive research design. Partial least squares regression (PLSR) was used to statistically analyze the data, as it was necessary to examine how each of the components explains and predicts readability on different models. The strength and stability of each LC index in all components were also identified using variable importance of projection (VIP).

4.2 Instruments

The instruments of the study were chosen and validated to answer the research questions. The computational analysis of the data was carried out using a variety of instruments, including Coh-Metrix, Lexical Complexity Analyzer (LCA), TAALED, P_Lex, letter counter, and RStudio. The basis for using the different instruments is that each one is designed to capture different linguistic aspects. Also this combination of instruments helps to enhance validity, reduce instrument bias, and deliver a more accurate measurement of LC and readability.

The letter count utility available at https://www.lettercount.com/ was used to obtain the total number of letters in each text. The letter count was mandatory in the calculation of the MEFLRI, along with other statistics such as word count, sentence count, and polysyllabic word count, which were extracted using the Gorby Text Analysis Tools.

For computational text analysis, the desktop version of Coh-Metrix was employed. The tool was provided through direct contact with its developers from Arizona State University. It was used to calculate one of the two readability indices, namely RDL2. In addition, LCA, accessed through its official online variant, was utilized to calculate the LD index as well as two lexical sophistication indices (LS1 and LS2).

Moreover, the Tool for the Automatic Analysis of Lexical Diversity (TAALED) was employed to calculate MATTR, MTLD, and HD-D indices for LV as well as one index for LD (lexical_density_types). The rationale for utilizing TAALED is that its indices are extensively validated and extensively tested in Zenker and Kyle's (2021) study. The Python variant of TAALED was accessed after following the setup guide provided by the LCR-ADS lab at the University of Oregon. Additionally, P_Lex version 3.0, accessed through its official online workspace, was used to measure the P_Lex index.

Lastly, statistical analysis was performed with R (version 4.3.3) in the RStudio integrated development environment. All statistical processes

involved in the PLSR were conducted using RStudio. These analyses enabled an accurate and reproducible investigation of the variables under analysis.

4.3 Indices Used in the Study

The indices of the current work were selected based on their construct validity and adherence to EFL contexts. For each of the analyzed constructs, multiple indices were utilized in order to achieve a more robust and comprehensive analysis. Table 1 illustrates the indices utilized in this study.

Table 1.

Lexical Complexity and Readability indices utilized in this Study

Construct	Measure
Lexical Density	Lexical density Lexical_density_types
Lexical Sophistication	LS1 LS2 P_Lex
Lexical Variation	MATTR MTLD HD-D
Readability	RDL2 MEFLRI

4.4 Procedures

The procedures of the current study began by compiling a corpus of 14 texts from the "Select Readings: Intermediate" textbook. The extracted corpus was directly copied and pasted from the respective source and got prepared for computational processing. Following the extraction, the text-cleaning procedures were carried out by excluding the noisy text segments, such as figures, line numbers, and glossaries. The cleaned texts were then submitted for manual revision by a jury of experts from the University of Basra, College of Arts, Department of English, to avoid any spelling mistakes. To efficiently manage the extracted data, texts were stored and systematically labeled Text 1, Text 2, and so on. Each file was then saved using the .txt extension in order to ensure recognition by the computational software.

After text preparation being completed, an initial step was mandatory before the large-scale statistical analyses. This step involved performing computational processing through the selected instruments to compute the values of LC and readability indices. The computation of LS1, LS2, and

LD indices, was done using the single mode of LCA. The LCA software does not support direct file uploads, so texts were manually copied and pasted from their already prepared .txt files. Being pasted, the texts were then processed using the interface of the tool after selecting the required indices.

In addition to LCA, the software TAALED was employed to calculate the three LV indices, including HD-D, MTLD, and MATTR, as well as the lexical_ density_ types index for LD. For the calculation of P_Lex, P_Lex v3.0 was accessed via the official workspace of the tool. The texts were copied and pasted into the input section from the prepared .txt files. The tool individually processed the texts and returned a lambda score as an output. The default settings of the tool were used throughout, including the exclusion of punctuation characters such as commas, periods, colons, semicolons, exclamation points, and question marks.

For readability measurement, two EFL-specific indices were calculated including the RDL2 and Miyazaki EFL index. The RDL2 index was computed using the interface of Coh-Metrix (desktop version). This variant of the tool is not open access, so to obtain it, the researcher contacted the developers via e-mail, and the .exe file for the tool was provided. To avoid any malware that might affect the functionality of the tool, the .exe file was installed on a fresh Windows operating system. As to the researcher's best of knowledge, there is no computational software capable of automating the MEFLRI. As a solution the index was computed using the RStudio after applying its mathematical equation and the required linguistic variables. The required variables for the index were obtained using valid word-counting tools and stored in an Excel spreadsheet. Subsequently, the obtained variables were input into the interface of RStudio, where the final values were automatically computed.

The resulting data were further processed in RStudio. However, before proceeding to the large-scale analyses, suitable statistical functions in RStudio were performed to verify the assumptions of linearity and normal distribution. The aim of the analysis then shifted to investigating the strength and predictive power of LC indices in accordance with readability.

4.5 Statistical Analysis

Before calculating LC and readability, the general linguistic features, such as the number of letters, words, sentences, and polysyllabic words, were calculated. The mean, standard deviation, minimum, and maximum values were recorded to give a general overview of the structural characteristics of the analyzed texts. Descriptive statistics were also

calculated for all LC and readability indices. These statistics summarize the distributions of the indices and function as a preparatory step before the conduct of subsequent correlation and regression analyses.

A small dataset presents a challenge of how to make the most of the available data. In order to overcome this challenge, PLSR was utilized to develop two separate predictive models. For each model, the Root Mean Squared Error of Prediction (adjCV) was calculated to evaluate model accuracy and precision. Moreover, the proportion of explained variance by the two models was gauged to testify how well they accounted for the variability in readability. Within each model, the values of LC indices served as predictor variables while readability scores as dependent variables. The importance of the predictor variables was then calculated using Variable Importance in Projection (VIP). The VIP scores exceeding the threshold of one were identified as the most influential predictors of readability. Finally, based on the criteria of the lowest adjCV and threshold exceeding VIP scores across models, the model of the present study was constructed.

5. Corpus Selection

The corpus for the present computational analysis is represented by 14 university-level reading texts, which represent the total number of texts in the textbook Select Readings: Intermediate 2nd edition (Lee & Gundersen, 2011). The aforementioned textbook is prescribed by the Ministry of Higher Education and Scientific Research for second-stage students at the University of Basra, College of Arts, Department of English. Overall, the selected texts vary in terms of length, ranging from 645 to 896 words, with letter counts between 3,259 and 4,348. Sentence counts range from 32 to 55, and the average number of words per sentence falls between 15.1 and 20.2. The number of polysyllabic words also varies, ranging from 81 to 136. These basic statistics were obtained using a valid word-counting tool available for free online access.

6. Results

6.1Statistical Description of Derived Indices

To illustrate the values of LC and readability indices, the selected corpus went through a computational analysis using the instruments of the study, as in Table 1.

Table 1.

Raw values of Lexical Complexity and Readability Indices in the selected Corpus

Index	Mean	SD	Minimum	Maximum
Lexical_Density_Types	0.72	0.03	0.65	0.77

LD Index	0.22	0.009	0.21	0.23
P_Lex	2.13	0.62	1.16	3.205
LS1	0.87	0.01	0.84	0.9
LS2	0.60	0.03	0.57	0.67
MATTR	0.81	0.02	0.75	0.85
HD-D	0.82	0.02	0.78	0.85
MTLD	80.70	16.09	58.54	116.74
MEFLRI	44.52	9.42	28.15	63.96
RDL2	16.72	3.91	10.57	22.92

Note. SD= standard deviation

The results in Table 1 provide the descriptive statistics for the readability and LC indices utilized in this study. Concerning readability, the RDL2 index exhibited a mean value of 16.72 with a standard deviation of 3.91, and its values ranged from a minimum of 10.57 to a maximum of 22.92. The statistics of RDL2 suggest that the readability of the selected corpus is low, indicating high reading difficulty for Iraqi EFL learners. The MEFLRI demonstrated a mean of 44.52, with a standard deviation of 9.42, and values ranging from 28.15 to 63.96, which means that readability is moderate, signifying relatively easy-to-read texts.

When it comes to the LD indices, the LD index showed a mean of 0.22, with a low standard deviation of 0.009, and values ranging between 0.21 and 0.23. The low values of the LD index suggest that the LD of the selected corpus is low, indicating that texts contain a small proportion of content word tokens. The lexical_density_types showed an average value of 0.73, with a standard deviation of 0.035 and scores ranging between 0.65 and 0.77, suggesting a large proportion of content word types. When it comes to the LS indices, LS1 averaged 0.87 (SD = 0.019), falling within a fairly narrower range of 0.84 to 0.90, revealing a large proportion of sophisticated, advanced, and domain specific lexical items. LS2, on the other hand, came out slightly lower, with a mean of 0.60 and a standard deviation of 0.032, ranging from 0.57 to 0.67. The results of LS2 imply that the selected reading texts contain an average number of sophisticated word types. Interestingly, P_Lex appeared to fluctuate more than the other two LS measures, as it showed the widest spread of values. Its mean stood at 2.14, but with a notably higher standard deviation of 0.63, ranging from 1.16 1.16 to 3.21. The results of P_Lex point to the existence of a moderate proportion of sophisticated words in different segments of the texts.

6.2 PLSR Models for Predicting Readability

Journal of Babylon Center for Humanities Studies: 2026, Volume: 16, Issue: 9

To address the research question of the present study, PLSR was utilized to develop predictive models that examined the influence of LC indices (predictors) on readability indices (criteria). Separate PLSR models were constructed for each readability index to isolate the contribution of individual LC indices in explaining variance in text readability. Besides, Root Mean Squared Error of Prediction (RMSEP) through adjusted cross-validation (adjCV), as well as variance explained on the Y axis, were computed for each model to determine prediction accuracy as shown in Table 2.

Table 2.

PLS Model Performance Metrics for Predicting Readability Indices across Latent Components

Number of Components	RDL2 adjCV	RDL2 Variance (%)	Miyazaki adjCV	Miyazaki Variance (%)
1	4.212	8.839	10.11	4.878
2	3.339	44.69	10.93	8.887
3	2.435	74.37	8.569	65.12
4	2.671	82.52	8.167	82.21
5	3.248	83.18	8.304	85.88
6	3.871	83.19	8.198	86.2
7	4.945	83.23	9.548	87.38
8	5.253	83.23	12.76	87.77

Table 2 demonstrates the performance of the RDL2 model across eight latent components. As the number of components increases from one to three, the adjCV decreases (from 4.236 to 2.444), while the explained variance rises substantially, from 8.83% to 74.37%. The model reaches its lowest adjCV at three components, indicating optimal predictive accuracy. When exceeding three components, the variance explained is around 83%, while the adjCV increases. The results here suggest that exceeding the optimal number of components might lead to potential model overfitting. As a result, the RLS2 model with three latent components is preforming well since it balances between complexity and prediction accuracy.

The performance of the MEFLRI model, on the other hand, fluctuates initially in the first component with an explained variance of 4.878. However, it starts to perform better beyond the first component. The MEFLRI does no show a clear gradation of performance throughout components. The model reaches its minimum at four components (adjCV=8.167) with an increase in the explained variance from 4.88% at

the first component to 82.21% at the fourth component. Beyond the four components, the RMSEP increases again, suggesting potential overfitting. When including additional components, the highest variance explained occurs at the eighth component (87.77%); however, with a high prediction error. The results here suggest that the inclusion of four components achieves the optimal balance between model complexity and predictive accuracy for the MEFLRI model.

Variable (X)	RDL2 Model (Y)	MEFLRI Model (Y)
lexical_density_types	1.49	1.54
LD Index	0.15	0.26
P_Lex	1.84	0.87
LS1	0.24	0.66
LS2	0.54	1.19
MATTR	0.84	1.14
HD-D	0.62	1.07
MTLD	0.93	0.66

To describe the relative importance of each LC index to the predictive power of the RDL2 and MEFLRI models, the Variable Importance for Projection (VIP) statistics were calculated as in Table 3.

Table 3.

The VIP Statistics for Lexical Complexity Indices in the RDL2 and MEFLRI Models

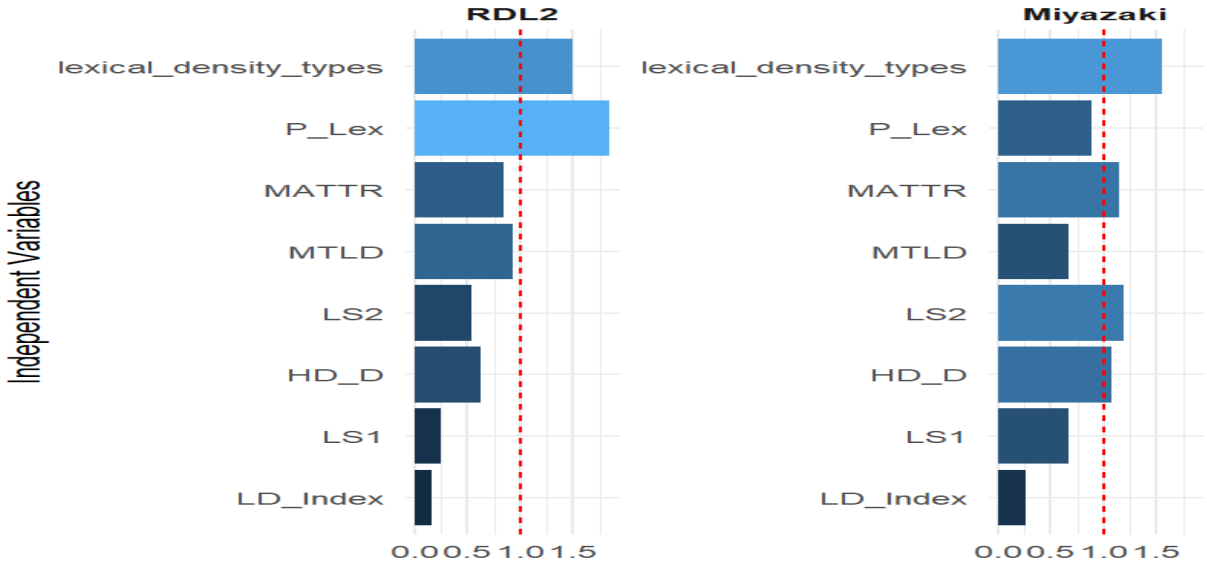
Note. X variables with VIP values greater than one are more important in predicting the Y variables.

The results in Table 3 show how each X variable (LC indices) predict the scores for the two readability models. In the RDL2 model, P_Lex (1.849) is the strongest predictor, while LS2 (0.544) and LS1 (0.244) have a lower relative importance. LD Index (0.158), MTLD (0.933), and MATTR (0.849) demonstrate moderate contribution. For MEFLRI, HD_D (1.075), LS2 (1.194), and MATTR (1.141), and are the most influential, while LS1 (0.664), MTLD (0.667), and P_Lex (0.879), and have moderate contribution. Moreover, LD Index (0.268) and lexical_density_types (1.546) have a limited impact on the MEFLRI model.

The abovementioned VIP scores for each LC index within the two predictive models are further visualized in Figure 1.

Figure 1.

The Graphical Visualization of VIP Scores across the Two Models



6.3 Model of Readability-Sensitive Lexical Complexity Indices

Table 4 presents a comparison between the two resulting models in terms of the optimal number of components and the lowest adjCV achieved.

Table 4.

Comparison between the RDL2 and MEFLRI Predictive Models

Model	Optimal Components	adjCV	% Explained Variance (Y)
RDL2	3	2.435	74.37%
MEFLRI	4	8.167	82.21%

Table 4 shows that the optimal number of components, adjCV, and the percentage of explained variance in the dependent variable (Y) revealed noticeable differences in model performance. The RDL2 model achieved a strong balance between accuracy and simplicity. It explained 74.37% of the variance with only three optimal components. Furthermore, it exhibited a moderate RMSEP (2.435), which reflects reliable performance with limited complexity. The MEFLRI model achieved a higher explained variance (85.88 %) compared to RDL2 with four optimal components. However, its higher RMSEP (8.167) indicates potential sensitivity and less stable predictions.

Prioritizing RLD2 as a more robust model over MEFLRI, as well as the VIP scores yielded from the two models, the following unified readability-based model of LC indices is constructed as in Table 5.

Table 5.

Readability-Based Model of LC Indices

Dimension	Selected Index
Lexical Density	Lexical_density_types

Lexical Sophistication	P_Lex
Lexical Variation	MATTR

Table 5 shows the final model, which encompasses lexical_density_types, P_Lex, and MATTR as the principal predictors of readability. When taken together, these indices account for the strongest and most consistent variance across models. They represent the essential lexical features that shape the readability of the selected EFL texts.

7. Discussion

This study aimed to develop a readability-based model of LC indices for Iraqi EFL texts. The results of the three research questions revealed that among the selected lexical complexity indices, a few key indices accurately predicted the readability of the selected EFL textbook. In particular, the indices P_Lex, lexical_density_types, and MATTR emerged as especially powerful predictors. This finding shows that lexical factors are part of the constraints which make reading difficult for the Iraqi EFL learners. Texts containing which have a greater proportion or a variety of content word types are found to be more difficult to process (Bloomfield et al., 2010). This suggests that teachers and material designers must pay close attention to vocabulary load and variety, as few unknown or infrequent words can lead to increases in comprehension demands. As Crossely et al. (2010) noted, the number of different words and content word frequency in texts gradually increase across different proficiency levels. Conversely, the findings here imply that such features make texts less accessible to lower proficiency Iraqi readers.

Across the resulting two models, two indices related to two dimensions of LC stood out overall. Measures of LS and LD, represented by P_Lex and lexical_density_types, repeatedly ranked among the top-most influential predictors. This aligns with Yang and Zheng's (2024) refined and concise model of LC, which similarly retained an LS and a lexical density index as central indices in their model. In fact, Yang and Zheng found that LS1 had a significantly larger effect size than the LS2 index, so it remained within the model as a more accurate indicator of sophisticated vocabulary. However, the P_Lex index, in contrast to LS1, proved more influential to the selected EFL dataset, even though not included in Yang and Zheng's writing-focused model. This signifies that, in addition to the overall number of sophisticated words, the segments of text encompassing advanced words can also slow down reading comprehension for Iraqi EFL readers. When taken together, the dominant indices point to the variation and sophistication of content vocabulary as

Journal of Babylon Center for Humanities Studies: 2026, Volume: 16, Issue: 9

the strongest linguistic drivers of reading difficulty for the respective EFL context.

In parallel with EFL reading and psycholinguistic theories, the current research emphasizes that encountering unfamiliar vocabulary increases processing effort. As Crossley et al. (2008) noted, EFL readability indices such as RLD2 that adopt psycholinguistic variables (e.g., word frequency) better predict reading difficulty when compared to the traditional approaches. The findings of the present work are in line with Crossley et al. (2008) in that LC indices associated with psycholinguistic variables are important in predicting readability as measured using RDL2. Moreover, the Lexical Quality Hypothesis suggests that the existence of a large proportion of unknown words disrupts comprehension (Perfetti, 2007). Additionally, the concept of lexical "threshold" (Laufer & Ravenhorst-Kalovski, 2010) suggests that the encounter with a large number of infrequent words constitutes an impediment, preventing learners from attaining the 98% of coverage required for a comfortable reading.

The significance of P_Lex during regression analysis signals that reading comprehension for Iraqi EFL learners is constrained by vocabulary frequency. Higher P_Lex scores indicate many difficult words per text segment, which in turn tax working memory and parsing processes. It is argued that better readability assessment involves decoding, syntactic parsing, and meaning construction, not just simple text-internal variables (Allen & McNamara, 2011). The resulting findings of the current study fit this view. By pinpointing LD, LS, and LV, these results capture exactly those deeper semantic and conceptual burdens Iraqi EFL learners face when establishing meaning.

Precision and interpretability are two fundamental factors in model development. The inclusion of multiple LC indices in a predictive model might probably lead to maximal statistical fit. However, such an approach would lead to an overfitted model with no generalizable results. Instead, the current study prefers a more streamlined approach. Mirroring the methodology of Yang and Zheng (2024), the present study emphasizes the few indices that best predict readability in the selected dataset. Their refined and concise model utilizes only six indices to "provide a more accurate and efficient tool" for LC assessment. Likewise, the PLSR analyses here suggest that a precise and concise model of predictors (e.g., focusing on P_Lex, lexical_density_types, MATTR, and perhaps one or two additional indices) explains most variance while remaining interpretable.

Practically speaking, a simpler model is also more useful for educators as teachers and curriculum designers can adequately track a handful of indices, whereas dozens of technical metrics would be impractical. The trade-off here is that the addition of more variables to a model may slightly improve its R^2 values, but at the cost of opacity. In a classroom setting, it is more important to diagnose the primary factors that influence reading comprehension. For example, a text with a high percentage of sophisticated words (as calculated by P_Lex) will likely require additional attention and support from the educators.

Building upon what is discussed, teachers should recognize that lexical factors drive reading difficulty for EFL learners rather than just shallow text-internal features like sentence lengths. For instance, a high value of LS1 in a text flags the existence of many advanced and domain-specific words. Suggesting pre-teaching or glossaries in this case is needed. Lexical density is a benchmark for the information load in a text. A text packed with nouns and adjectives may convey many ideas in just a few sentences; a scaffolding might involve deconstructing the text into smaller, manageable chunks. Computational tools that compute P_Lex or lexical_density_types, and MATTR can help curriculum designers and teachers select appropriately leveled reading texts. For example, texts with lower P_Lex and slightly lower density may be better for less proficient readers. In parallel, educators should not view LC as an all-inclusive construct; rather, understanding which lexical dimensions and features matter most leads them to adjust materials more effectively.

8. Conclusions

This study developed an EFL-specific readability-sensitive model based on three LC indices: lexical_density_types, P_Lex, and MATTR. The results of PLSR and VIP for Iraqi university-level EFL reading texts across models confirmed that each of these indices had a meaningful contribution to predicting reading difficulty. Including multiple LC indices in a model across the dimensions means it operates cohesively. The multidimensionality of LC is well documented in the current literature. Capitalizing on this, the model of this study targets the three dimensions of LC including LD, LS, and LV. This leads to the conclusion that all three contributed unique, statistically significant predictive power, which in turn reinforces the internal coherence of the model. Empirically speaking, the model provides an explanation for why Iraqi university-level EFL reading texts are more difficult for Iraqi learners. Texts with higher lexical_density_types, higher P_Lex, and higher MATTR were consistently rated as harder. In sum, the proposed model provides a model of readability-sensitive LC indices that is empirically grounded,

conceptually coherent, and gives an account for how LC drives reading challenges in the Iraqi EFL university-level context.

9. Pedagogical Implications

The findings of this study show that texts containing a large proportion of content words (high LD), rare vocabulary (high LS), and high vocabulary diversity (high LV) constitute most of the reading difficulty for Iraqi EFL learners. Teachers in the Iraqi EFL context need to prioritize these three factors when selecting or preparing texts for their lessons. For texts that contain a large proportion of advanced and domain-specific words, Instructors can implement glossaries and word lists to pre-teach key vocabulary to lower the cognitive load of the students. This aligns with the principles of the lexical quality hypothesis, which suggests that unknown words slow comprehension; therefore, ensuring familiarity with most content words is critical.

EFL-specific readability formulas like MEFLRI are a leap forward from the traditional methods. However, they still implement the same linguistic features found in the traditional formulas. When compared to the more advanced indices such as RDL2, MEFLRI proved to be less accurate in reflecting the underlying lexical features of the selected texts. Iraqi Curricula designers and teachers should prioritize readability indices with underlying cognitive and psycholinguistic assumptions (e.g., RDL2). Current research highlights that cognitively based readability indices, in this case RDL2, are more accurate in predicting readability than EFL-specific formulas. From a practical perspective, this means that evaluating texts with indices that account for word frequency, content word overlap, and syntactic similarity will yield better results that adhere to the lexical difficulties faced by Iraqi EFL readers. Teachers can capitalize on texts identified as hard by RDL2 contain many infrequent words, alerting teachers to provide additional support.

The present study found that specific indices such as MATTR, lexical_density_types, and P_Lex are the most predictive of EFL reading difficulty. Teachers can use computational tools or corpus scripts to calculate these indices for their materials. Texts with high values of P_Lex signal many advanced words that are distributed across different sections of the text. In their lesson preparations, teachers should treat such texts with caution by either simplifying them or breaking them into smaller sections and building background knowledge. In sum, tracking rare-word counts explicitly helps teachers identify which readings require extensive vocabulary scaffolding.

10. Limitations of the Study

As in all studies, this one also comes with limitations. First, there is the question of the generalizability of results. Although the results of PLSR were validated in terms of adjCV, their validity does not hold when using different readability or lexical complexity indices. Future studies would benefit from other validated indices not included in the present analysis to further expand the area of investigation, as different indices might exhibit different results. Second, there is the question of the corpus size, which consisted of 14 academic reading texts taken from an EFL textbook. Any index based on such a corpus with similar texts (i.e., academic) can be expected to show similar results (Crossley et al., 2008). However, future research would do well by using a larger corpus along with expert ratings or subjective judgments of learners.

References

Abdulla Boskany, S., & Ahmed, S. (2025). A Corpus Linguistic Study of Lexical Density and Readability of Selected Short Stories. (0AD). Iraqi Journal of Humanitarian, Social and Scientific Research Volume 5, Issue 17A (2025).

Allen, D., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1), 84–101. https://doi.org/10.64152/10125/66657

Bakuuro, J. (2024). In the belly of text complexity: Unravelling the nexus between lexical density and readability. *Athens Journal of Philology*, 11(3), 255–274. https://doi.org/10.30958/ajp.11-3-4

Bestgen, Y. (2023). Measuring lexical diversity in texts: The twofold length problem (arXiv:2307.04626). *arXiv*. https://doi.org/10.48550/arXiv.2307.04626

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). What makes listening difficult? *The University of Maryland*. https://apps.dtic.mil/sti/pdfs/ADA550176.pdf

Bozdağ, F., & Kilimci, A. (2023). Lexical complexity and language proficiency: An investigation of indices across CEFR levels.

Brown, J. D. (1997). An EFL readability index. *University of Hawai'i Working Papers in English as a Second Language*, 15(2), 1–17.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA (pp. 21–46). *John Benjamins*. https://doi.org/10.1075/lllt.32.02bul

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Chen, X. (2019). Automatic Analysis of Linguistic Complexity and Its Application in Language Learning Research.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. https://doi.org/10.1080/09296171003643098

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493. https://doi.org/10.1002/j.1545-7249.2008.tb00142.x
Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, 60(3), 573–605. https://doi.org/10.1111/j.1467-9922.2010.00568.x
Department of Linguistics and Phonetics, 53.
DuBay, W. H. (2004). *The principles of readability*. Impact Information.
Erarslan, A. (2021). Correlation between metadiscourse, lexical complexity, readability and writing performance in EFL university students' research-based essays. *Shanlax International Journal of Education*, 9(S1), 238–254. https://doi.org/10.34293/education.v9iS1-May.4017
Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202. https://doi.org/10.3758/BF03195564
Greenfield, J. (2003). The Miyazaki EFL readability index. *Comparative Culture, 9*, 41–49.
Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114(3), 357–374. https://doi.org/10.1037/0096-3445.114.3.357
Halliday, M. A. K. (1985). *Spoken and written language*. Deakin University.
Housen, A. (2014). Difficulty and complexity of language features and second language instruction. In Chapelle, C., editor, *The Encyclopedia of Applied Linguistics*, pages 2205-2213. Wiley-Blackwell.
Husna, F. H., Hartono, R., & Sakhiyya, Z. (2025). A corpus-based analysis of vocabulary load and coverage in Indonesian EFL textbook for 8th grade. *Acuity: Journal of English Language Pedagogy, Literature and Culture, 10*(3), 205–219. https://doi.org/10.35974/acuity.v10i3.4011
Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development, 9*(1–2), 67–84. https://doi.org/10.1080/01434632.1988.9994320
Johansson, V. (2009). Lexical diversity and lexical density in speech and writing. *Lund University, Department of Linguistics and Phonetics*, 53.
Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *Modern Language Journal, 102*(1), 120–141. https://doi.org/10.1111/modl.12447
Kiselnikov, A., Vakhitova, D., & Kazymova, T. (2020, August). Coh-Metrix readability formulas for an academic text analysis. *IOP Conference Series: Materials Science and Engineering*, 890(1), 012207. https://doi.org/10.1088/1757-899X/890/1/012207
Klare, G. R. (1963). *The measurement of readability*. Iowa State University Press.
Kyle, K., Crossley, S. A., & Jarvis, S. (2020). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly*, 18(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205
Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

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. https://doi.org/10.64152/10125/66648
Lee, B. W., & Lee, J. (2020, August 1). LXPER Index: A curriculum-specific text readability assessment model for EFL students in Korea. *arXiv*. https://doi.org/10.48550/arXiv.2008.01564
Lee, L., & Gundersen, E. (2011). *Select Readings: Student Book Intermediate* (2nd ed.). Oxford University Press.
Li, X., & Zhang, H. (2021). Developmental features of lexical richness in English writings by Chinese L3 beginner learners. *Frontiers in Psychology*, 12, 752950. https://doi.org/10.3389/fpsyg.2021.752950
Linnarud, M. (1986). Lexis in composition: A performance analysis of Swedish learners' written English. *CWK Gleerup*.
Liu, Z., & Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective. *Frontiers in Psychology, 14*, 1276705. https://doi.org/10.3389/fpsyg.2023.1276705
Lu, Xiaofei (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, 96(2), 190-208.
Majumdar, P. (2025). Type-token ratio (TTR) – A measure of lexical diversity. *Zenodo*. https://doi.org/10.5281/zenodo.15770103
McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. Language Testing, 24(4), 459–488. https://doi.org/10.1177/0265532207080767
McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd 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
McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, 12(8), 639–646.
Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect, 16*, 5–24. Retrieved from https://www.lognostics.co.uk/vlibrary/meara%26bell2001.pdf
Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. https://doi.org/10.1080/10888430701530730
Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The Acquisition of Reading Comprehension Skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook (pp. 227–247). Blackwell Publishing.* https://doi.org/10.1002/9780470757642.ch13
Rashotte, C., & Torgesen, J. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, 20, 180-188.
Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942
Ure, J. N. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics: Selected papers of the 2nd International Congress of Applied Linguists, Cambridge 1969* (pp. 443–452). Cambridge University Press.
Wafaa, S. I., & Rana, A. A (2018). Iraqi EFL university learners' vocabulary size in relation to their reading comprehension performance. *In Proceedings of the 9th International Visible Conference on Educational Science and Applied Linguistics*. https://doi.org/10.23918/vesal2018.a18

Yang, Y., & Zheng, Z. (2024). A refined and concise model of indices for quantitatively measuring lexical richness of Chinese university students' EFL writing. *Contemporary Educational Technology*, 16(3), Article ep513. https://doi.org/10.30935/cedtech/14707

Zainurrahman, Z., Yusuf, F. N., & Sukyadi, D. (2024). Text readability: Its impact on reading comprehension and reading time. *Journal of Education and Learning (EduLearn)*, 18(4), 1422–1432. https://doi.org/10.11591/edulearn.v18i4.21724