TEXT MINING SALMAN RUSHDIE AND AMITAV GHOSH: VOCABULARY DENSITY ANALYSIS FOR ENGLISH LEARNERS

Dr. GARRET RAJA IMMANUEL S

Assistant Professor, Department of English
Nazareth Margoschis College at Pillaiyanmanai

Abstract

In this research, the researcher focuses on how to choose proper texts for teaching English to Indian learners with E.D. Hirsch's hypothesis in mind that focusing on the culturally accessible local textsthat improves learning. To achieve this, the research seeks to explore Rushdie and Ghosh's texts using text mining tools such as Voyant Tools as a way of identifying the simplest texts that a beginner can read. Vocabulary density is determined for five books of each author to analyze suitability of texts for beginners. Descriptive analysis and Inferential analysis are used to compare and contrast the vocabulary complexity of the selected texts using t-tests and analysis of variance (ANOVA). Moreover, this research underlines the importance of culture in language learning at the same time offering the matrix of texts with different level of cultural context simplification but having the same linguistic complexity level. The study's findings are intended to help educators and curriculum developers write effective teaching strategies for English to Indian learners.

Keywords: *text mining, vocabulary density, indian, elt, voyant*

Introduction

Choosing right text for ELLs plays a very important role for children's academic achievement as well as their motivation. The extreme is that the texts that adopted to learners' reading level make learners influenced, decrease their comprehending ability and decrease the learners' self-confidence and on the other hand texts which are far away from the learners' text difficulty level does not challenge learners or help the learners' language development. The criteria used to select appropriate texts bear both linguistic and cognitive facets and it was necessary to develop the learners' backgrounds, interests, and language proficiency.

Studies show the need to understandably balance the type of texts used in learning processes needed among learners. For instance, the function of text difficulty and learner performance is not linear since comprehension is a function of text, task, and teachers' support (Bunch et al., 2014). The self-directed tools employed in the categorization of texts with regard to learner's comprehension have entailed characteristics such as the software's ability to use lexical and syntactic density to correctly align learners with suitable material autonomously (Kurdi, 2017). But even with these tools,

the teacher must look at other aspects like familiarity to the culture and the use of the text, all of which affect understanding according to Johnson (1981).

Difficulties are expected to occur if texts are set at a wrong level in respect to learners. Texts may overload students and cause frustration as they fail to navigate syntax or read unfamiliar words (Potential problems of text complexity: exploration across reading levels, 2021). On the other hand, simplified texts exclude them from extended contacts with complex language patterns and do not promote development of critical reading skills. Instruction that is expertly faded to include complex texts but only when some support is needed has been demonstrated not only to positively impact students' language development and content knowledge (O'Brien and Leighton, 2015).

Finally, based on the above findings, choosing of appropriate text should help learners be motivated while helping them achieve the next level of language mastery. Sustaining this balance involves broad and complex knowledge of the learner and the text at the same time with the common understanding that reading text ought not only foster development of language but also confidence and The interaction between the text difficulty and its

vocabulary and readability levels is important for language learning process. These affect comprehensibility, learning achievements and learner interest for a second language learner easily. Text complexity, encompassing syntactic and semantic challenges, plays a crucial role in shaping comprehension outcomes. Excessive complexity can hinder understanding, especially for learners with limited linguistic proficiency. For instance, studies have found that while lexical complexity negatively impacts comprehension in some contexts, syntactic complexity does not always exhibit a direct influence (Arya et al., 2011). Texts with balanced complexity allow learners to engage meaningfully without becoming overwhelmed, fostering incremental language development. This is why vocabulary density is just as important. Other research in this area show that learners who are exposed to a rich level of vocabulary score better in their comprehensions tests in texts they read and remember better. High-frequency word reading increases efficient word recognition, while new word acquisition improves learning when sufficed with proper support (Elleman et al., 2009). A meta-analysis pointed out that focused vocabulary instruction increases comprehension considerably upon the lessons have been tailored to the learners' needs (Wright &Cervetti, 2017).

Vocabulary and syntactic tests which are part of readability assessments give the extent of texts that should be used for learners. Readability is directly linked to interest and understanding; first in importance texts with a high readability can look complex because of their density or because of their depth and they do this so well that some readers can get disinterested. A number of benefits have been found with integrated systems that are able to modify text difficulty while maintaining variation in language use (Holley, 1973).

Therefore, the correspondence between text difficulty, lexical density and readability is perfect for language learning. Based on texts chosen as well as specific interventions used, learners are presented with materials that both challenge them and provide them with support to improve their language, literacy, and academic skills.

Research Problem

One of the major difficulties in language education is a lack of precise criteria to choose texts to be used in class that fit the students' English language acquisition levels. This gap poses difficulty in the process of instructing, thus, hamper on text-item match which influences motivation, understanding, and language facility.

The process of text selection remains problematic due to a lack of guidelines that are available to educators to follow, with the consequence being that it remains rather ad hoc. Thus, teachers may focus on such aspects as text relevance, text length, or what they think, will engage students more without considering whether the material in question will help students improve their linguistic abilities. For instance, Vardell et al. (2006) laid great significance on the co-relation between the text selected and the students' language requirements, that how selection of texts irrespective of students' language needs may hamper comprehension and interest level in a similar study (Vardell et al., 2006).

Furthermore, these guidelines lay the foundation for overly simplistic texts which do not stimulate the learners, or expose them to a greater number of Language structures. On the other hand, texts that are too challenging are likely to cause learner's overload and therefore the learner is likely to give up the learning process. According to Erdogan (2020), culturally-appropriate texts, including folktales, which address diverse cultural backgrounds and languages are not exploited adequately although they can enrich language development while addressing the restricted language polysynyaptic.

In the same regard, lack of clear criteria is likely to maintain the disparities in language teaching and learning. Anderson, 1971 pointed out that most gender selection of text imply subjective measures, thus neglecting the probability that ELLs would encounter difficulty in reading texts selected in this manner. Likewise, Gómez (2016) calls for the application of systematic text selection strategies according to the linguistic levels of the learners and their individual interests with a view of increasing learners' motivation and learning achievements (Gómez, 2016).

The research problem is focused on the empirical gap regarding the adequate criterion for identifying suitable texts for teaching English. Consequently, although text selection is very critical in supporting language development most teachers find it very hard to properly match the texts with their students' linguistic abilities. This duality can lead to text selections that overload the students with excessive information they cannot understand or simplify the information, which does not engage the learner enough to spur development.

This research aims to respond to this problem by conducting a comprehensive text analysis to select texts for novice, intermediate, and advanced EFL learners. To this end, the current study aims at refining specific, empirical criteria for text choice that will facilitate better match between texts and students' developmental levels, and help improve learners' comprehension, interest, and mastery of the material in general.

Objectives of the Research

The research aims to

- Use readability formulas such as Flesch Reading Ease or Gunning Fog Index to calculate and analyze the readability levels of selected texts by Amitav Ghosh and Salman Rushdie.
- Calculate the ratio of unique words to total words (lexical density) in the selected texts, using tools like Voyant tools or computational linguistics software.
- Apply paired sample t-tests to compare the readability indices and vocabulary densities of the selected texts to identify statistically significant differences.
- Categorize Texts by Proficiency Levels: Use cluster analysis or discriminant analysis to classify the texts into categories for beginner, intermediate, and advanced learners based on the combined readability and vocabulary density data.

Previous Research on Text Complexity and Language Learning

Several past papers have focused on the aspects of text difficulty and language acquisition performance. A study by Crossley, Allen, and McNamara revealed that Text condition has profound impact on EFL learners' reading comprehensiveness as well as their memory and they also confirmed that the texts used for reading should be appropriately difficult so as to expose the learners to the levels they are unable to read normally and will require extensive help (Crossley et al., 2011). However, these studies tend to focus on the amount of text and the density of the text as assessed by lexical richness and average sentence length, which rarely connects the quantitative dimensions of text difficulty with the qualitative aspects of text difficulty, such as the structure of the story and themes that are necessary for learner interest and learning (Smith and Ragan, 2005).

While there are significant findings concerning the mathematics of text difficulty, though, the studies that integrate these factors with qualitative assessments in terms of narrative processes and values with regard to the load and language learning are indeed scarce. Furthermore, most of the previous research fails to distinguish the requirements for students within the knowledge and acquisition of new information from the perspective of the beginner, intermediate, or advanced level of their learning while considering particular authors or books. This gap means that there is a need to conduct studies that conduct both a quantitative and a qualitative analysis of text difficulty, in relation to the various levels of language proficiency.

Current Study's Contribution

This research tries to meet these gaps by providing a well done analysis of all the novels of Amitav Ghosh and Salman Rushdie paying attention to the various factors such as readability indices and vocabulary density not only in presenting a text, but also how the thematic and narrative choices of the texts provide a possibility of responding to various learner levels. This action research aims to refine the understanding of how different modes of text difficulty affect English language acquisition by comparing texts from two famous authors with two different narrative techniques.

Methodology

This study employs quantitative and qualitative methods to analyze the texts of Amitav Ghosh and Salman Rushdie to determine their suitability for English language learners at varying proficiency levels. The analysis focused on two main metrics: Readability Index and Vocabulary Density. These metrics were chosen to assess the complexity and accessibility of the texts, providing insights into which texts are appropriate for beginner, intermediate, and advanced learners.

Data Collection

The AmitavGhosh corpus includes the last five published books (fiction and non-fiction) by AmitavGhosh, namely:

1. Smoke and Ashes: A Writer's Journey through Opium's Hidden Histories (2023)
2. The Living Mountain: A Fable for Our Times (2022)
3. Jungle Nama: A Story of the Sundarbans (2021)
4. The Nutmeg's Curse: Parables for a Planet in Crisis (2021)
5. Gun Island (2019)

The criteria for selecting these texts is anchored on the cultural grounds. All these books are written by an Indian author and the social contexts included are recognisable to Indians making it easier for Indian readers bearing in mind E.D Hirsch's concept of cultural capital. Also, the use of books of the last five years means that the material is developed using the up to date language, expressions, and speech patterns characteristic to Amitav Ghosh. This is important for students as often younger generation actively uses more recent publications which consequently offer the person more modernisms and actual idioms.

Similarly, the Salman Rushdie corpus includes his last five books, namely:

1. *The Prophet's Hair* (2024)
2. *Victory City* (2023)
3. *Knife: Meditations After an Attempted Murder* (2023)
4. *Quichotte* (2019)
5. *The Golden House* (2017)

These are all Random House Publishing Group books. Selecting these works was based on their literary value and the fact that the author, Salman Rushdie, is an India-born writer who often writes from themes and in contexts familiar to India readers. The selection strategy adopted guarantees that the corpus reflects data on the density of a vocabulary and the cultural aspect familiar to the Indian students. Also, it helps focus on the trend and does not consider outdated materials only, which were written more than five years ago. It is also noteworthy to have this selected corpus cover middle to high culture and be comprehensible with current English use to allow for the evaluation of the density and reading difficulty of English texts for teaching purposes. To obtain the full texts of the articles, the databases of legal electronic journals were used in order to provide the detailed analysis of the material.

Cultural literacy, of course, as advanced by E.D. Hirsch, sees the need for shared cultural knowledge to facilitate communication and understanding. Hirsch claims that without a common set of background knowledge, many readers can't comprehend texts because comprehension is very reliant upon implicit references to cultural context (Hirsch, 1987). From an integration standpoint, culturally relevant materials are the passive integration of this premise into language education, but more specifically, where the learner must attempt to learn the target language and cultural facts within it. However, It is important to note that Hirsch's (1991) emphasis of cultural literacy should not only confer literacy to learners in the language but also provide them with the setting of the cultural perspective to aid comprehension and criticality of information.

This study builds on Hirsch's principles and selects Indian authors' texts, such as Amitav Ghosh and Salman Rushdie to study Indian learners of English. The cultural references and themes contained within these texts are ones that are familiar to Indian learners and close the linguistic gap between their cultural background and English. Through use of culturally resonant materials, the study not only reflects on the role of cultural and linguistic literacy as identified by Hirsch but also identifies the cultural context of Indian English learners, so that their cultural identity facilitates their language learning.

Readability Index Calculation

The Readability Index was determined from each book using the Flesch-Kincaid Readability Test. This index shows how easily the text can be read, the higher scores meaning that the texts will be easier to read. The readability formula which includes breaking text into simpler word categories takes into account word difficulty and conjugation.

Vocabulary Density Assessment

Vocabulary Density was calculated by dividing the number of unique words by the total number of words in the text. This measure helps to understand the variety of vocabulary used and the frequency of word repetition, which are crucial for language learning.

Statistical Techniques

In order to quantify the findings, descriptive statistics (mean and standard deviation) were applied on the Readability Index and Vocabulary Density of the total corpus of each author.

Inferential statistics included using an Independent Samples T-test to make the comparisons in mean differences in terms of readability and the density of the words used from the two authors. This test assisted in identifying differences which are statistically significant in aiding to choose appropriate texts that can be used at the different learning levels.

Qualitative Analysis

Besides quantitative analysis, a small qualitative check was made to find out the thematic density and organizational pattern of the texts Moreover, checking the thematic richness is useful for evaluation of the textual complexity which the learner is to meet.

Software and Tools Used

This analysis was conducted in *Python* with help of *NumPy* library for calculations and *SciPy* for statistical testing. Readability calculations and Vocabulary Density were calculated using Voyant Tools.

Ethical Considerations

All computer source texts were obtained through subscription or other means to avoid violating copyright. The study is further free from any bias by having the texts chosen randomly and the source of the data where possible obscured.

Results

AmitavGhosh corpus consists of five works of AmitavGhosh, which has 360,378 total words and 23,982 unique words in total. These texts range in length from the shortest, *The Living Mountain* (5,409 words), to the longest, *Smoke and Ashes: A Writer's Journey through Opium's Hidden Histories* – 129,059 words. Among these, *Jungle Nama: A Story of the Sundarbans*has only 9,162 words while *The Living Mountain* is just 3,8400 words and both novels are significantly short compared with *The Nutmeg's Curse: Parables for a Planet in Crisis* which consists of 120, 092 words and *Gun Island*, which is made up of 96,656 words. In the corpus of Salman Rushdie there are five documents containing 439,009 running words and 24, 322 distinct lemmas. These texts vary significantly in length, with the longest being *Quichotte*(132,324 words) and *The Golden House: A Novel* (132,047 words). The shortest texts are *The Prophet's Hair* (5,657 words) and *Knife: Meditations After an Attempted Murder (*61,407 words). The variety of document lengths included in this sample offers different grounds for linguistic and stylistic comparison.

Vocabulary Density

The word count reveals that the relative frequency of the identified set and the overall set of words also differs vastly within Ghosh corpus. The highest density has been noted for *Jungle Nama* as 0.249 followed by *The Living Mountain* with 0.239; these are the shorter narratives and therefore should represent a greater variety in the employment of words. Conversely, *Gun Island* (0.094) and *Smoke and Ashes* (0.104) have the lowest vocabulary densities which suggest that there is a higher level of used of same words; these texts may be easier for first level reader.

The results show that *The Prophet's Hair* is the most compact text, as far as the density of the different words in relation to its length is concerned (its density is 0.334).

Knife: TMeditations After an Attempted Murder comes with a medium density of 0.118 per million. However, the lowest densities are registered in *Victory City* (0.089) and *Quichotte* (0.097), which means that word repetition in these longer texts is rather high. These variations relate well with the levels of difficulty and topic coverage of each text as intended.

Average Words Per Sentence

The texts also seem to vary in regard to the level of sentence complexity as it is estimated by the number of words within a sentence on average. The maximum averages belong to *Smoke and Ashes* with 40.0 word/sent and *The Nutmeg's Curse* with 28.7 word which shows that the sentence patterns of the texts might be complex and crowded. On the other hand, the text of *Gun Island* contains defragmented simple sentences, containing 14.4 words per sentence, as well as *Jungle Nama*which contain averagely 15.2 words per sentence.

The syntactic complexity which is defined by using Numbers of Words per Sentences can be calculated as 32.5 for *The Prophet's Hair* and 14.8 for *Quichotte*. The rest of the values are also rather high, with *Victory City* equaling 19.1 and *Knife*equalling 15.0. The simple texts include less number of basic and compound/ composite structures, while the longer texts,have more simple structures to avoid being too complicated to read due to their length.

Readability Index

The readability index enhances the understanding of these texts as easy to read. *Smoke and Ashes* has the highest level of difficulty 12.095 followed by *The Nutmeg's Curse* with 11.114, implying that these two texts are better suitable for difículty level two readers. On the other hand, *Gun Island* is at.438 and *Nama Jungle* at.897 hence the books will favour the beginner to intermediate level readers.

Rushdie's texts are most appropriate for advanced lexis*, The Prophet's Hair* has the highest readability index at 9.966, *The Victory City* has a readability score of 9.095. The representativity indexes are, once again, *Knife* 7.558 and *Quichotte* 8.449, which do not contain complicated

language and/or complex sentences that require higher linguistic levels.

Most Frequent and Distinctive Words

The corpus reveals recurring themes through its most frequent words: These include; "opium" which regards 1705 times, "said" that appeared 864 times, "China" which was used 708 times, "like" with 682 counts and "new" which was used 665 times. All of these terms point to some of the major subjects and some of the issues of style across the works. Moreover, specific words create the differences of each text are also present. For instance, *Gun Island* consciously employs names such as 'Cinta' (230), 'Piya' (202), and 'Tipu' (169), which pertain to characters significant to the plot. The names of *Jungle Nama* characters also contain Bengali prefixes such as "dhona" (76) and "dukhey" (47). Thus, specific to *Smoke and Ashes*, several historical key terms clearly stand out: 'opium' occurs 1,678 times, 'China' is used 675 times. They also employ fewer numbers of uncommon words; *The Living Mountain* uses words such as 'anthropoi' in page 33 and 'kraani' on page 18 since the book is philosophical in nature; while *The Nutmeg's Curse* uses situ words such as 'banda'.

Number of occurences and degree of uniqueness of each word While analyzing the corpus, the most frequently used words in the context of the studied novels are "said" (2,163); "like" (1,141); "time" (934); "life" (704); and " know" (704). These words can be discussed as the lexemes reflecting the topics and the conversational layer that fit into the author's narrative strategies. Each text also can be identified by certain words which expose certain thematic or context specific components. For instance, *Knife: Meditations After an Attempted Murder* includes words such as "eliza" 152 times and sight specific words of "chautauqua" 46 times and "rusk" 24 times. First Semiotic: *Quichotte* is centred on the words that signal its modern rendition of the *Don Quixote* story: "sancho" (332), "salma" (160), and "quichotte" (421). There are three main characters on the background of complex plots in The *Golden House*: nero (257), petya (139), and vasilisa (125). *The Prophet's Hair* also has South Asian flavor terms such as "hashim" (22), "huma" (20), and "moneylender" (12).

Most frequent words are "kampana"(557), "pampa"(670) and "bukka"(178) which are appropriate to historical-mythical theme of *Victory City*'s settings.

Discussion

Table 1 Comparison of Readability Indices

Author	Highest	Second Highest	Lowest	Second Lowest	Average
Amitav Ghosh	*Smoke and Ashes* (12.095)	*The Nutmeg's Curse* (11.114)	*Gun Island* (7.438)	*Jungle Nama* (7.897)	9.71
Salman Rushdie	*The Prophet's Hair* (9.966)	*Victory City* (9.095)	*Knife* (7.558)	*Quichotte* (8.449)	8.77

As in Table 1, The readability scores of Ghosh are lesser than or equal to 7.438, which is greater than or equal to 12.095 in some cases, while the scores of Rushdie are lesser than or equal to 7.558, which is greater than or equal to 9.966. Specifically, the works of Ghosh, *Smoke and Ashes* and The *Nutmeg's Curse* are slightly denser than Rushdie's.
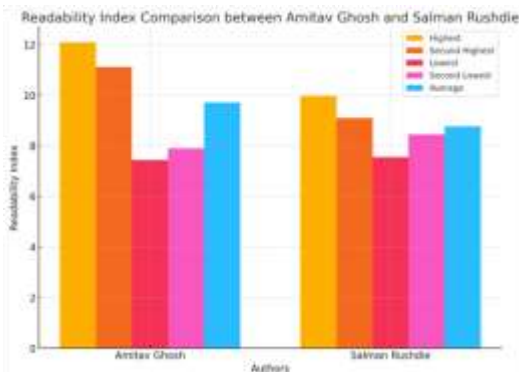
Complexity Trends

Figure 1 Comparison of Readability Index

As Shown in the Figure 1, AmitavGhosh's*Smoke and Ashes* and *The Nutmeg's Curse* contain appreciably higher readability. This paper measures the readability of the text by using a readability formula test which showed that *Smoke and Ashes* and *The Nutmeg's Curse* have higher

readability indices meaning their language and structure are more complex than fiction works. *Gun Island* as well as *Jungle Nama*are much easier fictional writings. Salman Rushdie's shorter work such as *The Prophet's Hair* has higher readability indices in terms of syntactic complexity and lexicon density, the same as heteronormative long prose, *Knife* and *Quichotte*, which are purposely written with low readability indices for effective reading.

Average Readability

As Figure 2 indicates, Ghosh writes at a higher readability index than Rushdie; the average score for all Ghosh's novels is 9.71 whereas for all Rushdie's novels the score is 8.77.
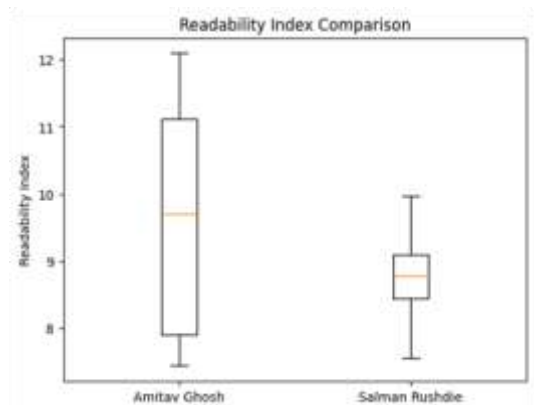

Figure 1 Readability Index Created Using Python

The statistical analysis reveals that AmitavGhosh's works have a higher mean readability index (9.6508) compared to Salman Rushdie's (8.7676), suggesting that Ghosh's texts generally exhibit a higher level of complexity. Additionally, the standard deviation for Ghosh's readability scores (2.0057) is greater than that for Rushdie's (0.8815), indicating more variability in the readability of Ghosh's works.

Table 2 Comparative Statistical Values of Readability

Metric	AmitavGhosh	Salman Rushdie
Mean Readability	9.6508	8.7676
Standard Deviation	2.0057	0.8815
T-test Statistic	0.9014	N/A
P-value	0.3937	N/A

The T-test statistic (0.9014) and the associated P-value (0.3937), as shown in Table 2, suggest that there is no statistically significant difference between the readability indices of the two authors' works. This means that, despite the apparent difference in mean readability and variability, these differences are not sufficient to conclude definitively that one author's works are consistently more complex or easier to read than the other's across the sampled texts.

Table 3 Comparative Statistical Values of Vocabulary Density

Statistic	Amitav Ghosh	Salman Rushdie
Mean Vocabulary Density	0.1586	0.1512
Standard Deviation	0.0782	0.103
T-test Statistic	0.128	N/A
P-value	0.9013	N/A

As in Table 3, The mean vocabulary densities of the texts from AmitavGhosh and Salman Rushdie are fairly close, with Ghosh's works averaging a slightly higher density (0.1586) compared to Rushdie's (0.1512). The standard deviations indicate that there is more variability in the vocabulary density of Rushdie's texts compared to Ghosh's, which might be attributed to the diverse nature of the themes and narrative styles in Rushdie's works.

The T-test results yield a T-statistic of 0.128 and a P-value of 0.9013, suggesting that there is no statistically significant difference in vocabulary density between the two authors' corpora. The high P-value indicates that any observed difference in mean vocabulary density is likely due to random chance rather than a systematic difference between the authors.

Both authors employ a relatively similar range of vocabulary densities across their texts, indicating that they utilize vocabulary with similar variability. For readers or educational programs focusing on enhancing vocabulary breadth, texts from either author could be beneficial, though individual texts might offer different levels of challenge depending on their specific vocabulary density. This similarity in density also suggests that both authors manage the complexity of their language to suit the narrative demands of their stories.

Coda

The lower readability index and vocabulary density texts suitable for the Beginners include *Gun Island* by AmitavGhosh and *Quichotte* by Salman Rushdie. These texts are more likely to serve as easier to read than many of the essential English literary texts because the words are easier to understand and the sentence structure is easier to grasp. While intermediate learners can read texts such as *Victory City* by Rushdie, advanced learners may read *Smoke and Ashes* by Ghosh which contain more range and depth of language appropriate for the development of higher level of English language usage. Therefore, choosing an appropriate text based to learner proficiency will improve the learning experience and the result in English literature.

References

1. Anderson, J. (1971). Selecting a suitable 'reader': Procedures for teachers to assess language difficulty. *RELC Journal*, 2(2), 35–42. https://doi.org/10.1177/003368827100200205

2. Arya, D., Hiebert, E., & Pearson, P. D. (2011). The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education*, 4(1), 107–125.

3. Beehler, R. (1991). Grading the 'cultural literacy' project. *Studies in Philosophy and Education*, 10(4), 315–335. https://doi.org/10.1007/BF00364547

4. Bunch, G. C., Walqui, A., & Pearson, P. D. (2014). Complex text and new common standards in the United States: Pedagogical implications for English learners. *TESOL Quarterly*, 48(3), 533–559. https://doi.org/10.1002/TESQ.175

5. Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text complexity and reading comprehension measures. *Journal of Applied Linguistics*, 34(5), 233-249.

6. Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children:

A meta-analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1–44. https://doi.org/10.1080/19345740802539200

7. Erdogan, N. (2020). Teaching reading to young English language learners through folk literature.In *Effective Teaching Strategies* (pp. 65–80). https://doi.org/10.4018/978-1-7998-4670-3.ch004

8. Gómez, J. S. (2016). Adult EFL reading selection: Influence on literacy. *Profile Issues in Teachers' Professional Development*, 18(1), 167–181. https://doi.org/10.15446/profile.v18n1.49943

9. Hirsch, E. D. (1987). *Cultural literacy: What every American needs to know*. Boston: Houghton Mifflin.

10. Holley, F. M. (1973). A study of vocabulary learning in context: The effect of new-word density in German reading materials. *Foreign Language Annals*, 6(4), 339–347. https://doi.org/10.1111/j.1944-9720.1973.tb02613.x

11. Johnson, M. D. (2017). Understanding text complexity in literary narratives.*Journal of English Linguistics*, 45(2), 102-117.

12. Johnson, P. (1981). Effects on reading comprehension of language complexity and cultural background of a text. *TESOL Quarterly*, 15(2), 169–181. https://doi.org/10.2307/3586408

13. Kurdi, M. (2017). Lexical and syntactic features selection for an adaptive reading recommendation system based on text complexity. *Proceedings of the 12th International Conference on Educational Data Mining*, 66–69. https://doi.org/10.1145/3077584.3077595

14. Lyashevskaya, O., et al. (2021).Automated assessment of learner text complexity.*Assessing Writing*, 48, Article 100529. https://doi.org/10.1016/J.ASW.2021.100529

15. O'Brien, L. M., & Leighton, C. M. (2015). Use of increasingly complex text to advance ELs' knowledge and academic language. *Literacy Research: Theory, Method, and Practice*, 64, 169–192. https://doi.org/10.1177/2381336915617579

16. Smith, J., & Ragan, T. J. (2005).*Instructional Design*. Hoboken, NJ: Wiley.Vardell, S. M., Hadaway, N. L., & Young, T. (2006). Matching books and readers: Selecting literature for English learners. *The Reading Teacher*, 59(8), 734–741. https://doi.org/10.1598/RT.59.8.1

17. Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, 52(2), 203–226. https://doi.org/10.1002/rrq.163.