https://doi.org/10.37815/rte.v38n1.1386
Original articles

Analysis of Readability and Lexical Complexity in Thesis Abstracts from Early Childhood and Basic Education

Análisis de la legibilidad y la complejidad léxica en resúmenes de tesis de educación inicial y educación básica

Andrea Ullaguari Piuri[1] https://orcid.org/0009-0007-4118-5266,
Rodrigo Tovar Viera[1] https://orcid.org/0000-0002-5960-3020

[1]*Universidad Técnica de Cotopaxi*, Latacunga, Ecuador
andrea.ullaguari9829@utc.edu.ec, rodrigo.tovar@utc.edu.ec

Sent: 2025/09/10
Accepted: 2025/10/17
Published: 2026/06/15

Abstract

The readability of academic texts is a key factor in ensuring reading comprehension, especially in educational contexts where textual clarity directly impacts learning. The main objective of this study is to identify levels of textual complexity and differences in discursive practices between two disciplines. To this end, a mixed-methods, descriptive design was adopted, combining quantitative and qualitative methods. The readability of English-language abstracts was assessed using traditional formulas, including Flesch Reading Ease, Flesch-Kincaid Grade Level, and Gunning Fog Index. Computational tools such as Text Inspector and VocabProfile were also used to analyse lexical diversity and discursive variables. The results revealed a high level of linguistic complexity in both groups of texts. The early childhood education abstracts were more difficult to read, with extreme scores across all the formulas applied. In contrast, the basic education texts showed greater lexical richness and sophistication. The analysis of discourse markers revealed a predominance of interactive resources, especially transitions, while interactional resources were scarce, evidencing an objective and impersonal academic style. In conclusion, writing patterns that hinder comprehension were identified, particularly in early childhood education. This study highlights the need to strengthen academic writing skills to improve the clarity, coherence, and accessibility of university texts.

Keywords: Academic skills, academic text, linguistic analysis, metadiscursive markers, reading comprehension.

Summary: Introduction, Materials and Methods, Results, Discussion, Conclusions.

How to cite: Ullaguari, A. & Tovar, R. (2026). Analysis of Readability and Lexical Complexity in Thesis Abstracts from Early Childhood and Basic Education. *Revista Tecnológica - Espol, 38(1), 123-140.* https://rte.espol.edu.ec/index.php/tecnologica/article/view/1386

Resumen

La legibilidad de los textos académicos es un factor clave para garantizar la comprensión lectora, especialmente en contextos educativos donde la claridad textual tiene un impacto directo en el aprendizaje. El objetivo principal de este artículo es identificar los niveles de complejidad textual y las diferencias en las prácticas discursivas entre dos disciplinas. Para ello, se adoptó un enfoque mixto con diseño descriptivo, combinando métodos cuantitativos y cualitativos. La legibilidad de los resúmenes escritos en inglés se examinó mediante fórmulas tradicionales, como de Flesch Reading Ease, Flesch-Kincaid Grade Level y Gunning Fog Index. Asimismo, se emplearon herramientas computacionales como Text Inspector y VocabProfile, que permitieron analizar la diversidad léxica y variables discursivas. Los resultados revelaron una elevada complejidad lingüística en ambos grupos de textos. Los resúmenes de Educación Inicial presentaron mayor dificultad de lectura, con índices extremos en todas las fórmulas aplicadas; por el contrario, los textos de Educación Básica mostraron mayor riqueza léxica y sofisticación. El análisis de los marcadores discursivos reveló un predominio de recursos interactivos, especialmente transiciones, mientras que los interaccionales fueron escasos, lo que evidenció un estilo académico objetivo e impersonal. En conclusión, se identifican patrones de escritura que dificultan la comprensión, siendo más pronunciados en Educación Inicial. Este estudio resalta la necesidad de fortalecer las competencias de escritura académica para mejorar la claridad, la coherencia y la accesibilidad de los textos universitarios.

Palabras clave: Análisis lingüístico, comprensión lectora, competencias académicas, marcadores meta-discursivos, textos académicos.

Introduction

Readability refers to the ease with which academic and non-academic texts can be processed and understood by readers, as it directly influences comprehension and information retention. According to Zainurrahman et al. (2024), readability pertains to the ease with which a reader can assimilate information from a text. Research in this field has shown that readability depends not only on linguistic aspects such as syntax and vocabulary but also on reader-related factors such as educational level and language skills. In this analysis, readability tools such as Coh-Metrix, along with traditional formulas, have been widely applied across various fields and disciplines. These include Flesch Reading Ease (FRE), which provides a numerical score of reading ease; the Flesch-Kincaid Grade Level (FKGL), which estimates the educational level required for comprehension; and the Gunning Fog Index (GFI), which assesses textual complexity based on word and sentence length. Together, these measurable formulas, in conjunction with intuitive methods (Alley, 2009), offer complementary and concrete perspectives for evaluating readability in written texts (Young, 1999). Researchers such as Fry (1968), Ash and Edgell (1975) played key roles in establishing readability indices. Over time, these tools became widely adopted, leading to the development of multiple formulas to assess and compare text readability, particularly in academic contexts (Crossley et al., 2014; McNamara et al., 2013; Li et al., 2025). Li et al. (2025) highlight the importance of adapting texts to the cognitive and linguistic abilities of the target audience to facilitate comprehension.

In recent years, readability has become a central topic in academic and applied linguistics due to the growing availability of digital information. Deutsch et al. (2020) argue that conventional academic writing styles often limit accessibility and exclude broader audiences. Complementing this view, Lee and Vajjala (2022) point out that readability is not determined solely by-word frequency or sentence length but also by coherence and fluency. Similarly, Alkaldi and Inkpen (2023) highlight how textual structure and lexical choices strongly shape the

accessibility of complex texts for readers with varied literacy levels. Advances in computational linguistics have further expanded the field. For instance, Meng et al. (2021) introduced *ReadNet*, a tool that integrates structural and semantic features to assess complexity, while Text Inspector evaluates lexical and linguistic dimensions to tailor texts to specific audiences. Cai et al. (2022) add that visual design, including typography, colour, and adaptive interfaces, enhances user experience by allowing customization that facilitates comprehension.

On the other hand, lexical richness is a measure of how varied, dense, and elaborate the vocabulary in a text is, indicating the author's ability to use diverse and complex words in written compositions. Lexical richness is a characteristic of academic language that combines the density, variety, and sophistication of the vocabulary used in a text to reflect its complexity and expressive precision (Tovar, 2022). In other words, lexical richness is understood as a construct that encompasses various dimensions of vocabulary, including both the variety and level of complexity that language learners are able to use in their productions (Zhang, 2020).

International research has increasingly applied computational and interdisciplinary approaches to the study of readability. Deutsch et al. (2020) combined linguistic features with deep learning to classify text difficulty, whereas Imperial (2021) integrated BERT with traditional indicators to improve assessments in English and Filipino. Beier et al. (2022) emphasize that both linguistic and visual elements shape reading fluency, while Vajjala (2021) calls for approaches that merge educational and computational perspectives. In the Latin American region, studies point to the pedagogical significance of readability. For instance, Stone et al. (2019) demonstrate that phonemic awareness interventions enhance decoding and comprehension in early education, and Barberán et al. (2024) report moderate links between comprehension and cognitive development among middle school students. At the higher education level, Mora et al. (2021) observed that university students achieve strong comprehension but struggle with slow reading speed, and Guerrero et al. (2025) show that virtual tutoring initiatives improve comprehension outcomes in basic education. Within this context, research on readability remains relatively scarce in Ecuador compared to international and regional studies.

The research works discussed above address the relevance of measuring text readability to improve educational quality (e.g., Barberán et al., 2024; Mora et al., 2021; Guerrero et al., 2025). Nevertheless, most existing work focuses on textbooks or general reading practices, leaving aside academic writing genres produced by students, particularly thesis abstracts in the fields of early childhood education and basic education. Although readability has been examined across educational materials (Hu, 2023), medicine (Long, 2022), finance (Besuglov & Crasselt, 2021; Yu, 2022), law (Arbel, 2024; Han et al., 2024), and tourism (Lazzeretti, 2021; Tiago et al., 2021), there is still little research on how Ecuadorian University students construct academic texts such as thesis abstracts. This gap justifies the need for local research that examines readability alongside deeper linguistic dimensions such as lexical richness and discourse markers. Hyland's (2015) model distinguishes between interactive markers, which organize the text for the reader, and interactional markers, which allow the writer to project a stance and engage with the audience. These markers, together with measures of lexical density, variation, and sophistication, are central to understanding how thesis abstracts are structured and how accessible they are to readers.

Accordingly, the present study aims to analyse readability, lexical richness, and discourse marker use in thesis abstracts from early childhood and basic education programs by identifying disciplinary differences in writing practices and their impact on accessibility. To guide this inquiry, the following research questions are posed:

What differences exist in the levels of readability, lexical richness, and use of discourse markers between thesis abstracts in early childhood education and basic education?

How might these differences influence the accessibility and comprehension of academic texts?.

Materials and Methods

This research adopts a quantitative approach with a descriptive design, supported by qualitative analyses. It combines quantitative and qualitative methods to analyse the readability and use of discourse markers in 60 thesis abstracts in early childhood education and basic education, submitted between 2019 and 2025. Thirty thesis abstracts written in English from their original Spanish versions in early childhood education and thirty in basic education were included in the analysis. Since the research examines the text as a finished product rather than the process of its creation, focusing on the quality of the written output rather than on the writer or their writing practices, a parallel analysis was not conducted for the Spanish version. However, when comprehension difficulties or inconsistencies were identified in the English text, the original Spanish version was consulted to clarify any doubts. This allowed the researchers to verify whether errors in the Spanish wording affected English writing quality. Following Noguera (2013), abstracts translated from Spanish into English often exhibit coherence and rhetorical issues stemming from literal translation or insufficient command of academic English. Consequently, it was expected that poor wording in the source language would likely result in similarly deficient quality in the translated version.

Corpus Collection

The thesis abstracts considered in this study were those written in English, taken from the digital repository of the Technical University of Cotopaxi. This institution offers bachelor's degrees in early childhood education and basic education. Only abstracts published between 2019 and 2025 were included, with a length of 200-250 words, and whose topics were related to language teaching or the educational field. Topics covered issues related to learning difficulties at school, recreational games, motor skills, nutrition, and the role of the teacher in the educational process. The rationale for choosing these mini texts lies in the fact that they are not purely narrative in nature; therefore, they are expected to display broader lexical and grammatical variation to convey information effectively (Mahadini et al., 2021). To ensure consistency and quality of the corpus, abstracts with incomplete content, serious formatting problems, or unrelated topics were excluded. The final data set, comprising 14,793 words, served as the basis for the analysis of readability, lexical richness, and elements of meta-discourse across different academic levels. The table below describes the corpus.

Table 1

Word count by discipline

Discipline	***Genre***	***Number of texts***	***Tokens***
Early childhood	*Academic text*	*30*	*7,666*
Basic education	*Academic text*	*30*	*7,127*
Total		30	14,793

Corpus Processing

Prior to analysis, the texts were cleaned, removing punctuation marks, numerical digits (converted to words), question marks, and other non-lexical elements to generate a continuous, uniform text for processing. After the cleaning process, the entire corpus was submitted to text analysis software. At this stage, the automatic tools Text Inspector (https://textinspector.com/) and VocabProfile (https://www.lextutor.ca/vp/eng/) gauged the lexical metrics and complexity of the texts. In particular, the classical readability formulas assessed text difficulty as follows:

(1) FRE evaluated reading ease on a scale from 0 to 100, the higher the score, the greater the clarity.
(2) FKGL illustrated the minimum educational level required to understand the text.
(3) GFI calculated the number of years of study required to understand the text, treating words with three or more syllables as complex.

These three formulas provide quantitative measures that estimate how easy or difficult a text is to comprehend, primarily based on sentence length and lexical complexity. Table 2 summarizes the difficulty levels associated with the FRE, FKGL, and GFI indices to facilitate interpretation of the results.

Table 2
Readability indices of traditional readability formulas

FRE	*FKGL*	*GFI*	*Level of Difficulty*
90 – 100	*4*	*5*	*Very easy*
80 – 89	*5*	*6*	*Easy*
70 – 79	*6*	*7*	*Fairly easy*
60 – 69	*7 - 8*	*9*	*Standard English*
50 – 59	*Some high school*	*12*	*Fairly difficult*
30 – 49	*High school to college*	*14 – 16*	*Difficult*
0 – 29	*College & up*	*>16*	*Very difficult*

Discourse Markers

Discourse markers —the linguistic devices that guide the reader throughout the text by signalling logical, organizational, and interactive relationships between ideas —are fundamental for enhancing reading comprehension. Following Tarzeen and Tahir's (2024) study, the current research highlights their importance in achieving coherence and cohesion in academic writing, as they organize information and clarify relationships between ideas and expressions within texts. Accordingly, adopting Hyland's (2015, p. 3) classification model, the analysis focused on the frequent use of interactive and interactional categories of discourse markers, as shown below.

Table 3

Classification of discourse markers

Category	Function	Example
Interactive	*Guide readers through the text*	
Code glosses	Clarify or rephrase for clarity	namely, that is
Endophorics	Refer to another part of the text	see section II
Evidentials	Cite information from other texts	according to X
Transitions	Link main clauses	therefore, however
Interactional	*Involve the reader in the text*	
Attitude	Express the author's attitude	regrettably, surprisingly
Hedges	Show limited author certainty.	possibly, could be
Boosters	Show strong author certainty	definitely, undoubtedly
Engagements	Engage the reader directly	consider, note that
Self-mentions	Author references in the text	I, we

Ethical Considerations

Following the ethical principles proposed by Kirk (2007) for educational research, rigorous measures were taken to protect the authors' identities in the analysed abstracts. Neither the authors' names nor any institutional or personal information that could link the texts to their creators was considered, as the entire procedure was carried out anonymously. This decision addressed the need to avoid potential conflicts of interest and to ensure that subjective judgments about the origin or authorship of the content would not influence the analysis results. In addition, anonymity made it possible to focus exclusively on the linguistic and discursive characteristics of the texts, thereby promoting a more objective, transparent, and respectful approach to the principles of confidentiality, integrity, and academic credibility throughout the study.

Results

Text Readability

Table 4 presents the average scores of three readability formulas applied to the abstracts in early childhood education and basic education. Although the mean values vary between the two groups, both indicate that the texts are difficult or very difficult to read, reflecting a high level of textual complexity in both cases.

Table 4

Average readability scores of thesis abstracts within disciplines

	Readability formulas			
Disciplines	**FRE**	**FKGL**	**GFI**	Range
Early childhood	-217.89	109.25	115.10	Very difficult
Basic education	–193.26	99.24	104.71	Difficult

Lexical Richness

Table 5 shows that both groups have identical lexical density (0.56), indicating no difference and suggesting that the proportion of content words is stable across the two disciplines. This means that text complexity, in terms of content words, does not increase with educational level. Regarding lexical variation, texts from basic education show slightly higher variation; they use a somewhat wider range of different words compared to early childhood texts. Although the difference is minor (+0.02), it suggests a gradual shift toward a more diverse vocabulary as students' progress. Similarly, basic education abstracts use a higher proportion of less frequent or more advanced words (+1.46), indicating an increase in lexical sophistication. This suggests that as education levels rise, texts tend to incorporate more complex and specialized vocabulary. While lexical density remains constant across the two groups, lexical variation and sophistication increase slightly in basic education, reflecting a modest rise in vocabulary diversity and complexity. However, the differences are relatively small, suggesting a gradual rather than abrupt increase in linguistic demands.

Table 5

Metrics of lexical richness in both disciplines

Metrics	**Early childhood**	**Basic education**	**Difference**	**P-value**
Lexical density	0.56	0.56	0.00	0.037*
Lexical variation	0.16	0.18	+0.02	***0.689***
Lexical sophistication	27.04	28.50	+1.46	**0.334**

Although the t-test did not reach significance across dimensions ($t = -1.480$, $p = 0.074 > 0.05$), further analysis using ANOVA reported statistical differences in lexical density ($F = 1.94$, $p = 0.037 < 0.05$), between the two disciplines. In contrast, both the t-test ($t = -1.480$, $p = 0.074 > 0.05$) and ANOVA ($F = 0.83$, $p = 0.689$) showed no significant differences for lexical variation, indicating that the range of vocabulary used in abstracts was comparable across groups. Similarly, lexical sophistication did not differ significantly, as shown by the t-test ($t = -0.736$, $p = 0.233 > 0.05$) and ANOVA ($F = 1.17$, $p = 0.334$), suggesting that the use of advanced or less frequent words remained consistent between the two disciplines.

Discourse Markers

Discourse markers are key resources for organizing and orienting academic discourse. Table 6 shows that the frequency of discourse markers varies between interactive and interactional types (e.g., hedges, transitions, attitude markers). In general, both groups show similar patterns of use, although with certain differences in the frequency of specific elements.

Table 6

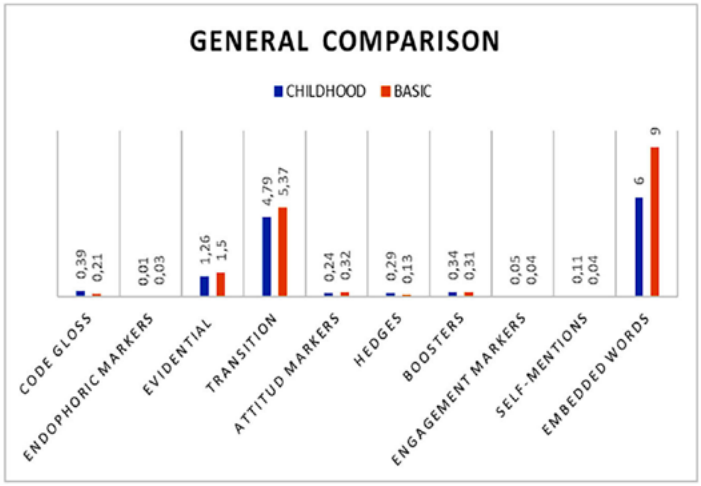
Frequency of discourse markers across disciplines

Discourse markers	Early childhood	Basic education	Difference	P-value
Interactive resources				
Code glosses	0.39	0.21	-0.18	
Endophorics	0.01	0.03	+0.02	*0.385*
Evidentials	1.26	1.50	+0.24	
Transitions	4.78	5.36	+0.58	
Interactional resources				
Attitude markers	0.24	0.32	+0.08	
Hedges	0.29	0.13	-0.16	
Boosters	0.34	0.31	-0.03	*0.387*
Engagements	0.05	0.04	-0.01	
Self-mentions	0.11	0.04	-0.07	

Although basic education shows slightly higher use of some markers (e.g., evidentials, transitions, and attitude markers), these differences in interactive ($p=0.385 > 0.05$) and interactional ($p=0.387 > 0.05$) discourse markers between early childhood and basic education texts are not statistically significant. The results suggest that both early childhood and basic education texts rely on similar patterns, with no clear statistical evidence of increasing complexity in their interactive or interactional functions. Transition markers emerged as the most frequent in both groups (5.36 in basic education and 4.78 in early childhood education), indicating a consistent reliance on logical connectors to structure discourse. These are followed by evidential markers (1.50 in basic education and 1.26 in early childhood education), which highlight the use of references to sources or prior knowledge. In contrast, boosters (0.34 vs. 0.31) and hedges (0.29 vs. 0.13) occur less often. Attitude markers also appear infrequently (0.32 vs. 0.24), as do code glosses (0.39 vs. 0.21) and self-mentions (0.11 vs. 0.04). Finally, engagement markers (0.05 vs. 0.04) and endophorics (0.01 vs. 0.03) represent the least frequent categories. Figure 1 illustrates the frequency of interactive and interactional discourse markers in the two disciplines.

Figure1

Frequency of discourse markers



Embedded words

Text complexity was also analysed using the embedded words index, which measures the number of words before the main verb in representative sentences. According to Barrot and Agdeppa (2021), this index is relevant for assessing syntactic complexity and its effect on reading comprehension. On average, the early childhood education texts had six embedded words, while those for basic education had nine, reflecting greater structural complexity in the latter. However, despite this difference, readability indices showed that early childhood education texts were more difficult to read, suggesting that other factors, such as specialized vocabulary or sentence length, also influence reading difficulty.

Discussion

In response to the first research question regarding differences in readability, lexical richness, and the use of discourse markers between thesis abstracts in early childhood education and basic education, the results indicate the following:

From a readability perspective, the abstracts reflect a high level of complexity, not only due to lexical load but also because of the absence of structural simplification strategies. Unlike Xu and Casal (2023), who observed that law students improved clarity by increasing lexical sophistication and simplifying syntax, this study found no evidence of such structural adjustments. This lack of syntactic refinement may contribute to the recorded reading difficulty, emphasizing the need for instruction that balances lexical richness and syntactic clarity to improve readability and comprehension.

Regarding the use of discourse markers, differences were observed in the use of transitional, evidential, clarifying, and hedging markers that contribute to coherence and precision in the transmission of ideas, which are fundamental for clarity in academic writing (Hyland, 2015). From a pragmatic-discursive perspective, interactive resources predominate, especially transitional markers (4.78 in early childhood and 5.36 in basic education) and evidentials (1.26 and 1.50), reflecting an effort to organize the text and reference external sources. In contrast, interactional resources such as attitude markers, hedges, boosters, engagement, and self-mentions are minimal or nearly absent, configuring an objective, impersonal, and distant style, as noted by Pasaribu et al. (2022). Therefore, the use of elaborative and inferential markers is vital not only to improve coherence and overall readability (Haninda & Bram, 2021), but also serves cohesive functions, reinforces authorship and prevents plagiarism by clearly attributing ideas (Ondondo & Rew, 2020). Nonetheless, in the current corpus, the use of such markers is limited, possibly related to academic level (undergraduate vs. graduate) and training in academic writing, which reinforces the hypothesis of depersonalized writing influenced by insufficient instruction in interactional metadiscourse (Abdullah, 2022).

Concerning lexical richness, although the proportion of lexical density is identical in both groups (0.56), there are differences at the significance level α=0.05. Statistically, lexical variation is slightly higher in basic education (0.18 vs. 0.16), as is sophistication (28.50 vs. 27.04); nonetheless, significance is scarce (see Table 5). These statistical differences, although small, reflect a considerable semantic load that contributes to the overall difficulty. This result contrasts with Müngen's (2023) findings, which identified notable disciplinary differences in lexical richness shaped by communicative requirements. While Müngen (2023) highlights how interdisciplinary exchanges often foster lexical adaptations, the present study reveals a

contrasting tendency toward stylistic uniformity, with little evidence of functional variation (see Bonsall et al., 2020; Kukafka & Arguello, 2021). The finding suggests that contextual demands did not translate into differentiated lexical use. This tendency could be attributed to differences in language proficiency and similar L1 linguistic backgrounds in producing the target language. Although lexical density emerged as the main distinguishing feature in the present study, likely shaped by institutional or disciplinary conventions, no substantial gains in lexical variation or sophistication were identified, in contrast to the improvements reported by Nasrollahi et al. (2022). One possible explanation for this divergence may stem from the participants' relatively homogenous linguistic profiles. Most shared similar first-language backgrounds and limited exposure to varied communicative contexts, which could restrict opportunities to broaden their lexical repertoires.

In relation to the second research question —the potential effect of these differences on accessibility and comprehension —the results revealed high textual complexity and low readability. These findings suggest that both groups produce texts that are not easily accessible to readers, including academic audiences. This pattern is comparable to Li et al. (2025), who note that syntactic complexity often increases at the expense of readability in academic materials. Regarding interactive resources, the most significant increase appears in transitions (+0.58) and evidentials (+0.24), suggesting that basic education texts may rely more on explicit logical connections and clearer source attribution. Code glosses decline (-0.18), indicating fewer explanatory reformulations at higher levels. Overall, differences are minor and not statistically significant ($p = 0.385$). Concerning interactional resources, attitude markers show a slight increase (+0.08), indicating a somewhat stronger evaluative stance in basic education texts. Hedges (-0.16) and self-mentions (-0.07) decrease, indicating less tentativeness and reduced author visibility in later texts. This tendency is consistent with Wang and Zeng's (2021) findings, which show that novice writers use more hedges and self-mentions, whereas advanced texts typically reduce these features in favour of a more assertive stance. Boosters and engagements remain relatively stable with minimal variation, which can serve as an indicator of conciseness in academic discourse (Al-Harthi et al., 2022). The differences are minor and not statistically significant ($p = 0.387$). Taken together, basic education texts exhibit slight increases in transitions, evidentials, and attitude markers, suggesting a move toward greater logical structuring, clearer source attribution, and a more evaluative stance. However, decreases in hedges and self-mentions indicate less cautious, less personal writing (Li et al., 2025; Wang & Zeng, 2021).

Put another way, the combination of high lexical density and lack of syntactic simplification may cause cognitive overload, hindering effective comprehension, especially for non-specialist readers. Moreover, the scarcity of discourse markers that promote reader interaction, such as self-mentions or engagement markers, can reduce clarity and communicative effectiveness, making the texts appear more distant and less accessible (Pasaribu et al., 2022). These results indicate that, although lexical richness is important, it does not guarantee comprehension. Therefore, instruction should focus on strategies that integrate lexical richness, syntactic clarity, and the effective use of discourse markers to improve accessibility in academic texts.

Comparative Interpretation

In general, the results obtained from the readability indices, the measures of lexical richness, and the analysis of discourse markers indicate consistent patterns. These findings allow us to affirm that both groups of abstracts —those in early childhood education and those

in basic education —present high linguistic complexity and difficulty in reading. However, the texts produced in the early childhood education program tend to be more difficult to process overall. This tendency is mainly supported by the readability indices, which reveal that the early childhood education abstracts exhibit more extreme values (FRE: -217.89; FKGL: 109.25; GFI: 115.10). These findings indicate dense syntactic structures characterized by long sentences and a high level of complexity in the lexicon. Although the basic education texts also reflect a high level of difficulty, their slightly lower scores (FKGL: 99.38; GFI: 104.71) suggest a comparatively more accessible syntactic structure. These results are consistent with those of Rafatbakhsh and Ahmadi (2023), who identified the content-word overlap index and the FRE metric as the strongest predictors of reading difficulty in graduate entrance-exam passages. This reinforces the relevance of the readability indices applied in the present study, particularly FKGL, in accurately reflecting the degree of complexity present in academic texts.

From a lexical perspective, both disciplines exhibit high density, slight lexical variation, and sophistication, contributing significantly to the overall textual complexity. Abstracts in basic education show higher lexical variation (0.18 vs. 0.16) and a marginally greater proportion of sophisticated words (28.50 vs. 27.04), suggesting a broader, but also more demanding, lexicon. This finding aligns with Maamuujav (2021), who notes that EFL writers typically produce texts with low lexical variation and limited academic vocabulary, often relying on repeated basic words. The slight lexical variation and sophistication observed in the basic education abstracts suggest a diverse core lexical repertoire, which correlates with increased textual complexity and higher-quality writing. It also aligns with Bashori (2021), who highlights intra-individual interplay between lexical diversity and sophistication over time, indicating that such growth often reflects greater lexical creativity rather than a purely linear advance. As for discourse markers, there is a predominant use of interactive resources, especially transitional markers, reflecting an effort to organize the discourse and maintain textual coherence. However, the limited use of interactional resources, such as self-mentions and engagement markers, suggests an objective, impersonal writing style, potentially reducing reader engagement. This observation echoes findings from studies on academic writing, such as those of Alfinda et al. (205), who report that L2 writers often use self-referential devices infrequently, favouring a formal and impersonal style. In summary, although both corpora exhibit linguistic features that hinder comprehension, the early childhood education texts show greater overall syntactic complexity, underscoring the need to reinforce academic writing practices in degree-level work.

The results revealed clear differences between the abstracts of early childhood education and basic education in terms of readability, lexical richness, and use of discourse markers. The early childhood education texts present greater reading difficulty, especially at the syntactic level. By contrast, basic education texts show greater lexical variation and sophistication. In both cases, the limited use of interactional resources suggests objective writing that could reduce reader involvement.

Qualitative Stage

In the abstracts of early childhood education and basic education, differences emerge in the use of discourse markers. In early childhood education, sentences containing frequent words, such as *since*, fall into the category of transition markers. These elements establish logical connections between ideas within a sentence. For example, in abstract No. 6, the use of *since* appears in a sentence with nine embedded words preceding the main verb, serving as evidence of a complex syntactic structure. By contrast, in basic education abstracts, one of the most

recurrent words is *determined*, which functions as an evidential marker. Such markers indicate the source of information — that is, they show that the authors ground their claims in existing evidence or prior analysis. In abstract No. 5, *determined* appears in a sentence with nine words before the main verb. Here, it serves both as the passive-voice main verb and as an evidential marker, introducing a statement based on a theoretical review. This use underscores objectivity and a deliberate distance from the author.

Likewise, the expression *such as*, a discourse marker classified under the code gloss category, was identified in the same abstract. Code glosses consist of expressions or words that clarify, explain, or exemplify ideas within the text, thereby facilitating readers' comprehension. In this case, the sentence contained nine words preceding the main verb, further reflecting syntactic complexity. Additionally, hedges were identified in childhood education texts. These markers mitigate the strength of a statement by indicating caution or possibility rather than absolute certainty. The following excerpts illustrate the function and the categories to which they belong:

1) "As far as the theoretical review is concerned, it can be determined that attachment is the affective bond that is consolidated from birth, *since* the first attachment bond is the mother, in whom she will feel safe and protected before any threat..."
2) "As far as the theoretical review is concerned, it can be determined that attachment is the affective bond that is consolidated from birth..."
3) "It was determined that other factors also influence, *such as* the infant's personality, physical or emotional abandonment and attachment, which is the difficulty of becoming familiar with the baby..."
4) "In conclusion, it was possible to identify the social and educational that children with dyslexia face."

As can be seen in excerpt 1, the word *since* belongs to the category of transitional discourse markers and, according to the interactional category proposed by Hyland (2015), it functions as a causal connector, introducing a reason or cause. If *since* is replaced by because, the causal sense is maintained, which confirms its explanatory function within the discourse. In excerpt 2, the expression 'can be determined', contains the main verb of the sentence and is classified under the hedge type, within the interactional category. Its function is to soften the assertion and protect the author against possible counterarguments because it avoids an absolute statement. In excerpt 3, the expression *such as* acts as an explanatory discourse marker that introduces an enumeration of examples and is placed in the code gloss markers. These markers facilitate readers' understanding by clarifying, reformulating, or exemplifying the idea. Finally, in excerpt 4, the expression, 'it was possible', indicates possibility, tentativeness, or absence of absolute certainty.

According to Hyland's (2015) classification, *possible* is considered an interactional resource that functions as a hedge, attenuating the strength of the assertion and allowing a more cautious stance from the author. These findings show how abstracts in early childhood education and basic education use different types of discourse markers to structure and qualify their academic texts, reflecting varying levels of syntactic complexity and communicative strategies. This finding is connected with those of Aparici et al. (2024), who identified that educational level directly influences the use and variety of discourse markers, thereby affecting, in some cases, the coherence, cohesion, and argumentative density of academic texts. The use of transitional,

evidential, clarifying, and hedging markers helps organize ideas more coherently and accurately (Kapranov, 2020). These features make texts easier to follow, which is fundamental for clarity and rigor in academic writing. By examining how metadiscourse markers are used in thesis abstracts, this analysis provides readers with a detailed view of the writing styles and discursive practices within each academic program.

Conclusions

The analysis of the thesis abstracts in early childhood education and basic education, using traditional readability formulas (FRE, FKGL, and GFI), yielded consistent results that show a high level of textual complexity in both cases. The interpretation of these indices aligns reasonably with classical reference tables, with minimal margins of error, validating their usefulness for comparative analysis in the academic environment. Nevertheless, the values obtained, well above expected school levels, reveal significant difficulty for the average reader, especially in early childhood education texts, characterized by greater syntactic and lexical density. Although these formulas offer an overview of textual complexity, they have limitations in capturing finer, contextual aspects of academic writing, such as discursive genre, modality, or communicative intention. Thus, it is recognized that reading comprehension depends not only on structural factors but also on variables such as linguistic competence, the reader's cognitive abilities, and familiarity with the content. This study, therefore, represents an initial stage in the assessment of readability in the field of teacher education and suggests the need to complement it with qualitative approaches that include the reader's perception and the context of use.

Regarding lexical richness, the findings confirm that both lexical density and lexical variation pose significant challenges to readability and clarity in academic texts. Although this study focused on early childhood and basic education disciplines, previous research in other disciplines, such as the English Literature program, has also indicated that lexical density and readability are critical factors in the comprehension of academic texts (see Siregar et al., 2024). This suggests that challenges related to textual accessibility may span multiple fields of study. Consequently, this research not only provides empirical evidence on the current state of academic writing in these disciplines but also establishes a solid foundation for future formative actions and comparative research in other educational contexts.

Additionally, the analysis of the metadiscourse confirms a limited use of interactive and interactional resources, which are fundamental for organizing the discourse, guiding the reader, and establishing a clear relationship with the text. This finding reinforces the need to strengthen academic writing skills among university students, particularly in producing clear, coherent, and adequate texts for their communicative purpose. In this sense, it is recommended to incorporate pedagogical strategies into teacher training that promote the conscious use of metadiscourse, together with the evaluation of readability and lexical richness from a multidimensional approach (AbManan & Raslee, 2020). The integration of these elements would not only improve the quality of academic texts but also develop writing skills transferable to future pedagogical practice. In this process, it is recommended to promote the use of technological tools such as Text Inspector or VocabProfile, which can be incorporated as support in academic writing classes to help students identify and correct areas of poor readability. This would contribute to the development of more conscious, measurable, and reader-oriented writing.

Technological tools such as Text Inspector and Vacabprofile can be effectively integrated into university writing courses to strengthen academic writing among nonnative speakers. For example, in a *Lexical Development Workshop*, students analyse their essays with Text Inspector to identify overused high-frequency words and replace them with academic vocabulary, improving lexical diversity and readability scores. In a *Sentence Structure and Cohesion module*, students use the tool to examine sentence length, clause density, and metadiscourse devices (e.g., connectors), then revise their paragraphs to balance syntactic complexity and clarity, tracking progress through readability indices such as Flesch-Kincaid. Finally, in a *Genre-Based Research Writing Project*, students compare their drafts with published papers in their disciplines using VocabProfile to examine lexical sophistication and Text Inspector to examine passive-voice use and cohesion patterns, refining their texts to match disciplinary conventions. These activities help students see measurable improvements in their writing, raise awareness of linguistic features that affect readability, and encourage independent revision skills supported by data-driven feedback.

Given the limited specific research on readability in early childhood education and basic education, it is hoped that the data obtained in this study will serve as a solid foundation for future research in this area. Although the results focus on these two, similar findings have been reported in other disciplines. This opens the door to expanding the analysis of readability to various educational and academic areas, promoting better understanding and access to texts for different audiences. These research findings are expected to encourage future studies that expand the number of corpora analysed and include other university programs. A specific case is that, in Ecuador, research related to readability in academic texts is scarce, especially in the field of teacher training. This situation raises questions about whether students in these majors receive sufficient orientation in academic writing and how this could influence the degree modality they choose or develop. In addition, the findings invite reflection on whether future teachers, responsible for teaching reading and writing, are adequately prepared to produce legible texts. This ability not only affects their academic performance but also their future classroom pedagogical practice.

Acknowledgements and Declarations

This manuscript is based on the research project titled “Academic Discourse: A Transdisciplinary Study between Education and Linguistics in Linguistic-Communicative Contexts for Teaching English.” The authors sincerely thank all those who supported its completion and confirm that this paper has not been submitted to any other journal for consideration. Furthermore, “the authors declare that no artificial intelligence tools were used in the preparation of this article”.

References

Allen, D. (2009). A study of the role of relative clauses in the simplification of news texts for learners of English. *System, 37*(4), 585–599. https://doi.org/10.1016/j.system.2009.09.004

Ash, R., & Edgell, S. (1975). A note on the readability of the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 60*(6), 765–766. http://dx.doi.org/10.1037/0021-9010.60.6.765

Abdullah, S. S. (2022). A pragmatic analysis of metadiscoursal markers (hedges and boosters) in linguistics and biology MA theses: A lingua scientific corpus-based study. *International Journal of Early Childhood Special Education, 14*(1), 1173– 1185. https://doi.org/10.9756/INT-JECSE/V14I1.221134

AbManan, N. A., & Raslee, N. S. (2020). Explicit discourse marker instruction to improve coherence and cohesion in academic writing. *International Journal of Academic Research in Business and Social Sciences, 8*(1), 250–263. https://doi.org/10.6007/IJARBSS/v8-i1/3820

Alfinda, R., Pranowo, D. D., & Widyantoro, A. (2025). Corpus analysis of Spanish L2 exam-based writing: Lexical analysis and indications of translingualism. *International Journal of Pedagogy and Teacher Education, 9*(1), 170–178. https://doi.org/10.20961/ijpte.v9i1.86337

Al-Harthi, M., Alshahrani, H. J., & Hamed, D. M. (2022). Evidentials, Code Glosses, Hedges and Boosters in Academic Articles: A Cross-Disciplinary Study. *Eurasian Journal of Applied Linguistics, 8*(3), 181–200. https://doi.org/10.32601/ejal.803015 https://files.eric.ed.gov/fulltext/EJ1386005.pdf

Alkaldi, W., & Inkpen, D. (2023). Text simplification to specific readability levels. *Mathematics, 11*(9), 2063. https://doi.org/10.3390/math11092063

Allen, D. (2009). A study of the role of relative clauses in the simplification of news texts for learners of English. *System, 37*(4), 585–599.

Aparici, M., Rosado, E., & Vilar, H. (2024). The influence of students' linguistic condition, school level, and pedagogical input on analytical essay features. *Frontiers in Language Sciences, 3*. https://doi.org/10.3389/flang.2024.1480422

Arbel, Y. A. (2024). The readability of contracts: Big data analysis. *Journal of Empirical Legal Studies, 21*(4), 927-978. https://doi.org/10.1111/jels.12400

Barberán, S. X., Mendoza Mejia, J. L., Torrealba Peña, M. J., Vásconez Rivera, I. N., & Hurtado Guevara, R. F. (2024). Reading Comprehension and Cognitive Processes in Ecuadorian Middle School Students. *Journal of Ecohumanism, 3*(8), 6305–6314. https://doi.org/10.62754/joe.v3i8.5232

Barrot, J. S., & Agdeppa, M. A. (2021). Investigating syntactic complexity and language- related error patterns in EFL students' writing: corpus-based and epistemic network analyses. *Cercles Journal of English Studies, 13*, 1–25. https://doi.org/10.1515/cercles-2023-2009

Bashori, M. (2021). The development of intra-individual variability in academic writing: A study on lexical diversity and lexical sophistication. *Studies in English Language and Education, 8*(2), 745-758. https://doi.org/10.24815/siele.v8i2.16843

Beier, S., Berlow, S., Boucaud, E., Bylinskii, Z., Cai, T., Cohn, J., ... & Wolfe, B. (2022). Readability research: An interdisciplinary approach. *Foundations and Trends in Human–Computer Interaction, 16*(4), 214-324. https://doi.org/10.1561/1100000089

Besuglov, E., & Crasselt, N. (2021). The effect of readability and language choice in management accounting reports on risk-taking: an experimental study. *Journal of Business Economics, 91*(1), 5-33. https://doi.org/10.1007/s11573-020-00980-4

Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. M. (2020). The impact of readable financial disclosures on the corporate bond market. *Contemporary Accounting Research, 37*(2), 789–817. https://doi/10.1111/1911-3846.12544

Cai, T., Wallace, S., Rezvanian, T., Dobres, J., Kerr, B., Berlow, S., ... & Bylinskii, Z. (2022). Personalized font recommendations: Combining ml and typographic guidelines to optimize readability. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (pp. 1-25). https://doi/abs/10.1145/3532106.3533457

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*(1), 92-113. https://files.eric.ed.gov/fulltext/EJ1031308.pdf

Deutsch, T., Jasbi, M., & Shieber, S. (2020). Linguistic features for readability assessment. In *Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications*, (pp.1–17), Seattle, WA, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.bea-1.1

Fry, E. (1968). A Readability Formula That Saves Time. *Journal of Reading, 11*(7), 513–578. http://www.jstor.org/stable/40013635

Guerrero, S. E., Astudillo, M., Jurado, V. O., Robalino, G., & Riera, F. (2025). Online Academic Tutoring for English Language Learning: A Case Study of Receptive Skills Development in Ecuadorian Students. *World Journal of English Language, 15*(1), 413-413. https://doi.org/10.5430/wjel.v15n1p413

Han, Y., Ceross, A., & Bergmann, J. H. (2024). The use of readability metrics in legal text: A systematic literature review. *Computer Science: Computation and Language*, 1-22. https://doi.org/10.48550/arXiv.2411.09497

Haninda, M. H., & Bram, B. (2021). Academic writing of EFL students' undergraduate theses: A discourse marker analysis. *ELS Journal on Interdisciplinary Studies in Humanities, 51*(1), 109–114. https://doi.org/10.34050/elsjish.v5i1.19719

Hu, T. (2023). Readability assessment of advanced English textbooks: A corpus- linguistic study. *Theory and Practice in Language Studies, 13*(1), 251-256. https://doi.org/10.17507/tpls.1301.29

Hyland, K. (2015). Metadiscourse. In Tracy, K. (ed.) *International Encyclopaedia of Language and Social Interaction*. Oxford: Wiley-Blackwell

Imperial, J. M. (2021). BERT Embeddings for Automatic Readability Assessment. *Computer Science: Computation and Language*, 1-8. https://doi.org/10.48550/arXiv.2106.07935

Kapranov, O. (2020). The use of discourse markers in academic writing by in-service primary school teachers of English. *Polish Journal of English Studies, 6*(1), 37–52. https://doi.org/10.2478/pjes-2020-0009

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

Kirk, S. (2007). Methodological and ethical issues in conducting qualitative research with children and young people: A literature review. *International Journal of Nursing Studies, 44*(7), 1250–1260. https://doi.org/10.1016/j.ijnurstu.2006.08.015

Kukafka, R., & Arguello, M. (2021). Assessing the readability of policy documents on the Digital Single Market of the EU. *European Policy Analysis, 7*(2), 235–250. https://doi.org/10.1002/epa2.1110

Lazzeretti, C. (2021). Communicating Sustainable Tourism in English and Italian: A Contrastive Analysis. *Linguæ &-Rivista di lingue e culture moderne, 19*(2), 133-154. https://doi.org/10.7358/ling-2020-002-lazz

Lee, J., & Vajjala, S. (2022). A neural pairwise ranking model for readability assessment. *Computer Science: Computation and Language*, 1-12. https://doi.org/10.48550/arXiv.2203.07450

Li, C., Wang, X., & Qian, L. (2025). Exploring syntactic complexity and text readability in an ELT textbook series for Chinese English majors. *SAGE Open, 15*(1), 1-17, https://doi.org/10.1177/21582440251323619

Long, M. (2022). Effects of lexical richness on text readability of medical journal articles. *International Journal of Social Science and Education Research, 5*(5), 80–87. https://doi.org/10.6918/IJOSSER.202205_5(5).0011

Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing, 47*, 100540. https://doi.org/10.1016/j.asw.2021.100540

Mahadini, M. K., Setyaningsih, E., & Sarosa, T. (2021). Using conventional rubric and Coh-Metrix to assess EFL students' essays. *International Journal of Language Education, 54*, 260-270. https://doi.org/10.26858/ijole.v5i4.19105

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behaviour Research Methods, 45*, 499-515. http://dx.doi:10.3758/s13428-012-0258-1

Meng, C., Chen, M., Mao, J., & Neville, J. (2021). Readnet: A hierarchical transformer framework for web article readability analysis. In A*dvances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42* (pp. 33-49). Springer International Publishing. https://arxiv.org/pdf/2103.04083

Mora, F., Quito, R., & Macías, L. (2021). Reading comprehension and reading speed of university English language learners in Ecuador. *Journal of English Language Teaching and Applied Linguistics, 3*(11), 11-31. https://doi.org/10.32996/jeltal.2021.3.11.3

Müngen, A. (2023). Exploring Lexical Richness in English-Language theses Across Disciplines: A Comparative Analysis. *Journal of Scientometric Research, 12*(2), 459-468. https://doi.org/10.5530/jscires.12.2.042

Nasrollahi, A., Amini, D., & Nourdad, N. (2022). The effects of interactive writing strategy on lexical complexity in EFL students' written performance. *Journal of Language and Linguistic Studies, 18*(1), 280–293. https://doi.org10.48185/tts.v6i1.1364

Noguera, C. (2013). English-written abstracts for Spanish publications: A challenge in the globalization of science. *LFE: Revista de Lenguas para Fines Específicos, 19*, 365–384. https://ojsspdc.ulpgc.es/ojs/index.php/LFE/article/view/24

Ondondo, E. A., & Rew, F. O. (2020). Discourse markers and plagiarism in the literature review section of a research thesis: A study in Kenya. *Communication and Linguistics Studies, 6*(2), 27–33. https://doi.org/10.11648/j.cls.20200602.13

Pasaribu, A. N., Pasaribu, T. K., Sinambela, E., & Manullang, V. R. (2022). Interactive metadiscourse markers in Indonesian EFL students' academic writing. English Review. *Journal of English Education, 10*(3), 883-890. https://doi.org/10.25134/erjee.v10i3.6817

Rafatbakhsh, E., & Ahmadi, A. (2023). Predicting the difficulty of EFL reading comprehension tests based on linguistic indices. *Asian-Pacific Journal of Second and Foreign Language Education, 8*(41), 1-15. https://doi.org/10.1186/s40862-023-00214-4

Riznanda, W. A. (2021). Discourse Markers in Tertiary Level Students' Essay Writing: Ability and Problems. *English Franca: Academic Journal of English Language and Education, 5*(1), 1–16. https://doi.org/10.29240/ef.v5i1.1193

Siregar, M. S., Hara, N. T., & Sinar, T. S. (2024). Lexical density and readability in students' thesis introductions. *KLAUSA: Kajian Linguistik, Pembelajaran Bahasa, dan Sastra, 8*(1), 71-79. https://jurnal.machung.ac.id/index.php/klausa/article/view/948/395

Stone, R., de Hoop, T., Coombes, A., & Nakamura, P. (2019). What works to improve early grade literacy in Latin America and the Caribbean? A systematic review and meta-analysis. *Campbell Systematic Reviews, 16*(1), e1067. https://pmc.ncbi.nlm.nih.gov/articles/PMC8356313/

Tarzeen, R., & Tahir, A. (2024). Analyzing the Impact of Discourse Markers in Academic Writing: A Corpus-Based Approach. *Pakistan Languages and Humanities Review, 8*(3), 138-166. https://doi.org/10.47205/plhr.2024(8-III)13

Tiago, F., Gil, A., Stemberger, S., & Borges-Tiago, T. (2021). Digital sustainability communication in tourism. *Journal of Innovation & Knowledge, 6*(1), 27-34. https://doi.org/10.1016/j.jik.2019.12.002

Tovar, R. (2022). Lexical richness of abstracts in scientific papers in Anglophone and non-Anglophone journals. *MEXTESOL Journal, 46*(2), 1-12. https://doi.org/10.17576/3L-2022-2802-15

Vajjala, S. (2021). Trends, Limitations and Open Challenges in Automatic Readability Assessment Research. *Computer Science: Computation and Language*, 1-12. https://arxiv.org/abs/2105.00973

Wang, J., & Zeng, L. (2021). Disciplinary recognized self-presence: Self-mention used with hedges and boosters in PhD students' research writing. *SAGE Open, 11*(2), 1-13, https://doi.org/10.1177/21582440211005454

Xu, Y., & Casal, J. E. (2023). Navigating complexity in plain English: A longitudinal analysis of syntactic and lexical complexity development in L2 legal writing. *Journal of Second Language Writing, 62*, 101059. https://doi.org/10.1016/j.jslw.2023.101059

Young, D. J. (1999). Linguistic Simplification of Second Language Reading Material: Effective Instructional Practice? *The Modern Language Journal, 83*(3), 350-366. https://www.jstor.org/stable/330258

Yu, Z. (2022). Financial report readability and accounting conservatism. *Journal of Risk and Financial Management, 15*(10), 454. https://doi.org/10.3390/jrfm15100454

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

Zhang, W. (2020). Lexical richness in L2 writing: An investigation of advanced EFL learners' lexical use. *System, 96*, 102402. https://doi.org/10.1016/j.system.2020.102402