Use Eye Movement Features to Explore the Impact of Language Features on Text Difficulty

Yi Ju Chan, National Taiwan University of Science and Technology, Taiwan
Hou-Chiang Tseng, National Taiwan University of Science and Technology, Taiwan
Yao-Ting Sung, National Taiwan Normal University, Taiwan

The Asian Conference on Education 2024
Official Conference Proceedings

Abstract
Lifelong learning emphasizes continuous learning throughout one's life, with reading enriching elders' lives and aiding their adaptation to aging. However, aging is often associated with a decline in cognitive functions and an increased risk of impairment, making providing appropriate reading materials for elders a crucial matter. Scholars have explored the correlation between eye movement indicators and text difficulty. However, previous research has rarely explicitly linked eye movement indicators to linguistic features, which may limit authors' ability to comprehensively assess their texts' readability. Therefore, this study utilizes an eye tracker to observe elders' reading behavior when engaging with texts of varying difficulties, extracts eye movement indicators to examine challenges faced by elders with different reading abilities, and employs CRIE (Chinese Readability Index Explorer) to analyze the computational linguistic features. The aim is to understand why certain linguistic features pose reading challenges for elders. The findings revealed that when elders read difficult texts, their dwell time and regression count increased, particularly with longer sentences and unfamiliar vocabulary. Linguistic indicators exhibited moderate correlations with several eye movement indicators, suggesting that language structure significantly influences reading behavior. This study, beginning with the reader's reading process, identifies specific correlations between eye movement indicators and text difficulty, offering a novel perspective for research in eye movement and computational linguistics. Furthermore, it has implications for the cognitive health management of elders in medical and psychological fields, ultimately enhancing their quality of life and promoting lifelong learning.

Keywords: Eye Tracker, Linguistic Feature, Text Difficulty, Text Readability, Elders

1. Introduction

Since Taiwan entered an aging society in 1993, the proportion of people over 65 years old has been increasing year by year. According to statistics from the Ministry of Interior (2024), the population over 65 years old in Taiwan has reached 4.158 million, accounting for 17.8% of the total population. Given this data, aging has gradually become an important issue in Taiwan. For example, lifelong learning has been the focus of aging in recent years. In contemporary education and psychological research, lifelong learning is widely considered to profoundly impact individual cognitive, psychological, and social well-being. For the elderly, reading is not only a means of acquiring knowledge, but also a key activity for maintaining cognitive functions, promoting mental health, and social cipation (Rigg & Kazemek, 1983; Zhang et al., 2022). In light of the growing elderly population, it is imperative to provide materials that are appropriate for the reading abilities of elders.

Reading is a cognitively demanding activity that requires the coordination of both linguistic and non-linguistic cognitive functions, such as attention, abstract reasoning, working memory, long-term memory, and lexical and conceptual retrieval (Kweldju, 2015; Afifah et al., 2008). As people age, there are natural declines in memory, language, and other cognitive functions, which is a normal aspect of the aging process. Aging leads to cognitive decline and memory deterioration, among other cognitive impairments (Pearman & Storandt, 2005). Therefore, reading materials for elders should consider cognitive load to ensure that they can engage with reading while also receiving an appropriate cognitive challenge.

Many studies have started to focus on text difficulty, with research commonly falling into two categories. First, some studies use multiple-choice questions and cloze tests to assess text difficulty. For example, Bormuth (1969) utilized readers' performance on cloze tests as an objective measure of text readability. Fanguy and colleagues (2004) used cloze tests to determine whether readers understood a company's privacy policy. Similarly, the McCall-Crabbs Standard Test Lessons in Reading designed multiple-choice questions based on the writing and content of the texts, requiring readers to read the passages and respond to questions (McCall & Crabbs, 1961). Second, traditional readability models assess difficulty based on vocabulary quantity and complexity. For instance, Lively and Pressley (1923) argued that textbooks filled with technical terminology made it difficult for students to learn, while the Flesch Reading Ease formula measures readability by considering the average number of syllables per word and sentence length (Flesch, 1948). Although research on text difficulty is abundant, these readability assessment standards often predict readability using surface-level linguistic features, which fail to reflect the deeper cognitive mechanisms involved in reading (Benjamin, 2012; Bormuth, 1966; Davison & Kantor, 1982).

In comparison to the limitations of the traditional readability models previously discussed, eye-tracking technology provides a relatively objective method for investigating how readers engage with texts (Rayner, 1998). Eye movements are frequent and highly variable (Shan & Zhou, 2020), making eye-tracking an effective tool for providing real-time information about a user's psychological state (Djamasbi et al., 2010) and reflecting cognitive processing through eye-tracking indicators (Just & Carpenter, 1976; Martin et al., 2011). This method has been widely used to explore topics related to reading processes, such as eye movement characteristics, perceptual span, and information integration. Furthermore, eye-tracking data is frequently utilized to investigate cognitive processes across diverse cognitive tasks (Rayner, 1998), and eye-tracking devices provide sufficient indicators to capture real-time cognitive processing during reading (Jian, 2022). As this technology has matured, it has continued to be applied in various fields of learning research, particularly in process-oriented studies. For instance, Wang and Jian (2022) reviewed nearly 30 years (1990-2020) of research on the reading processes of scientific text and images, revealing that research in this field has primarily focused on university students, with a few studies involving elementary school students (Guo & Jian, 2022). However, there is a notable lack of studies examining the cognitive processes of elders. Therefore, this study utilizes eye-tracking to observe the eye movement indicators of elders as they read texts of varying difficulty, aiming to explore the relationship between these eye movement indicators and text difficulty. The results can be applied to cognitive health management for elders, contributing to improving their quality of life and promoting lifelong learning.

2. Literature Review

Many studies have explored the relationship between eye-tracking indicators and text difficulty, although the methods used by each researcher vary. For example, Ko and colleagues (2005) examined the effects of term frequency and word class on eye movement patterns during the Chinese reading processes of university students. The research found that when analyzing words with available word frequency data, higher-frequency words tended to have shorter fixation times in both the first fixation duration and rereading gaze metrics when reading expository texts. This indicates that word frequency significantly influences word recognition during reading (Ko et al., 2005; Carpenter & Just, 1983; Rayner & Duffy, 1986). Ko and colleagues (2005) also examined the relationship between word class and fixation time, discovering that postpositions, a type of function word (e.g., "after" or "between"), had the highest fixation rates and the longest first fixation duration. They suggested that this could be related to temporal and referential coherence in narratives, which are crucial indicators of whether readers can understand the story content. This phenomenon also appears in English. Graesser and colleagues (1994) and Zwaan and colleagues (1995) found that postpositions in English bear a significant cognitive load for conceptual understanding. Liu and Zhou (2019) explored translation difficulty using fixation and saccade durations. Just and Carpenter (1993) investigated sentence comprehension among university students by examining pupil dilation, finding that more complex sentences required longer processing times and resulted in more significant pupil dilation.

In studies using eye-tracking to assess reading task difficulty, the results show that readers tend to have longer fixation times when processing the following types of words: long words (Just & Carpenter, 1980; Rayner et al., 1996), low-frequency words (Just & Carpenter, 1980; Inhoff, 1984; Rayner & Fischer, 1996; Rayner & Raney, 1996), novel or unfamiliar words (Chaffin et al., 2001; Williams & Morris, 2004), ambiguous words (Rayner & Duffy, 1986; Sereno et al., 2006), and words that are contextually unrestricted or difficult to predict based on the surrounding context (Ehrlich & Rayner, 1981; Zola, 1984; Rayner & Well, 1996; Ashby et al., 2005). Table 1 summarizes the above-mentioned studies, showing that scholars have explored text difficulty through eye-tracking indicators, but most of the participants in these studies were university students. In addition, researchers also conducted statistical analyses on the range of values for eye movement indicators. For example, Rayner (1998) reviewed 20 years of eye-tracking research (1978-1998) on reading and other information processing tasks and found that fixation duration typically ranges from 100ms to 500ms, and saccadic latency falls between 150ms and 250ms. Bentivoglio and colleagues (1997) observed that adults blink at a frequency of 10-20 times per minute. In summary, the variables considered in these studies were relatively limited, primarily focusing on basic linguistic features such as word frequency and word class. Important factors such as syntactic complexity, contextual relevance, and the reader's prior knowledge were not fully addressed, limiting the studies' ability to accurately reflect the true difficulty of the texts (Collins, 2014).

Table 1: Research on Text Difficulty Using Eye-Tracking Indicators

Eye tracking indicators	Linguistic features	Text difficulty	Participants	References
Fixation Duration Saccade		Word Translation Difficulty	College Students	Liu & Zhou, (2019)
Fixation Duration Character per minutes		Classical Chinese Difficulty	College Students High School Students	Chen (2014)
First Fixation Duration Rereading Gaze	Word frequency Word class		College Students	Ko et al. (2005)
Fixation Duration	Word class		College Students	Ko et al., (2005)
Pupil Dilation		Sentence Complexity	College Students	Just & Carpenter, (1993)

The analysis of eye-tracking indicators is highly complex. Researchers not only examine eye-tracking data from a macro perspective but also conduct detailed analyses of Areas of Interest (AOIs), which may include individual words, sentences, or broader regions (Wang & Chien, 2022). This study references and integrates the eye-tracking indicators used in previous literature (Wang & Chien, 2022; Chen et al., 2010; Chien & Wu, 2012; Alemdag & Cagiltay, 2018; Hyönä et al., 2003; Mason et al., 2018). Although there is a wide variety of eye-tracking indicators, due to space limitations, Table 2 lists only five commonly used indicators. These indicators help us more precisely observe participants' reading behavior and identify their cognitive characteristics and challenges when processing texts.

Table 2: Common Eye-Tracking Indicators

Eye tracking indicators	Definition/Meaning	References
Total fixation count, TFC	The total number of fixation points within the AOI. Generally, a higher fixation count indicates a higher level of cognitive processing and is highly correlated with total fixation duration.	Eitel (2016); Schnotz & Wagner (2018)
First-pass fixation duration	The total duration of all fixations from the first entry into an AOI until leaving it, typically reflecting initial processing, such as semantic processing after word recognition.	Hyönä et al. (2003); Mason et al. (2013)
Regression count	The total number of times a fixation leaves an AOI and then re-enters it, typically reflecting late-stage processing, particularly when the reader encounters content they do not understand.	Chen et al. (2010); Mason et al. (2016); Schüler (2017)

Percentage of fixation duration	Other fixation duration metrics (e.g., total fixation duration on an AOI, first fixation duration, or re-fixation duration) divided by the total fixation duration during the learning period, usually reflect the reader's selective attention allocation.	Jian & Wu, (2012); Alemdag & Cagiltay, (2018); Mason et al., (2013)
Total fixation duration	The total duration of fixations within the AOI generally reflects the level of cognitive processing during reading. Longer fixation durations indicate that the reader needs to expend more cognitive resources to process the information in that area.	Chen et al., (2010); Jian & Wu, (2012)

Since different eye-tracking indicators capture various layers of cognitive activity during the reading process, this study aims to explore how elders with different reading abilities perform when reading texts of varying difficulty. To achieve this goal, this study utilizes the Diagnostic Assessment of Chinese Competence (DACC) developed by Li and colleagues (2021) to evaluate the reading ability of participants. Eye-tracking technology is used to record participants' eye movements, and eye-tracking indicators are extracted to analyze the corresponding sentences or words. Additionally, the CRIE (Chinese Readability Index Explorer) system (Sung et al., 2016) is employed to analyze the computational linguistic indicators of text. This study aims to explore the relationship between eye-tracking indicators and computational linguistic indicators. This study not only establishes a concrete connection between eye-tracking indicators and text difficulty in the fields of linguistics and eye-tracking but also provides new perspectives and methodological support for future eye-tracking research. In the fields of medicine and psychology, it can be applied to cognitive health management for elders, contributing to improving their quality of life and promoting lifelong learning.

3. Methodology

This study observes the reading behavior of elders using an eye tracker while they read texts of varying difficulty. Eye-tracking indicators are extracted to examine the challenges faced by elders with different reading abilities during the reading process. Furthermore, computational linguistic features of non-proper nouns, proper nouns, sentences, and passages are analyzed to explore how these features impact reading challenges for elders. The proper nouns are diabetes-related terms defined by the International Diabetes Federation (2024), the American Diabetes Association (2024), and SA Health (2024), while non-proper nouns are words excluding these diabetes-related terms. This study conducts eye-tracking experiments from the reader's perspective, using heat maps for visual analysis and defining Areas of Interest (AOI) based on non-proper nouns, proper nouns, and sentences. Various eye-tracking indicators are captured and analyzed, and the extracted words and sentences are input into the Chinese Readability Index Explorer (CRIE) (Sung et al., 2016) to observe the relationship between computational linguistic indicators and eye-tracking indicators. By aligning these two types of indicators, the study aims to achieve its research objectives.

3.1 *Participants*

The participants in this study were 15 elders aged 65 and above from the "Dinning Together for the Elderly" program in Taipei City. All participants were native Chinese speakers with normal or corrected-to-normal vision. One participant did not meet the age requirement, two participants had partial data loss due to failure in eye-tracking during the experiment, and one participant skipped pages while reading. Therefore, a total of 4 datasets were excluded, leaving 11 usable datasets.

3.2 *Apparatus*

This study used the SR EyeLink 1000 eye-tracking system, with a sampling rate of 1000 Hz (i.e., 1,000 samples per second). The reading texts were displayed on a 24-inch LCD monitor with a resolution of 1920 x 1200, and the text size was set to 28. The distance between the participants and the monitor was 65 cm.

3.3 *The Experimental Text*

According to the report《Taiwan Diabetes Report 2000-2045》by the International Diabetes Federation (IDF) Diabetes Atlas, Taiwan has the highest prevalence of diabetes in Asia, with more than one in ten people potentially facing diabetes-related issues. Therefore, this study selected diabetes as the topic for the experimental texts. The texts provided to the participants were divided into three difficulty levels: easy (6th grade), medium (9th grade), and difficult (12th grade). The diabetes-related content was written by six professional reading teachers. The readability of the texts was confirmed using the CRIE (Chinese Readability Index Explorer) readability model (Sung et al., 2016), and the final difficulty levels were approved by the six reading experts. The accuracy of the medical content was also reviewed and verified by professional doctors.

3.4 *Diabetes Knowledge Questionnaire*

Before conducting the eye-tracking experiment, participants in this study were asked to complete the Diabetes Knowledge Questionnaire (DKQ) (Gracia et al., 2001) to assess their prior knowledge of diabetes. This step aimed to determine whether the participants' eye-tracking performance was driven by their reading ability or influenced by their familiarity with diabetes-related background knowledge. By administering this questionnaire, the study sought to investigate whether participants' eye-tracking data during the diabetes-related texts were associated with their prior knowledge, thereby further examining how differences in prior knowledge might impact reading behavior.

The Diabetes Knowledge Questionnaire (DKQ) is a standardized tool used to assess participants' understanding of diabetes knowledge and has demonstrated good internal consistency, with a Cronbach α coefficient of 0.78. The Traditional Chinese version of the questionnaire was translated by a native Mandarin speaker with a Bachelor's degree in English and back-translated by a Taiwanese professional translator with a Master's degree in English. The translation process was reviewed by two experts: a psychometrician and a doctor certified in diabetes education. The Simplified Chinese version of the DKQ, developed by Hu and colleagues (2013), was also referenced to ensure linguistic accuracy and cultural appropriateness. Through this questionnaire, the study can compare whether there are differences in prior diabetes knowledge among the participants.

3.5 *Diagnostic Assessment of Chinese Competence (DACC)*

During the experiment, participants will complete the Diagnostic Assessment of Chinese Competence (DACC) (Lee et al., 2021) to evaluate their reading abilities. This assessment is conducted in the form of a computer-adaptive test that comprehensively evaluates participants' reading proficiency. The DACC system assesses reading comprehension skills, including understanding (e.g., vocabulary, literal and inferential comprehension), contextual integration, analysis, and evaluation. The assessment is designed for students from second to twelfth grade. The DACC test items were drafted by school teachers, PhD students in psychology, and professionals in Chinese language research. All drafters must undergo and pass training before they are allowed to contribute test items to the DACC system. The DACC consists of multiple-choice questions, with each question corresponding to one of five dimensions: vocabulary, literal understanding, contextual integration, inferential comprehension, analysis, and evaluation. This design allows for a comprehensive measurement of student performance across these dimensions, providing an in-depth analysis of their reading ability upon completion of the DACC.

3.6 *Procedure*

The experimental procedure of this study is illustrated in Figure 1. Before the eye-tracking experiment begins, participants are first asked to sign an informed consent form, indicating that they understand the study details and agree to participate. After completing the informed consent form, participants are given the Diabetes Knowledge Questionnaire (DKQ) to assess their prior knowledge of diabetes. Next, participants take the Diagnostic Assessment of Chinese Competence (DACC) to evaluate their reading ability. After completing this assessment, participants are given a 10-minute rest before proceeding to the eye-tracking experiment. Before the formal experiment, the researcher asked participants to position themselves comfortably on the eye-tracker mount and adjusted their positions accordingly. Calibration and validation tests, necessary for data collection using the eye-tracker, were then conducted. The researcher used a nine-point calibration procedure, during which participants were instructed to steadily gaze at each calibration point until all nine were completed. If the calibration results showed no significant errors, the eye-tracker computed the correspondence function between the screen and eye movements, converting eye movement distances directly into screen coordinate displacements. The text was divided into three levels of difficulty—easy, medium, and hard. After reading each section, participants answered simple questions to ensure comprehension. To reduce cognitive load and eye fatigue, participants rested for five minutes after completing each section before proceeding to the next.

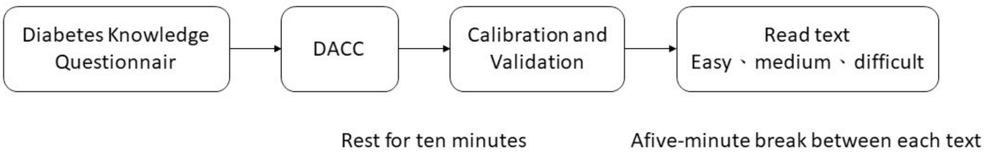

Figure 1: Experimental Procedure

4. Results

4.1 *Analysis Based on Areas of Interest: A Comparison of Sentences, Non-proper Nouns, and Proper Nouns*

4.1.1 *Diabetes Prior Knowledge Assessment*

In this study, participants were divided into two groups based on their reading ability: the low reading ability group (below 7th grade) and the high reading ability group (8th grade and above). The Diabetes Knowledge Scale (DKS) was used to assess participants' prior knowledge of diabetes. The average number of correct answers on the scale for the low reading ability group was 15.00, while the average for the high reading ability group was also 15.00. The results showed no significant difference between the two groups in their diabetes-related background knowledge, indicating that both groups had a similar level of understanding of the subject. Therefore, if significant differences are observed in eye-tracking indicators in subsequent analyses, these differences are likely attributable to variations in reading ability rather than differences in background knowledge about diabetes.

4.1.2 *Choose AOI*

In this study, participants were first divided into two groups based on their reading ability: low (below 7th grade) and high (8th grade and above). Heat maps for participants with different reading abilities were generated (as shown in Figures 2-4). Based on the heat maps of these two reading ability groups, non-proper nouns, proper nouns, and sentences that showed differences between the groups were selected as Areas of Interest (AOI). This selection aimed to identify which text components were processed differently by participants with varying reading abilities.

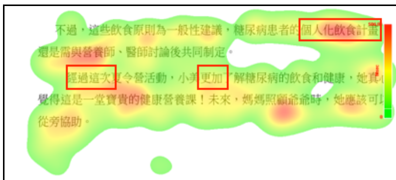	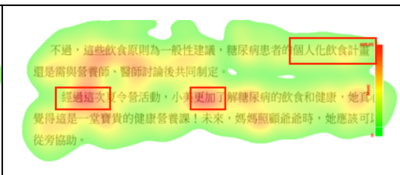	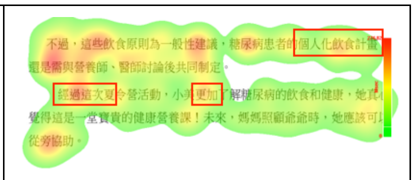
Figure 2: Heat Map for Low Reading Ability With Low Text Difficulty	Figure 3: Heat Map for Medium Reading Ability With Low Text Difficulty	Figure 4: Heat Map for High Reading Ability With Low Text Difficulty

4.1.3 *Data Analyze*

Previous reading studies have often focused on analyzing individual word-level features (such as word frequency and word length) or used sentences as the sole unit of analysis. While this approach has provided valuable insights, it has not fully explored how readers simultaneously process both lexical and sentence-level information during reading, particularly the differences between proper nouns and non-proper nouns, which have been less examined in past research. To address this gap, this study divides the Areas of Interest (AOI) into three categories for data analysis: "non-proper nouns," "proper nouns," and "sentences."

This study consists of three parts, each focusing on a different unit of analysis: non-proper nouns, proper nouns, and sentences. After the experiment is completed, the collected eye-tracking data and assessment results from each participant will be analyzed. The data analysis includes:

- Summarizing the reading results of the participants.
- Converting each participant's eye-tracking data while reading the three texts of varying difficulty into mp4 or avi files for post-experiment review.
- Using the Data Viewer software to define AOI regions and extract the AI report data for further analysis.

In eye-tracking studies focused on reading plain text, AOI (Areas of Interest) can be analyzed at the word, sentence, or paragraph level. The eye-tracking indicators selected for this study include Dwell Time, Fixation Count, Dwell Time Percentage (Dwell Time%), First Fixation Duration, and Regression Count, as detailed in Table 2, which provides a comprehensive explanation of each indicator's meaning.

4.1.4 *Analysis Results Based on Non-proper Nouns*

Following the experimental design of Jian and Ko (2017), this study employed a two-way mixed design ANOVA to examine the effect of vocabulary across different text difficulty levels and its interaction with reading ability on eye-tracking indicators. The results, as shown in Table 3, revealed that text difficulty had a significant main effect on total fixation count, $F(2, 9) = 4.26$, $p < .05$, $\eta^2 = .32$, indicating that text difficulty significantly influenced the total number of fixations. Post-hoc comparisons revealed that participants had significantly more fixations when reading simple texts compared to medium-difficulty texts. Although the interaction between reading ability and text difficulty did not reach statistical significance, an interesting pattern emerged from the raw data: in the low reading ability group, two participants had a higher fixation count for simple texts than for medium-difficulty texts. Similarly, three participants in the high reading ability group exhibited the same pattern. Even though the simple texts appear to be easier on the surface, participants may have needed more time to verify and comprehend the content, leading to a higher number of fixations. This outcome suggests that participants, when faced with simpler texts, might spend more time confirming information.

Table 3: Post-hoc Comparison Table of Fixation Count for Non-proper Nouns Across Different Text Difficulties

SV	*SS*	*df*	*MS*	*F*	η^2	Post-hoc Comparison
Reading Ability	0.01	1	0.01	0.01	0.00	
Error	9.99	9	1.11			
Text Difficulty	2.62	2	1.31	4.26*	0.32	Easy > Medium
Text Difficulty * Reading Ability	0.33	2	0.17	0.54	0.06	
Error	5.53	18	0.31			

*p<.05

4.1.5 *Analysis Results Based on Proper Nouns*

This study examined the effects of proper nouns on eye-tracking indicators about different text difficulty levels and their interaction with reading ability. The results, as shown in Table 4, revealed that text difficulty had a significant main effect on total dwell time, $F(2, 18) = 5.12$, $p < .05$, $\eta^2 = .36$, indicating that text difficulty significantly influenced total dwell time.

Post-hoc comparisons showed a significant difference between difficult and easy texts, with participants spending significantly more time on higher-difficulty texts. This finding aligns with Rayner and colleagues (2006), who suggested that processing time increases with text difficulty. The results indicate that when participants faced more challenging texts, they needed more time to comprehend the content, resulting in extended total dwell time. This outcome further supports the view that increased cognitive load during complex text processing significantly affects reading behavior.

Table 4: Post-hoc Comparison Table of Dwell Time for Proper Nouns Across Different Text Difficulties

SV	*SS*	*df*	*MS*	*F*	η^2	Post-hoc Comparison
Reading Ability	85509.63	1	85509.63	0.20	0.02	
Error	3845811.66	9	427312.41			
Text Difficulty	983292.20	2	491646.10	5.12*	0.36	Hard > Easy
Text Difficulty * Reading Ability	145779.50	2	72889.75	.76	0.08	
Error	1728420.79	18	96023.38			

*p<.05

As shown in Table 5, the results for first fixation duration indicate that text difficulty had a significant main effect, $F(2, 18) = 5.52$, $p = .01$, $\eta^2 = .38$, demonstrating that text difficulty significantly influenced the first fixation duration. Post-hoc comparisons revealed a significant difference between high and low-difficulty texts, indicating that participants spent significantly longer on their first fixation when reading high-difficulty texts compared to low-difficulty texts. First fixation duration reflects the initial understanding and response to words, and reading more difficult vocabulary tends to take longer, as participants require more time for initial cognitive processing (Jian et al., 2013). In this study, proper nouns are considered low-frequency words, which aligns with the "word frequency effect" in computational linguistics. The word frequency effect suggests that readers process high-frequency words more quickly, while low-frequency words demand more processing time (Inhoff & Rayner, 1986). Consequently, when encountering low-frequency words, participants showed significantly increased total fixation time and first fixation duration, reflecting the increased difficulty of processing these words (Inhoff & Rayner, 1986).

Table 5: Post-hoc Comparison Table of First Fixation Count for Proper Nouns Across Different Text Difficulties

SV	*SS*	*df*	*MS*	*F*	η^2	Post-hoc Comparison
Reading Ability	769.30	1	769.30	0.98	0.10	
Error	7101.06	9	789.01			
Text Difficulty	8776.67	2	4388.34	5.52*	0.38	Hard > Easy
Text Difficulty * Reading Ability	537.491	2	268.75	0.34	0.04	
Error	14298.40	18	794.36			

*p<.05

Previous studies have mostly focused on the reading behavior of children or university students and have consistently observed a common phenomenon: when confronted with high-difficulty texts or low-frequency vocabulary, both total fixation time and fixation count increase. However, this study, which focuses on elders, found a similar pattern—elders also exhibited increased total fixation time and fixation count when reading high-difficulty texts. This result suggests that text difficulty significantly impacts readers across different age groups, especially when processing low-frequency proper nouns. This finding extends the current understanding of reading behavior in elders, showing that despite cognitive changes associated with aging, elders still exhibit reading patterns similar to those of younger participants. It highlights that text complexity, particularly the challenge posed by low-frequency vocabulary, affects readers regardless of age, indicating a universal aspect of cognitive processing during reading.

4.1.6 *Analysis Results Based on Sentences*

This study examined the effects of sentence-level text difficulty and its interaction with reading ability on eye-tracking indicators. The results showed that for regression count, first fixation duration, dwell time percentage, total dwell time, and total fixation count, neither reading ability, text difficulty, nor their interaction reached statistical significance. This suggests that in the current sample, these variables may have a weaker impact on eye-tracking behavior, or the sample size may have been insufficient to detect significant differences. Although some indicators approached significance, the overall results did not achieve statistical significance, particularly about text difficulty. Therefore, future research should consider increasing the sample size to further examine whether these variables significantly affect eye-tracking behavior. This would also allow for a more accurate assessment of the potential impact of text difficulty on reading behavior.

4.2 *Analysis Based on Passages*

At the vocabulary and sentence levels, AOI are primarily analyzed using heat maps, which provide localized insights from a micro-level perspective, allowing for a detailed examination of specific regions. In contrast, passages are analyzed from a macro-level perspective, helping researchers gain a comprehensive understanding of how readers process and comprehend the overall structure of the text. This broader approach allows for a more holistic exploration of readers' understanding of the entire language structure and their cognitive processing throughout the reading experience.

To correlate eye-tracking indicators with linguistic features, this study follows the approach of Ozeri-Rotstain and colleagues (2020) by utilizing Pearson correlation analysis. Based on Cohen's (1988) guidelines, the Pearson correlation coefficient r is used to measure the strength of the relationship between two variables. Specifically, an r-value of 0.1 indicates a small correlation, suggesting a weak relationship; an r-value of 0.3 represents a medium correlation, indicating a moderate relationship; and an r-value of 0.5 reflects a large correlation, suggesting a strong relationship with significant practical implications. This approach enables a clear interpretation of how linguistic complexity and cognitive load interact during reading, providing deeper insights into the connections between language features and eye-tracking data.

The passage analysis inputs the text into CRIE to analyze the computational linguistic indicators and extract the participants' eye-tracking metrics for Pearson correlation analysis, to examine the relationship between the computational linguistic indicators and the eye-tracking metrics. According to the statistical analysis results, there is a significant correlation between eye-tracking indicators and linguistic features. Table 6 defines these moderately correlated linguistic features. In terms of negative correlations, the log-transformed average word frequency of the corpus showed significant negative correlations with total fixation time, total fixation count, and total saccade count (*r* = −.30, *r* = −.33, *r* = −.33, respectively). The ratio of noun phrase modifiers also showed a significant negative correlation with total fixation time (*r* = −.31), while the average number of prepositional phrases was negatively correlated with total fixation time, total fixation count, and total saccade count (*r* = −.30, *r* = −.32, *r* = −.32, respectively). On the positive correlation side, the number of intentional words had significant positive correlations with total fixation time, total fixation count, and total saccade count (*r* = .31, *r* = .33, *r* = .33). Additionally, the number of causal conjunctions was positively correlated with total fixation count and total saccade count (*r* = .32, *r* = .32), and the ratio of conjunctions in complex sentences also showed positive correlations with total fixation count and total saccade count (*r* = .33, *r* = .33). These results suggest that there are moderate positive and negative correlations between linguistic features and eye-tracking behavior, reflecting the potential influence of linguistic structure on reading behavior (Ozeri-Rotstain et al., 2020).

Table 6: Correlation Values Between Linguistic Features and Eye-Tracking Indicators

Linguistic Features	Definition/Meaning	Corresponding Eye-Tracking Indicators	Correlation Value
SD of average of word freq. in logcorresponding to external database	The average word frequency in the Academia Sinica Balanced Corpus is calculated logarithmically for all the words in the text, to align with the cognitive processes of word recognition	Dwell Time	-.30
		Fixation Count	-.33
		Saccade Count	-.33
Modifiers per NP	The average occurrence of modifiers in noun phrases. Studies indicate that the longer the modifiers, the more likely they are to cause comprehension difficulties	Dwell Time	-.31
		Fixation Count	-.33
		Saccade Count	-.33
Average propositional phrase	The average number of prepositional phrases per sentence in the entire text.	Fixation Count	-.32
		Saccade Count	-.32
Intentional words	The number of words in the text with an 'intent' meaning, such as 'attempt' or 'force'.	Dwell Time	.31
		Fixation Count	.33
		Saccade Count	.33
Causal conjunctions	The total number of causal conjunctions in the text used to indicate a causal relationship between two sentences, such as 'because,' 'therefore,' and 'so that'.	Fixation Count	.32
		Saccade Count	.32
Conjunction ratio in complex sentences	The proportion of compound sentences in the text that contain conjunctions.	Dwell Time	.31
		Fixation Count	.33
		Saccade Count	.33

This study explores the relationship between linguistic features and reading difficulty by analyzing readers' actual eye-tracking data. By examining the correlation between eye movement patterns and specific linguistic features, such as sentence structure or vocabulary, writers can more accurately assess the difficulty level of their text. These data can help authors gain clearer insights into which linguistic features significantly affect readers' perception of text difficulty, enabling them to make appropriate adjustments to improve the overall readability and comprehension of their articles.

5. Conclusions

The results of the study show that text difficulty has a significant impact on the processing of both non-proper nouns and proper nouns during reading. Specifically, while simple texts are easier to comprehend, participants exhibited a significantly higher number of fixations on these texts compared to medium-difficulty texts. This may be because elders are more willing to carefully scan and verify information when dealing with simpler content. In contrast, for difficult texts, participants spent more total fixation time processing proper nouns, which aligns with the "word frequency effect," indicating that low-frequency vocabulary requires more cognitive resources for processing.

This study also explored the correlation between eye-tracking indicators and linguistic features, particularly the relationship between different linguistic characteristics (such as word frequency and grammatical structure) and reading behavior. The results indicate that the linguistic features listed in Table 6 significantly influence readers' eye movement patterns, suggesting a strong connection between language structure and cognitive processing during reading.

The results indicate of log-transformed frequency of corresponding parent words, noun phrase modifier ratio, and average number of prepositional phrases are negatively correlated with total fixation duration and saccade count. This suggests that higher values for these linguistic indicators are associated with shorter fixation durations and fewer saccades, implying that these linguistic structures may reduce cognitive load. The indicators of intentional vocabulary count, causal connective count, and compound sentence connective ratio are positively correlated with eye-tracking metrics. This indicates that readers allocate more cognitive resources to process these complex grammatical structures and logical connections, resulting in increased fixation and saccade counts. These findings underscore the significant impact of syntactic structures on reading behavior, particularly as cognitive load markedly increases when readers engage with more complex linguistic units.

Overall, these results underscore the profound impact of linguistic features on reading behavior. Simpler language units, such as high-frequency vocabulary and more straightforward grammatical structures, reduce cognitive load for readers, while more complex language units increase processing difficulty. These findings not only support previous research but also provide valuable empirical evidence for future studies on language processing and eye-tracking behavior.

Future research could extend these findings to readers from different age groups or language backgrounds to validate the results of this study. Additionally, further analysis of the impact of passage-level linguistic structures on overall eye movement patterns could provide a more comprehensive understanding of the cognitive processes involved when readers engage with various types of discourse. This has particular significance for the field of education, as it can inform the design of reading materials tailored to the needs of different readers, ultimately enhancing learning outcomes.

6. Limitations

This study provides initial insights into the relationship between eye-tracking indicators and linguistic features, but several limitations should be noted. First, the small sample size, which included only a limited group of elders, may reduce the representativeness of the statistical analysis and impact the reliability and generalizability of the results. This is particularly relevant when considering differences across various age groups, cognitive abilities, educational backgrounds, and reading habits within the elderly population. The small sample size may not fully capture these variations. Therefore, future studies should increase the sample size to enhance the external validity of the findings, ensuring that the observed relationships between eye-tracking indicators and linguistic features are more broadly applicable.

Acknowledgments

This project was financially supported by the "Empower Vocational Education Research Center" of National Taiwan University of Science and Technology (NTUST) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. This project was also be supported by funding from the National Science and Technology Council, Taiwan (111-2410-H-011 -039 -MY2).

References

About Diabetes Common Terms. (n.d.). American Diabetes Association. https://diabetes.org/about-diabetes/common-terms (Accessed September 2024)

Afifah, U. M. R., Wiyono, N., Hastami, Y., & Munawaroh, S. (2008). The Effects of Reading Activity on Episodic Memory in Elderly. *Berkala Kedokteran*, *17*(1), 47-54.

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education, 125*, 413-428.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., ... & ACTIVE Study Group. (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *Jama, 288*(18), 2271-2281.

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review, 24*, 63-88.

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement disorders, 12*(6), 1028-1034.

Bormuth, J. R. (1969). Development of Readability Analysis.

Cambria, J., & Guthrie, J. T. (2010). Motivating and engaging students in reading. *New England Reading Association Journal*, *46*(1), 16-29.

Carpenter, P. A., & Just, M. A. (1983). What Your Eyes Do while Your Mind Is Reading. In Rayner, K. (Ed.), *Eye movements in reading*: *Perceptual and language processes* (pp. 275-307). Academic Press.

Chen, H. C., Lai, H. D., & Chiu, F. C. (2010). Eye tracking technology for learning and education. *Journal of Research in Education Sciences, 55*(4), 39.

Chuang, H. H., & Liu, H. C. (2012). Effects of different multimedia presentations on viewers' information-processing activities measured by eye-tracking technology. *Journal of Science Education and Technology, 21*, 276-286.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Lawrence Erlbaum Associates google scholar, 2*, 263-278.

Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics, 165*(2), 97-135.

Davis, T. C., Wolf, M. S. (2004). Health literacy: implications for family medicine. *Family Medicine*. *36*(8). 595–598.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading research quarterly*, 187-209.

Diabetes Glossary. (n.d.). International Diabetes Federation. https://idf.org/diabetes-glossary/(Accessed September 2024)

Djamasbi, S., Siegel, M., & Tullis, T. (2010). Generation Y, web design, and eye tracking. *International journal of human-computer studies, 68*(5), 307-323.

Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction, 41*, 70-84.

Fanguy, R., Kleen, B., & Soule, L. (2004). Privacy policies: cloze test reveals readability concerns. *Issues in Information Systems, 5*(1), 117-123.

Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology, 32*(3), 221-233.

Fuller, D., & Sedo, D. R. (2014). " And Then We Went to the Brewery": Reading as a Social Activity in a Digital Era. *World Literature Today*, *88*(3), 14-18.

Garcia, A. A., Villagomez, E. T., Brown, S. A., Kouzekanani, K., & Hanis, C. L. (2001). The Starr County Diabetes Education Study: development of the Spanish-language diabetes knowledge questionnaire. *Diabetes care, 24*(1), 16-21.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review, 101*(3), 371.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary educational psychology, 4*(2), 95-123.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of memory and language, 32*(6), 717-742.

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced?. *Psychological science in the public interest, 9*(1), 1-65.

Höffler, T. N., Koć-Januchta, M., & Leutner, D. (2017). More evidence for three types of cognitive style: Validating the object-spatial imagery and verbal questionnaire using eye tracking when learning with texts and pictures. *Applied cognitive psychology, 31*(1), 109-115.

Hu, J., Gruber, K. J., Liu, H., Zhao, H., & Garcia, A. A. (2013). Diabetes knowledge among older adults with diabetes in Beijing, China. *Journal of clinical nursing, 22*(1-2), 51-60.

Hyönä, J., Lorch Jr, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In *The mind's eye* (pp. 313-334). North-Holland.

IDF Diabetes Atlans. (n.d.). International Diabetes Federation. https://www.diabetesatlas.org/data/en/country/194/tw.html (Accessed September 2024)

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & psychophysics, 40*(6), 431-439.

Jian, Y.-C., & Wu, C.-J. (2012). The effect of arrows in an illustration when reading scientific text: Evidence from eye movements and reading tests. *Chinese Journal of Psychology, 54*, 385–402.

Jian, Y. C. (2022). Using an eye tracker to examine the effect of prior knowledge on reading processes while reading a printed scientific text with multiple representations. *International Journal of Science Education, 44*(8), 1209-1229.

Jian, Y. C., & Ko, H. W. (2017). Influences of text difficulty and reading ability on learning illustrated science texts for children: An eye movement study. *Computers & Education, 113*, 263-279.

Jian, Y. C., Su, J. H., & Hsiao, Y. R. (2019). Differentiated processing strategies for science reading among sixth-grade students: Exploration of eye movements using cluster analysis. *Computers & Education, 142*, 103652.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441-480.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 47*(2), 310–339.

Ko, H. W., Chen, M. L., & Liao, C. N. (2005). Frequency Effect, Word Class and Eye Movements: Evidence from Text Reading. *Chinese Journal of Psychology, 47*(4), 381–398.

Kweldju, S. (2015). Neurobiology Research Findings: How the Brain Works during Reading. *Pasaa: Journal of language teaching and learning in Thailand*, *50*, 125-142.

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review, 10*, 90-115.

Lee, Y. H., Chou, Y. T., & Sung, Y. T. (2021). Development and Validation of a Diagnostic Assessment of Chinese Competence System. *Bulletin of Educational Psychology, 53*(2), 285-306.

Liu, Y., Zheng, B., & Zhou, H. (2019). Measuring the difficulty of text translation: The combination of text-focused and translator-oriented approaches. *Target, 31*(1), 125-149.

Lively, B. A., & Pressey, S. L. (1923). A method for measuring the "vocabulary burden of textbooks". in *Educational Administration and Supervision, 9,* (pp. 389-398).

Luyt, B., & Ann, H. S. (2011). Reading, the library, and the elderly: A Singapore case study. *Journal of Librarianship and Information Science*, *43*(4), 204-212.

Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In *Engineering Psychology and Cognitive Ergonomics: 9th International Conference, EPCE 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011. Proceedings 9* (pp. 592-597). Springer Berlin Heidelberg.

Mason, L., Pluchino, P., & Tornatora, M. C. (2013). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly, 48*(2), 199-214.

Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye□tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology, 47*(6), 1083-1095.

McCall, W. A., & Crabbs, L. M. (1926). *Standard Test Lessons in Reading*. (No. 4). Teachers College, Columbia University, Bureau of Publications.

Ministry of Interior Statistical Inquiry Network (2024). https://statis.moi.gov.tw/micst/webMain.aspx?sys=100&funid=defjsp (Accessed April 2024)

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental psychology: Human Perception and performance, 10*(5), 667.

Ozeri-Rotstain, A., Shachaf, I., Farah, R., & Horowitz-Kraus, T. (2020). Relationship between eye-movement patterns, cognitive load, and reading ability in children with reading difficulties. *Journal of psycholinguistic research, 49*, 491-507.

Pearman, A., & Storandt, M. (2005). Self-discipline and self-consciousness predict subjective memory in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(3), P153-P157.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition, 14*(3), 191-201.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Rayner, K. & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ:Lawrence Erlbaum Associates.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging, 21*(3), 448.

Rigg, P. and R. Kazemek. (1983). 'Literacy and adults: what we know and what we need to know'. Educational Gerontology 9 (5 & 6): 417-424.

Rooney, M. K., Santiago, G., Perni, S., Horowitz, D. P., McCall, A. R., Einstein, A. J., ... & Golden, D. W. (2021). Readability of patient education materials from high-impact medical journals: a 20-year analysis. *Journal of patient experience, 8,* 2374373521998847.

Rush, R. T. (1985). Assessing readability: Formulas and alternatives. *The Reading Teacher, 39*(3), 274-283.

Schnotz, W., & Wagner, I. (2018). Construction and elaboration of mental models through strategic conjoint processing of text and pictures. *Journal of Educational Psychology, 110*(6), 850.

Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction, 49*, 218-231.

Shan, H., & Zhou, Q. (2020, October). Research on the correlation between eye movement feature and thematic structure label. In *Proceedings of the 19th Chinese National Conference on Computational Linguistics* (pp. 141-152).

Smith, M. C. (1993). The reading abilities and practices of older adults. *Educational Gerontology: An International Quarterly, 19*(5), 417-432.

Sung, Y. T., Chang, T. H., Lin, W. C., Hsieh, K. S., & Chang, K. E. (2016). CRIE: An automated analyzer for Chinese texts. *Behavior research methods*, *48*, 1238-1251.

Wang, T. N., & Jian, Y. C. (2022). A Systematic Review of Eye-Tracking Studies on Text-Diagram Science Reading. *Bulletin of Educational Psychology, 53*(4), 773-800.

Wolf, R. (1977). 'What is reading good for? Perspectives from senior citizens'. *Journal of Reading 21*(1): 15-17.

Zhang, W., Zhang, Y., & Wang, J. (2022). Effect and mechanism of reading habits on physical and mental health among the elderly: evidence from China. *Frontiers in Public Health*, *10*, 1031939.