English Teaching, Vol. 79, No. 2, Summer 2024, pp. 33-55
DOI: https://doi.org/10.15858/engtea.79.2.202406.33
http://journal.kate.or.kr

The Cognitive Process Involved in Young EFL Learners' English Word Recognition: An Eye-Tracking Study*

Suh Keong Kwon**

Kwon, Suh Keong. (2024). The cognitive process involved in young EFL learners' English word recognition: An eye-tracking study. *English Teaching, 79*(2), 33-55.

This paper investigates the cognitive processes involved in English word recognition among young EFL learners using eye-tracking methodology. A quasi-experimental mixed method design was used to investigate how young L2 learners engage with basic words, with or without pictorial cues. A total of seventeen 6th-grade pupils from two schools participated in the experiment. The participants were presented with a list of 20 words and were asked to read them aloud while their eye movements were tracked to discern their viewing patterns. Immediately after the reading task, stimulated-recall interviews were conducted to triangulate and validate the participants' viewing behaviors. Results indicate that participants focused significantly more on the text than the accompanying pictures yet demonstrated better performance in recognizing and reading the words presented in a picture-based mode. Some participants reported that the pictures were not viewed because the words were easy to read. In contrast, others struggled to read certain words due to an over-reliance on their background knowledge, which sometimes led to misinterpretation. These results emphasize the importance of integrating visual cues with word recognition instruction in early language learning contexts, highlighting when and how these cues should be utilized effectively.

Key words: word recognition, basic reading skills, multimodal language instruction, eye tracking

*This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A8081230).

**Author: Suh Keong Kwon, Professor, Department of English Education, Chinju National University of Education; 365 Bungil 3, Jinju, Gyeongsangnam-do, Republic of Korea; Email: sokekwon@cue.ac.kr

Received 31 March 2024; Reviewed 22 April 2024; Accepted 1 June 2024

1. INTRODUCTION

Word recognition has been deemed to be of importance for fundamental reading skills. The key elements required for reading include phonemic awareness, reading fluency, vocabulary knowledge, and text comprehension (Neuman, Copple, & Bredekamp, 1998). Among the many important elements dealing with fundamental reading skills, a lack of phonemic awareness skill has long been pointed out as a common phenomenon among young English as a Foreign Language (EFL) learners. A number of studies focusing on English underachievers in South Korea suggest that a common cause of their underperformance in English is a deficiency in fundamental reading skills (Kim & Kwon, 2023; Kwon & Kim, 2021).

One of the commonly used instructional methods for developing fundamental reading skills for early language learners is the teaching of sight words (Helman & Burns, 2008). This approach involves introducing young learners to frequently exposed essential words, allowing them to practice reading them during the early learning stages. *Sight word recognition* was first introduced by Dolch (1941), who defined it as high-frequency words used in text that early language learners are encouraged to learn or recognize. He introduced 'The Dolch Word List', which contains 220 'service words' and 95 'high-frequency nouns.' The list, consisting of 315 words, is now categorized into different grade levels, from pre-kindergarten to third grade. It is yet debatable whether recognizing sight words is sufficient for developing young EFL learners' basic reading skills. In fact, according to the 2022 Revised National English Curriculum, the number of basic words listed at the primary school level is 800, which is much more than the existing sight word list (Ministry of Education, 2022). Although enough consensus has been established on the importance of teaching and learning basic words, little is yet known about what strategic methods for teaching high frequency basic words are effective for young EFL learners, and how they process them. With the multimodal demands of engaging with digital technologies introduced today, the present study seeks to explore innovative strategic approaches that can be employed from an early age to teach fundamental reading skills. Specifically, the aim of this study is to investigate the effectiveness of incorporating basic words and pictorial cues into the teaching and learning process.

2. LITERATURE REVIEW

2.1. Word Recognition

Word recognition is known as "the most basic and most critical process in reading"

(Chikamatsu, 2006, p. 67). Gough and Tunmer (1986) argued that successful reading occurs when word recognition and language comprehension are both developed. In the context of L2 language reading, it has been posited that possessing effective visual word recognition skills is essential (Kida, 2016), a concept initially introduced by Ehri (1998) in the pre-alphabetic phase.

Studies have suggested that word recognition skills are central to developing reading skills and are a critical ability to ensure proficient reading (Ehri, 1998; Grabe, 1991). Word recognition is associated with sub-components such as 'letter knowledge,' 'phonological awareness,' and 'decoding.' Letter fluency is defined as knowledge of letter names, which enables learners to get the gist of the sounds of the letters (Carroll, 2000). Phonological awareness is associated with the learner's ability to recognize the sound of phonemes, which is essential in reading texts. In terms of learning English in the Korean context, a study investigated 114 Korean elementary school students' phonological awareness and revealed that it could significantly predict their reading skills (Kang, 2009). Decoding is a strategy that supports learners in reading unknown words, and it involves identifying the sounds of the presented graphemes and blending them into the pronunciations of words. Decoding simple vowel and consonant sounds enables readers to recognize words, and learners often rely on spelling-sound cues to try reading unfamiliar words as well. According to Ehri's (1998) four phases of reading development, such decoding starts to occur in the partial-alphabetic phase, where learners recognize the written language symbols and gain letter-sound skills. In general, proficient reading involves the automatic decoding of words with small or no cognitive efforts required (National Institute of Child Health and Human Development, 2000). It was suggested that automatic word recognition is a fundamental skill required by early language learners to obtain an automatic decoding skill, which is as important as vocabulary knowledge (Bernhardt, 2000).

Systematic phonics instruction has been proven to be an effective way to develop decoding strategies to facilitate basic literacy skills (Ehri, Nunes, Stahl, & Willows, 2001). In the Korean context, Lee (2017) asserted that students in an EFL context need to develop decoding strategies to manage the complexity of English spelling sounds to achieve success in learning to read. One effective strategy might be to provide visual input that gives semantic cues to facilitate one's word recognition skill and allows learners to identify the words successfully. However, little is yet known about what specific cognitive processes are involved in such word recognition processes, particularly when visual cues are presented together.

2.2. Multimodal Resources and Dual Coding Theory

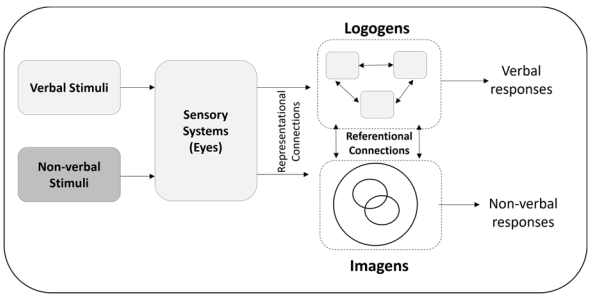
The present study hypothesizes that multimodal input facilitates young EFL learners'

word recognition performance. This hypothesis draws upon Mayer's (2009) cognitive theory of multimedia learning and Paivio's (1990) dual coding theory, both of which explain the ways in which multimodal input functions in our cognitive system. Mayer (2009) introduces the concept of a dual channel, which is one of the three theory-based assumptions on the cognitive processes of learning through words and visuals. The dual-channel assumption posits the existence of two separate channels that process and manipulate received information: the verbal channel and the visual-pictorial channel (Paivio, 1990). Mayer (2002) adds that when these multimodal presentations are processed in our cognitive system, learners mentally arrange the received words and visual cues together into a "coherent mental representation in their working memory" (p. 60). He provides a significant observation that multimodal presentations enhance learners' comprehension by enabling them to mentally integrate both visual and verbal explanations, thereby facilitating the learning process. Hence, the overarching aim of delivering multimedia presentations to learners is to foster the process of learning.

Paivio's (1990) dual coding theory explains that "there are two classes of phenomena handled cognitively by separate subsystems" (p. 3). Specifically, one class of phenomena pertains to processing language, while the other class concerns processing non-verbal components. He argues that although each subsystem, namely logogens and imagens, functions independently in our cognitive system, they are referentially interconnected by initiating each other. In other words, both verbal and non-verbal components are triggered by each other in our cognitive system to process multimodal input into a coherent meaning. The following Figure 1 provides more intuitive illustrations of these dual coding concepts.

FIGURE 1

Dual Coding Theory (Paivio, 1990)



According to Lotherington and Ronda (2012), multimodality has emerged as a "concomitant of human communication" and is now extensively integrated into language teaching and learning" (p. 107). Language teachers today need to acknowledge visual input as one of the essential components of multimodal communication. In this regard, existing

studies have examined the potential benefits of engaging visual cues in helping EFL learners' language acquisition (Alghonaim, 2019; Chang, 2011; Mestres & Pellicer-Sánchez, 2020; Park & Lee, 2022; Tragant Mestres, Baró, & Garriga, 2019; Tragant Mestres & Vallbona, 2018). However, it should be noted that most of these studies examined the cognitive processes and perceptions found in reading-while-listening tasks. Only one study from the above examined learners' eye movements using a research method such as eye tracking (Mestres & Pellicer-Sánchez, 2020). To my knowledge, no studies have examined the cognitive processes involved in the viewing behavior of young EFL learners in recognizing basic words under both text-only and visual-supported conditions. Based on these research gaps found in the existing studies, this study addresses the following research questions:

RQ 1. To what extent do young EFL learners benefit from viewing pictorial cues when recognizing and reading basic English words?
RQ 2. How does the viewing behavior of young EFL learners differ when picture cards are provided compared to when they are not?
RQ 3. To what extent do young EFL learners rely on pictorial cues when recognizing and reading basic words?

3. METHODOLOGY

The present study employed Creswell and Creswell's (2022) convergent mixed methods research design. The findings for both the quantitative and the qualitative analyses were merged by comparing and relating them. By doing so, a more comprehensive understanding was gained and an overall interpretation of the findings to answer each research question was possible.

3.1. Participants

The participants of this study were 17 elementary school students in the 6th grade selected from two distinct schools. School A, situated in a mid-sized city, has an enrollment exceeding 1,000 students, and class sizes typically range from 25 to 30 students. School B, on the other hand, is a small school in a rural agricultural area. The selection of these two groups of participants were done through purposive sampling technique to ensure a balanced representation across various environment, including mid-to-large cities and rural areas. The participants exhibited varying levels of English proficiency, influenced predominantly by their experiences in learning English outside the traditional classroom setting. None of the students had prior exposure to living or studying abroad before they participated in this study.

3.2. Instruments

3.2.1. Word list and picture cards

To select the words for testing the participants' fundamental English reading skills, the author used the basic word list of the National English curriculum (2022 revision). From the 800 words recommended for the elementary level, 20 words were selected under three key standards: 1) words that start with each different letters, 2) words with various phonics elements (short/long vowel, consonant, vowel digraphs, blend, etc.), and 3) words that are easily identifiable through a visualized form. Considering the aims of this study, it was required to look at how participants identify the meanings and sounds of words that are frequently taught and used in early English learning classrooms. During the selection stage, words starting with letters 'I', 'Q', 'U', 'V', and 'Y' were excluded because there were no suitable words from the basic word list that met the three conditions. As a result, a total of 20 words were selected and reviewed by a head elementary teacher before finalizing the list.

A generative artificial intelligence tool, *Adobe's Firefly*, was used to minimize the researcher's subjective preference in making a picture-based word list. When each word was entered into *Firefly*, four candidate images were generated, and the researcher selected the most clearly identifiable image from them. The list of 20 words is presented in Appendix.

3.2.2. Eye tracker

For the eye-tracking experiment, the Tobii Nano eye tracker was used, which has a 60 Hz sampling rate. This sampling rate can capture a participant's eye movements 60 times per second, which is fast enough to record and identify their viewing behavior of reading large texts and images that are used in this study. For eye-movement data analysis, *Tobii Pro Lab* (v.1.181) was used.

3.3. Data Collection

The data collection was carried out at the participants' schools during school hours. The eye-tracking device was set up in an isolated, eclosed space in each school. In order to keep a consistent testing environment across the two schools, the sizes of the desks and chairs and the room lights were carefully considered. Each participant was tested individually, and each test was approximately 10 minutes long.

3.3.1. Eye-tracking test

The eye-tracking test started with a calibration stage. In this stage, the participants' eye movements were calibrated with the eye tracker. Re-calibration was carried out with a few participants who showed a data loss rate of more than 10%. As a result, the participants' gaze sample rate was 91.7% on average, which is a highly acceptable level. The test began with the 20-word list shown to the participants automatically. The same word list was projected on a screen twice, once with a picture and once with a word only. The participants were asked to read aloud the words that were shown on the screen and their meanings in Korean. Of the 17 participants, 8 of them received the picture-based word list first and then the text-only list. The remaining 9 participants received the text-only list first and then the picture-based word list. As such, the order of this condition was counterbalanced to minimize the order effect. The following Figure 2 illustrates the counter-balancing method:

FIGURE 2

Counterbalancing of the Data Collection Phase

Group 1 (n = 8)	**Picture-based**	→	**Text-only**
Group 2 (n = 9)	**Text-only**	→	**Picture-based**

3.3.2. Stimulated-recall interviews

Immediately after the eye-tracking word-reading test, all participants were involved in a retrospective stimulated-recall interview. For this interview, the researcher showed the visualized gaze plot and heat map of each participant's eye movement and asked questions about their word-reading processes. Specifically, semi-structured interview questions were used, which consisted of three key questions: 1) Where and why did you look at _____? 2) What were you thinking when identifying/reading this word? 3) Was looking at _____ (either picture or word) helpful or distracting in reading the word? In addition to these semi-structured questions, more questions were posed when interesting eye movements or reading behaviors were found during the test. For this, the researcher used the field notes kept during the test and also the participants' visualized eye movements. The duration of each interview ranged from three to seven minutes, varying based on the participants' reading performance and viewing behavior observed during the test.

3.4. Data Analysis

In order to examine the degree to which the presence of pictures might have an impact on their test scores, the test scores of both groups were compared. Based on the normality test

results, non-parametric Wilcoxon Signed Rank tests were conducted.

For eye-movement analysis, the areas of interest (AOI, hereafter) were defined first, which is an essential step in designing an eye-tracking study (Holmqvist et al., 2011). The present study defined two AOIs in total: Picture and Word. The sizes of picture AOIs were equivalent across all 20 items. Figure 3 below illustrates how the AOIs were defined on the test screen.

FIGURE 3

Sample Screenshots of the AOI Design

For eye-movement analyses, total fixation duration and fixation count measures were used. The former measures the cumulative duration of viewing the AOIs (also known as 'dwell'), and the latter measures the frequency of viewing the AOIs. Usually, average fixation duration for silent reading is 225 milliseconds and for visual search is 275 milliseconds (Rayner, 1998). The present study compares the degree to which participants view the picture and the text differently in terms of length of time and frequency. For this, non-parametric Wilcoxon Signed Rank tests were computed, followed by a normality test.

For a qualitative analysis of the participants' viewing behavior, stimulated-recall interview data were first transcribed using *Naver Clova Note*. Then, the author cross-checked the recordings of the interviews and the initial transcriptions and edited the final draft transcript. All interview data were entered into *Microsoft Excel*, and a thematic analysis was carried out to capture common patterns found in the participants' viewing behaviors and the rationales behind them (Braun & Clarke, 2006). The coding of each segment was carried out based on the three semi-structured interview questions. During the initial coding process, key chunks of data were labeled with three main themes – 1) reasons for viewing text/image, 2) reasons for failing to read, 3) general perceptions of having the image in reading the text – and inferential sub-themes were coded for further analysis. The following Table 1 presents the structure of the thematic analysis.

TABLE 1

Themes and Sub-themes of Qualitative Findings

Themes	Sub-themes
1. Reasons for viewing text/image	1.1 Ability to read words
	1.2 Reliance on background knowledge
	1.3 Helpfulness of visual cues
2. Reasons for failing to read the word	2.1 Difficulty of reading words (Phonics, Capitalization)
	2.2 Reliance on background knowledge
3. General perceptions of having the image for word reading	3.1 Beneficial
	3.2 Confusing

The frequency of each theme and sub-themes found in the data were calculated to demonstrate the overarching trends of the findings. Also, excerpts from the participants' interview data were translated into English, and key examples were carefully selected to present qualitative findings.

4. RESULTS

4.1. Test Score Difference

Descriptive statistics presented in Table 2 below show that the mean test score of the picture-based test was 18.65, and the mean test score of the text-only test was 16.24. These results indicate that the participants performed better in reading the words when the picture was present. In order to examine whether this result was statistically significant, inferential statistics were carried out.

TABLE 2

Test Score Differences Between the Picture-based and the Text-only Conditions (*N*=17)

Picture-based				Text-only			
Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
18.65	1.69	15	20	16.24	4.27	5	20

According to the Shapiro-Wilk normality test, the scores of the picture-based condition were not normally distributed ($p < .001$). Therefore, the Wilcoxon Signed Rank test was conducted, and the results showed that the score difference between the two conditions was statistically significant: $z(16) = 2.70$, $p = .007$. Therefore, it can be confirmed that the participants performed significantly better when there was a picture provided with the word than when there was text only.

4.2. Eye-Movement Difference

To further investigate the extent to which the participants' viewing behavior might differ when picture cards are provided compared to when they are not, eye-movement data were analyzed. Descriptive statistics presented in Table 3 indicate the total fixation duration on the text AOI.

TABLE 3

Total Fixation Duration on Text: Picture-based vs. Text-Only (*N*=17)

Unit: Seconds

	Picture-based		Text-only		*z*	*p*
	Mean	S.D.	Mean	S.D.		
Entire test	47.77	2.39	63.94	15.39	-3.43	<.001
Per each item	2.39	0.83	3.20	0.77		

The mean total fixation duration of the text in the picture-based condition was 47.77 seconds and 63.94 seconds in the text-only condition. In other words, the participants viewed the text longer when there was no picture than when the picture was presented alongside the text. Followed by the normality test, the Wilcoxon-Signed Rank test results showed that the mean difference between the two conditions was statistically significant ($p < .001$).

The following Table 4 shows the total fixation duration of the picture and the text in the picture-based test.

TABLE 4

Total Fixation Duration: Picture vs. Text (*N*=17)

Unit: Seconds

	Picture		Text		*z*	*p*
	Mean	S.D.	Mean	S.D.		
Entire Test	17.60	12.49	47.77	2.39	-3.24	<.001
Per item	0.88	0.625	2.39	0.83		

Descriptive statistics show that the participants viewed the picture for 17.60 seconds on average and viewed the text for 47.77 seconds on average for the entire test. When looking at the duration for each item, the participants looked at the picture for 0.88 seconds on average and the text for 2.39 seconds on average. In summary, the participants viewed the text substantially longer than the picture, and this was statistically significant ($p < .001$).

Descriptive statistics presented in Table 5 below present the fixation count on the picture and the text in the picture-based test.

TABLE 5

Fixation Count on the Picture and the Text (*N*=17)

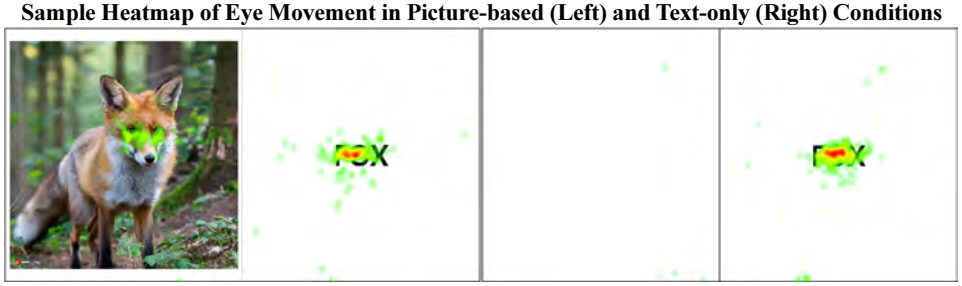
	Picture		Text		*z*	*p*
	Mean	S.D.	Mean	S.D.		
Entire Test	81.12	54.75	188.24	54.99	-3.05	< .01
Per item	4.06	2.74	9.41	2.75		

The result shows show that the participants viewed the picture 81.12 times on average and viewed the text 188.24 times on average for the entire test. When the fixation count data for each item was analyzed, the participants looked at the picture 4.06 times on average and the text 9.41 times on average. As the Shapiro-Wilk normality test result showed that the data were not normally distributed ($p < .05$), the Wilcoxon Signed Rank test was computed. The results showed that the fixation count difference between the Picture and the Text AOIs was statistically significant: $z(16) = -3.05$, $p < .01$. Thus, the participants viewed the text significantly more frequently than the picture.

The following Figure 4 shows a heatmap of participant's eye movements that illustrates how much the participants viewed the text and the picture during the test. It should be noted that all fixations are indicated in green, and the longer fixations are indicated in red.

FIGURE 4

Sample Heatmap of Eye Movement in Picture-based (Left) and Text-only (Right) Conditions



4.3. Results from the Stimulated-Recall Interviews

Table 6 below shows the number of references identified in the participants' interviews for each theme and sub-theme.

TABLE 6

Number of References for Each Theme and Sub-theme

Theme/ sub-theme	Sub-themes	No. of references	No. of participants who made references
1. Rationales for viewing text/image	1.1 Ability to read words	14	8
	1.2 Reliance on background knowledge	11	6
	1.3 Helpfulness of visual cues	10	10
2. Reasons for failing to read the word	2.1 Difficulty of reading words (Phonics, Capitalization)	6	5
	2.2 Reliance on background knowledge	9	7
3. Other perceptions	3.1 Novel, no reason	4	4
	3.2 Confusing	4	4

According to Table 6, the most frequently mentioned theme was "rationales for viewing either the text or the images." This theme is directly relevant to the first semi-structured question and was asked to all participants. Responses to the question related to Theme 1 varied depending on the participant's viewing behaviors; that is, whether or not they viewed the text or the images during the test. A total of eight participants spoke of their ability to read the words with or without the images, and six participants mentioned their reliance on their background knowledge while viewing the text or the images. Also, ten participants reported that they viewed the images because they found them helpful. The second theme, "reasons for failing to read the word," is closely related to the rationales for failing to read the words during the test. Those who were not able to read some of the words during the test were asked to report the reasons why they could not read them during the stimulated recall interviews. A total of five participants reported that they couldn't read the words because of their lack of English reading skills. Specifically, the lack of English reading skills found among these participants was related to their phonics knowledge or ability to identify capitalized words. In addition, seven participants reported that they relied on viewing the images to recall their background knowledge instead of reading the words. Four students responded that they viewed the text or the images because they were novel or for no reason. Four participants also reported that the two different ways of presenting the words (with and without the images) were confusing.

4.3.1. Rationales behind the viewing behavior

The first theme deals with the rationales behind viewing behavior. During the stimulated-recall interviews, all participants were asked to report on why they viewed or did not view the images or the text. According to the eye-movement analysis, it was found that the participants viewed the text more than the images. Through the stimulated-recall interviews, the rationales behind the participants' viewing behavior could be better understood.

The main reason why the participants viewed the text substantially longer than the image was because they were able to read the sight words without any help from the visual cues. The following excerpts are examples:

> Interviewer: You looked at the text "apple" much longer than the picture. Why did you do that?
> Participant: Because I know the word. I know how to read the word.
> (Participant KJ23)

> Interviewer: You looked at the images just once, but you read the words most of the time. Any reason for this?
> Participant: They were easy (to read).
> Interviewer: How did you find reading the words when there were no images? Like this one.
> Participant: I read the words that I knew.
> (Participant KJ12)

Next, when participants were asked why they viewed the images, ten participants reported that they found the images to be helpful. The following is what participants mentioned during the interview:

> Interviewer: You didn't look at the text a lot for this word. How did you find it when there was a picture versus when there was no picture?
> Participant: I found it easier with the picture.
> (Participant KJ14)

> Interviewer: How was it when there was no picture?
> Participant: When there was the word rabbit, the same rabbit as last time (with picture), I couldn't read it because there was no picture. So, I read this word by looking at the picture.
> Interviewer: I see. That's why you were only able to read the first sound, "rab," when there was no picture.
> (Participant KC02)

Another reason for viewing the images was associated with the reliance on their background knowledge. A total of eleven references were made from six participants who reported that they relied on their knowledge of the vocabulary that was stimulated by the images shown to them.

Interviewer: So, you are saying that looking at these images helped you read the word? Tell me, what is this (image of a carrot) in English?
Participant: Carrot.
Interviewer: Right. But you couldn't read it when there was no image. You also didn't look at the image much. Were you able to read the word, but you were just confused?
Participants: No, I didn't know how to read the word. But I looked at the picture and said, "Carrot."

(Participant KJ11)

The participant KJ11 above reported that he was unable to read the word without the carrot image. However, the eye-movement data of the participant showed that he only looked at the picture twice for a short period of time (See Figure 5 below). Nevertheless, KJ11 reported that he was able to read the word by looking at the image. This finding means that he did not "read" the word aloud but rather used his background knowledge about the name of the vegetable.

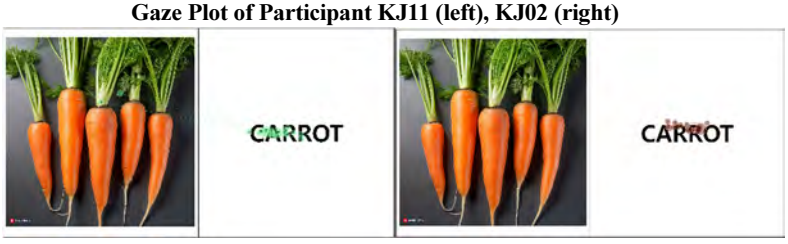
There were a few other cases where self-reported data added more specific information to the eye movements in terms of the degree to which they viewed the images. It was found that many participants used their peripheral vision to "peek" at the images rather than fully focusing on them using their foveal vision. The following excerpt is an example from a participant.

Interviewer: But you were not looking here (at the image).
Participant: I wasn't paying attention to it, but I still peeked at it using a sideways glance. I can still see what it is.
Interviewer: So you just didn't pay attention to the image but still knew that there was a carrot over here.

(Participant KJ02)

FIGURE 5

Gaze Plot of Participant KJ11 (left), KJ02 (right)



As presented above, the cross-checking of the findings between the eye-movement data and the stimulated recall interview data was carried out to validate the findings.

4.3.2. Rationales for failing to read words

Based on the field notes, those who failed to read certain words were asked about the reasons why they failed to read the words aloud. There were largely two reasons: a lack of English reading skills and a reliance on background knowledge. The following excerpt is one example of a participant who lacked English reading skills.

> Interviewer: And next, when you read this one (TAIL), this is how you looked. When there was a picture, you only looked at the letters *t* and *a*. When there was no picture, you read the word as "tell." Why don't you try reading it again now?
> Student KJ11: Terry? Telli? Terry.
> Interviewer: What do you think it means?
> Student KJ11: *Kko-ri*. (meaning tail in Korean).
> Interviewer: That's right. It's 'tail.' So, you knew it started with a 't' sound but didn't know how to read the word to the end.
>
> (Participant KJ11)

> Interviewer: Why do you think you couldn't read this word (PAPER) when there was no picture?
> Student KJ12: I got confused because the word was in all capital letters.
> Interviewer: I see. But when there was this picture, then you were confident enough to say 'paper.'
> Student KJ12: Yes.
>
> (Participant KJ12)

The following excerpt is one example of a case where the participants' use of their background knowledge distracted them from reading the words.

> Interviewer: How about this one with a person kicking a ball? What do you think the word is?
> Student KC02: Kick.
> Interviewer: That's correct. Why couldn't you read it a few minutes before?
> Student KC02: I thought the word was difficult.
> Interviewer: Wasn't the picture helpful?

Student KC02: I thought, soccer? And I just couldn't read it after all.
(Participant KC02)

The participant KC02 reported that he could not read the word 'kick' when there was text only. In other words, the participant did not have knowledge of the vocabulary word, but instead used his background knowledge to guess what the word was. The participant first came up with the word 'soccer' but later changed it to 'kick.' It can be speculated that the participant guessed any relevant words that came to his mind with the help received from the pictorial cue. The next excerpt is an example of a participant who did not rely on the pictorial cue to read the word.

Interviewer: When there was no picture, you read this word as 'Rabbly.'
Student KC03: I got confused with something else. I didn't know it was a rabbit.
Interviewer: Then what did you think it was when there was only the text?
Student KC03: I just questioned what word it might be…
Interviewer: As you can see here (picture-based test screen), you didn't really look at the picture of the rabbit. But you still read it as 'rabbit' when this picture was given. Do you remember how you read this?
Student KC03: I didn't look at it (the picture).
(Participant KC03)

The participant above failed to read the word 'rabbit' in the text-only condition but read it properly in the picture-based condition. The participant reported that he did not look at the picture of the rabbit to read the text, but the eye-movement data showed that he looked at the picture three times. It can be assumed that the participant looked at the picture unconsciously and did not recall any memory of it or receive clues from it. However, the short glances at the picture may have helped the participant to come up with the word in the picture-based condition.

4.3.3. Perceptions of viewing the images with the words

Most participants favored to have the picture next to the word. Regardless of whether they need them to read the word or not, showing a picture led to positive perceptions from the participants. The following is one of the excerpts from an interview:

Interviewer: When you read this word ('RABBIT'), you seem to hesitate a little. You read it as 'robit.'
Student KC01: I looked at the picture.

Interviewer: You looked at the picture of the rabbit for quite a while. Why did you do so?

Student KC01: I just found it interesting. I don't know how to say that (the rabbit) in English.

(Participant KC01)

During the test, a noticeable hesitation was found when participant KC01 was reading the word "rabbit." Unlike how it was observed during the test, the participant reported that he looked at the picture more than the target word and recognized what it was, but did not have the vocabulary knowledge to say the word in English. Also, the participant looked at the picture for a relatively long period of time, not because he was trying to decode the word using his background knowledge but because the picture seemed interesting. Therefore, language teachers need to identify what difficulties learners have when recognizing words, even when pictures are provided together.

5. DISCUSSION

5.1. Key Findings and Interpretations

The present study revealed that the participants performed better in recognizing and reading basic words when the pictures were present. This finding answers Research Question 1, "To what extent do young EFL learners benefit from viewing pictorial cues when recognizing and reading basic English words?" It can be argued that the presence of pictorial cues aids EFL learners to read the words successfully. Regardless of whether they can 'understand' the meaning of the word or not, it can be asserted that presenting the pictures of the given words is helpful in identifying and reading the word. However, a careful interpretation is required as this assertion does not necessarily indicate their phonics ability. Instead, it may be their background knowledge facilitated by the pictures presented. In fact, during the stimulated-recall interview, several participants (6 out of 17) reported that they relied on their background knowledge facilitated by pictorial cues when reading aloud the word.

In answer to Research Question 2 "How does viewing behavior of young EFL learners differ when pictures are provided compared to when they are not?" the participants' eye movements were analyzed. It was found that the participants spent significantly longer amounts of time viewing the text and tended to view it more frequently compared to the pictures. To a certain extent, this result contradicts the finding that the presence of pictorial cues was found to be beneficial. However, it should be acknowledged that the eye tracker

can only capture the test-takers' foveal vision (the core area where eyes are directed). In the stimulated recall interview data, the researcher found that some participants actually peeked at the visual cues while reading the word using their peripheral vision. Also, looking at a picture to identify what it is does not take a long period of time compared to reading a word without a picture (Rayner, 1998). Therefore, it can be speculated that the longer fixation duration and more frequent fixations on the text over visual cues are associated with the cognitive load required to process the word and the picture.

The findings of the stimulated recall interviews shed further light on the specific processes and rationales behind the eye movements shown by the participants. In answer to Research Question 3 "To what extent do young EFL learners rely on pictorial cues when recognizing and reading basic words?" a thematic analysis was conducted on the transcribed qualitative data. It was found that the participants who did not rely on the pictorial cues were those who were capable of reading the words without any aids. On the other hand, those who relied on the pictorial cues reported that they often relied on their background knowledge of the objects that were shown in the pictures. Hence, it was not their reading ability that helped them successfully read the word, but they used their knowledge about the word based on its image. Some participants failed to read the word either because of their lack of English reading skills or because their background knowledge misled them. Some participants lacked both an English phonics skill and vocabulary knowledge, which therefore made them unable to read the word with or without the presence of a pictorial cue. Also, some participants were misled by the pictorial cues provided; for example, they read the word "house" as "home" by looking at the picture only. From these findings, it can be asserted that there is a need for more sophisticated basic reading tests that can examine early EFL learners' reading skills more validly.

5.2. Limitations

The limitations of this study are twofold. First, since the eye-movement data only provide information about the test-takers' foveal vision (the visual area where test-takers' eyes are directed) and not their peripheral vision (the area surrounding the main foveal vision), the findings of the participants' viewing behaviors need careful interpretation. For this reason, the present study triangulated by adding stimulated-recall interviews and identified that there was some missing information about the test-takers' cognitive processes that were not captured by the eye-tracking method. For example, there were several instances where participants 'peeked' at the pictures while reading the words. Still, such viewing behaviors were not taken into account when eye movements were analyzed quantitatively. However, Rayner and Pollatsek (2006) state that speed and fluidity of reading are not indicators of reading fluency, but rather, the duration spent looking at each word in the text is a more

tangible measure. Hence, the short 'peeking' of the pictures may not have distracted the learners from reading the word carefully. Second, the small sample size with homogeneous characteristics restricts the generalizability of the findings of this study. Inevitably, non-parametric test was used due to non-normal distribution of the data, but different findings could have been found if parametric test with a larger sample was conducted. Nevertheless, it should be noted that sampling a large number of participants in an eye-tracking study is practically difficult, considering that a retrospective stimulated recall interview is an essential step in triangulating the data. The findings of this study have been effectively cross-checked, providing valuable insights into participants' reading processes and viewing behaviors when reading basic words with and without the help of pictorial cues.

5.3. Implications and Conclusion

The findings of this study provide credible evidence to support Mayer's (2009) cognitive theory of multimedia learning and Paivio's (1990) dual coding theory. It was found that presenting pictures with words aids young EFL learners in activating their pictorial channel to use their background knowledge in order to recall the word more effectively. However, this finding does not mean that their word recognition is improved because technically, it cannot be considered as 'reading' the word. Instead, it simply helps those who are confused about how to read the word. Therefore, systematic phonics instruction must be prioritized as the initial step in developing phoneme awareness and decoding strategy (Ehri et al., 2001). Visual cues can be used as a supplement to support learners in recognizing the words that they were already exposed to and learned to read. This recognition can be one of the effective decoding strategies that Lee (2017) has called for.

Based on the findings revealed from this study, three key implications can be asserted. Firstly, integrating pictorial cues alongside words can enhance young EFL learners' word recognition skills. However, careful approach is also required as some learners may guess and 'pretend to read' using the pictorial cues. Therefore, teachers need to utilize visual aids strategically to scaffold reading tasks. While pictorial cues may not aid in improving learners' phonics skills, as found in this study, they can still serve as effective aids in word identification. Therefore, it is important to carefully select and present images that align with the learners' background knowledge to facilitate their reading skills.

Secondly, it is recommended that EFL teachers balance the cognitive load of reading activities. To achieve this balance, a better understanding of the cognitive load involved in processing both text and visual cues is needed for various reading-while-viewing activities. While participants in this study showed longer total fixation durations on text than on pictures, it should be acknowledged that this finding may just be attributed to the cognitive effort required to decode words. With this finding in mind, teachers need to carefully

consider the cognitive load of young EFL learners when designing reading tasks in order to strike a balance between text and visual stimuli. This balance can be achieved by controlling the complexity of texts and relying on pictorial cues.

Thirdly, it is necessary to develop and employ a variety of assessment methods to evaluate young EFL learners' basic reading skills more accurately. Currently, the most commonly used basic English reading test in public schools of the Republic of Korea is the *System for Diagnosis and Correction on Basic Academic Skills.* This test is developed and administered by the Center for Supporting Basic Academic Skills of each province. However, schools in Korea rely too heavily on this one single test for diagnosing learners' basic English skills. Another problem with this test is that it can only examine a learner's reading ability indirectly. For this reason, the test may not fully capture learners' true reading abilities, especially when test-takers rely solely on pictorial cues or when background knowledge influences word identification. Instead, more authentic tasks, such as performance-based tasks and observational assessments should be developed and implemented to gain insights into learners' reading processes and strategies. By using diverse assessment methods, teachers can better identify individual learners' strengths and areas for improvement, leading to more targeted instructional interventions.

Applicable levels: Early childhood, elementary, secondary

REFERENCES

Alghonaim, A. S. (2019). Impact of watching TV English cartoons on EFL language acquisition: A case study of an Arabic young child. *Arab Journal for Scientific Publishing*, *10*(2), 23-54.

Bernhardt, E. B. (2000). Second-language reading as a case study of reading scholarship in the 20th century. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research* (vol. 3, pp. 791-811). New York: Routledge

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101

Chang, A. C. S. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching, 21*, 43-64.

Carroll, J. B. (2000). The analysis of reading instruction: Perspectives from psychology and linguistics. *Scientific Studies of Reading, 4*(1), 3-17.

Chikamatsu, N. (2006). Developmental word recognition: A study of L1 English readers of L2 Japanese. *The Modern Language Journal, 90*(1), 67-85.

Creswell, J. D., & Creswell, J. W. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: Sage.

Dolch, E. W. (1941). *Teaching primary reading*. Champaign, IL: Garrard Press.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). New York: Routledge.

Ehri, L., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*, 393-447.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25*(3), 375-406.

Helman, L. A., & Burns, M. K. (2008), What does oral language have to do with it? Helping young English-language learners acquire a sight word vocabulary. *The Reading Teacher, 62,* 14-19.

Kang, Y. (2009). The role of phonological awareness in Korean elementary EFL learners' word reading. *English Teaching, 64*(2), 29-45.

Kida, S. (2016). Automatization and orthographic development in second language visual word recognition. *Reading in a Foreign Language 28*(1), 43-62.

Kim, D., & Kwon, S. (2023). Action research of written language instruction methods for elementary English underachievers. *Primary English Education, 29*(2), 23-46.

Kwon, S., & Kim, T. (2021). The factors and processes of becoming an underachiever in learning English: A longitudinal case study. *Primary English Education, 27*(3), 45-79.

Lee, Y.-A. (2017). English word recognition skills of elementary school students. *Primary English Education, 23*(3), 27-50.

Lotherington, H., & Ronda, N. S. (2012). Multimodal literacies and assessment: Uncharted challenges in the English classroom. In C. Leung & B. V. Street (Eds.), *English–A changing medium for education* (pp. 104–128). Bristol, England: Multilingual Matters.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge, England: Cambridge University Press.

Mestres, E. T. & Pellicer-Sánchez, A. (2020). Young EFL learners processing of multimodal input: Examining learners' eye movements. *System, 80*, 212-223.

Ministry of Education. (2022). *English department curriculum* (No. 2022-33, supplementary volume 14). Sejong, Korea: Ministry of Education.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel - Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). U.S. Government Printing Office.

Neuman, S. B., Copple, C., & Bredekamp, S. (1998). *Learning to read and write: Developmentally appropriate practices for young children.* Washington, DC: National Association for the Education of Young Children.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Park, J. & Lee, J. (2022). Effects of multimodal English materials on EFL elementary students' English proficiency: Audiobooks vs. videos without captions. *Multimedia-Assisted Language Learning, 25*(3), 88-116.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In M. Traxler & M. A. Gerbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 613-657). London: Academic Press.

Tragant Mestres, E., Llanes Baró, À., & Pinyana Garriga, À. (2019). Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English. *Reading and Writing, 32*(3), 819-838.

Tragant Mestres, E., & Vallbona, A. (2018). Reading while listening to learn: Young EFL learners' perceptions. *ELT Journal, 72*(4), 395-404.

APPENDIX

The basic word test used in this study

No.	Word	Phonics Elements*	Dolch Sight Word
1	Apple	Consonant, Short vowel	○
2	Book	Consonant, Short vowel	×
3	Carrot	Consonant, Short vowel	×
4	Dog	Consonant, Short vowel	○
5	Egg	Consonant, Short vowel	○
6	Fox	Consonant, Short vowel	×
7	Golf	Consonant, Short vowel	×
8	House	Consonant, Vowel digraph	○

9	Juice	Consonant, Vowel digraph	×
10	Kick	Consonant, Short vowel	×
11	Lemon	Consonant, Short vowel	×
12	Milk	Consonant, Short vowel	○
13	Nose	Consonant, Long vowel	×
14	Open	Consonant, Long vowel	○
15	Paper	Consonant, Short vowel	○
16	Rabbit	Consonant, Short vowel	○
17	Snake	Blend, Short vowel	×
18	Tail	Consonant, Vowel digraph	×
19	Window	Consonant, Vowel digraph	○
20	Zoo	Consonant, Vowel digraph	×