SECOND LANGUAGE LEARNERS' STRATEGIES OF READING MULTIMODAL TEXTS: THE EFFECT OF SOCIAL MEDIA USE ON READING

DOI:10.18136/PE.2023.861

Sheida Marzban
Doctoral School of Applied Linguistics
University of Pannonia, Veszprém
sheida.marzban@phd.uni-pannon.hu

Doctoral (PhD) Dissertation

SECOND LANGUAGE LEARNERS' STRATEGIES OF READING MULTIMODAL TEXTS: THE EFFECT OF SOCIAL MEDIA USE ON READING

By

Sheida Marzban

Supervisors:

Prof. Dr. Gyöngyi Fábián

Prof. Dr. Béla Weiss

Multilingualism Doctoral School

Faculty of Modern Philology and Social Science

University of Pannonia

Veszprém, 2023

STATEMENT

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of Modern Philology and Social Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The content and research methodologies presented in this work represent the work of the candidate alone.

Sheida Marzban …,….,2023

Candidate Date

Dissertation Committee:

_____ …,….,2023

Chairperson Date

_____

First Reader

SECOND LANGUAGE LEARNERS' STRATEGIES OF READING MULTIMODAL TEXTS: THE EFFECT OF SOCIAL MEDIA USE ON READING

Thesis for obtaining a PhD degree in the Doctoral School of Multilingualism of the University of Pannonia

in the branch of Applied Linguistics

Written by Sheida Marzban

Supervisors: Prof. Dr. Gyöngyi Fábián & Prof. Dr. Béla Weiss

Propose acceptance (yes / no) …….…….

(supervisor)

…….…….

(supervisor)

As a reviewer, I propose acceptance of the thesis:

Name of Reviewer: …… yes / no …….…….

(reviewer)

Name of Reviewer: …… yes / no …….…….

(reviewer)

The PhD-candidate has achieved ……% at the public discussion.

Veszprém, ……/……. 2023 …….…….

(Chairman of the Committee)

The grade of the PhD Diploma …..... (……. %)

Veszprém, ……/……. 2023 …….…….

(Chairman of UDHC)

ABSTRACT

Visual displays are rapidly increasing in our daily lives. With the advent of technology and widespread use of the internet and social media, students are constantly connected and engaged in multimodal texts. Educational practices have also evolved, incorporating more visuals to improve communication and learning (Schnotz, 2002). Researchers are now compelled to study the complex relationship between modes in meaning-making. This pilot study employs a mixed-method approach (online and eye-tracking experiments) to investigate reading habits, modal preferences, and multimodal strategies among different age groups of second language learners (L2).

Seventy Hungarian L2 learners with B1 English language proficiency attended the online experiment to read and respond to the multimodal texts. Each task item included a multimodal text and a question with two answers, which represented either the visual or verbal information in the multimodal text. A reading habit questionnaire and follow-up questions were also used to collect information concerning the participant's reading habits and their opinions towards multimodal texts, respectively. While the findings of the reading habit questionnaire show that L2 learners were generally involved in reading social media, the results of the online multimodal reading test reveal that they preferred verbal over visual mode to answer the questions. Moreover, most of the participants perceived that images and their interactions with the texts helped their reading comprehension.

Dividing the population into Group A (13-29) and B (31-46), the results indicate that Group A looked at pictures and watched videos on social media more frequently than Group B did. In contrast, Group A preferred verbal over visual responses more frequently. As for the follow-up questions, Group B conceived that image and its interplay with text improved their reading comprehension while Group A found verbal information more beneficial to their learning gains. Moreover, the reading and response speed increased from adolescence (13-17) to young adulthood (24-29), however, they started to decrease towards the middle age (38-46).

The eye-tracking experiment seeks to examine second language learners' cognitive processes while reading multimodal texts. Eye movements were recorded during Hungarian L2 learners' (N=9) processing of multimodal texts in order to examine their gaze patterns. The findings reveal that the participants spent most of their time reading the text and preferred to have an overview of the text before fixating on the image. Some interconnected stages in multimodal reading were also inspected when the number of

fixations and fixation durations changed. Moreover, visual-verbal relations (e.g., text-subordinate-to-image and image-subordinate-to-text) are likely to determine the degree of the participants' involvement with the image. The present research may contribute to the development of a more comprehensive model of L2 multimodal and multimedia learning.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Fábián Gyöngyi. Her guidance, motivation and knowledge have helped me since the first day of my PhD study. Thank you for sparing the time and the knowledge in the process of our cooperation.

I am also extremely grateful to my co-supervisor, Dr. Weiss Béla, for his professional and valuable assistance during writing my dissertation. He generously shared his knowledge with me and his support was significant.

I want to extend my profound appreciation to Professor Judit Navracsics and Dr. Bátyi Szilvia, all teachers and faculty members, who have guided me through my PhD journey. Their kind support and patience have inspired me from the first day of my PhD journey.

I am thankful to the Hungarian Government for providing financial support during the study period.

I am also grateful to the principles of the language school and the secondary school and all my Hungarian participants who were involved in the study and assisted me in collecting the data despite all the problems.

Last but not least, I would like to express my gratitude to my beloved husband, dear sisters, and loving parents. Their unwavering support, affection, and assistance were pivotal in helping me overcome the challenges and hurdles throughout my academic journey.

Completing my dissertation would have been impossible without all these people. I thus dedicate this work to them and their support and care.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AOI	Area of Interest
CEFR	Common European Framework of Reference
EFL	English as a Foreign Language
ESL	English as a Second Language
L1	First Language
L2	Second Language
LDA	Latent Dirichlet Allocation
OGAMA	Open Gaze and Mouse Analyzer
SLA	Second Language acquisition
TESOL	Teaching English to Speakers of Other Languages

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Multimodality

Multimodality is a concept that has received increasing attention in recent years, particularly in the fields of linguistics, communication, and media studies. It refers to the use of multiple modes of communication to convey meaning, including text, image, sound, and gesture (Kress & van Leeuwen, 2001). According to Kress and van Leeuwen (2001), multimodal communication can be divided into three categories: linguistic, visual, and audio modes. The linguistic mode includes spoken and written language, and is the most commonly used mode of communication. However, it is not always sufficient to convey complex or abstract concepts, as it relies heavily on syntax and vocabulary. The visual mode includes still and moving images, and is particularly useful for conveying spatial and temporal relationships, as well as emotional and affective meanings (Jewitt, 2009). The audio mode includes sound and music, and can enhance the emotional impact of a message, as well as provide additional information or cues.

One of the key advantages of multimodal communication is its ability to convey complex ideas more effectively. By using multiple modes of communication, a message can be made more engaging, interactive, and immersive, which can help the audience understand and remember the information better (Machin, 2007). This is particularly important in the age of digital communication, where people are constantly bombarded with information and have shorter attention spans.

Multimodality also has implications for language learning and teaching, as it can provide a more comprehensive and authentic learning experience for learners and enhance learners' motivation, engagement, and comprehension (Varaporn & Sitthitikul, 2019)

However, multimodality also presents challenges, such as the need for careful consideration of how different modes interact and contribute to the overall message, and the potential for information overload or confusion if not executed effectively (Jewitt, 2009). In addition, the interpretation of multimodal messages can vary depending on the cultural and social contexts of the audience.

In sum, multimodality is a powerful tool for communication, with its ability to convey complex ideas and engage audiences in more meaningful ways. Understanding the principles and strategies of multimodal communication can help teachers and material

developers convey their messages more effectively and engage with their language learners in more meaningful ways.

As the exploration of multimodality comes to a close, it becomes evident that the impact of various modes of communication goes beyond traditional channels. The subsequent section discusses an intersection where multimodality converges with the powerful realm of social media. Therefore, the subsequent section delves into the pros and cons of social media as a multimodal platform in the era of technology.

1.2. Reading habits in the age of technology

The advent of technology has led to a significant transformation in the way people read and consume information. The traditional means of reading, which involved printed materials such as books and newspapers, have been replaced by digital technologies, including e-books, online articles, and blogs.

One of the most significant changes that technology has brought about is the ease of access to information. The internet has revolutionized the way we gather information, and we can access a vast variety of resources within seconds which has resulted in a preference for digital media over printed materials. Digital texts offer readers greater convenience and accessibility. Digital texts can be accessed from anywhere and at any time, and readers can carry hundreds of books with them on a single device. Moreover, digital texts often come with added features such as searchability, hyperlinks, and annotations, which can enhance the reading experience for some readers (Mangen et al., 2013).

The emergence of social media platforms has had a significant impact on reading habits in the digital age. Social media now serves as a prominent source of information and news worldwide. The ease of information sharing on these platforms has resulted in a higher consumption of concise articles and quick updates. Furthermore, the integration of various communication modes, such as text, images, videos, and audio, has made social media an alluring space for users. This combination of multimodality and social media has surpassed traditional boundaries, offering an immersive experience where individuals can express themselves more deeply and establish richer connections.

However, the shift to digital texts has also raised concerns about the impact on readers' comprehension and attention. Some studies have suggested that people tend to skim and scan digital texts rather than reading them thoroughly (Mangen et al., 2013). Furthermore, the constant distractions and interruptions that come with digital devices may lead to a

decrease in reading comprehension and retention (Carr, 2011). Therefore, it is important for readers to be mindful of their reading habits and to develop strategies to minimize distractions and enhance their comprehension of digital texts.

It is important to note that not everyone has embraced digital technologies, and some individuals still prefer traditional print media. The preference for printed materials may be due to a variety of reasons, including the tactile experience of holding a book or the lack of distractions that come with digital reading.

The advent of technology has revolutionized reading and information consumption, replacing traditional print media with digital technologies. The ease of access to vast online resources has encouraged a preference for digital media, offering convenience, accessibility, and enhanced features. Social media platforms have further reshaped reading habits, allowing for the consumption of concise content and fostering immersive experiences through multimodal communication. However, concerns about comprehension and distractions in digital reading persist, necessitating mindful reading strategies. While digital technologies dominate, it's important to acknowledge that some individuals still value the tangible and focused experience provided by traditional print media. Taking into account this foundation, my research aims to explore multimodal reading strategies within the context of social media use.

1.3. Background for the research

Teaching second language learners of different age groups can be challenging, as each group brings unique characteristics, learning styles, and preferences. As a language teacher with years of experience, I have observed significant differences in the way adult learners and younger learners approach their language learning journey. One of the most noticeable differences is their reading habits, particularly when it comes to reading multimodal texts.

Adult learners typically prefer to use EFL/ESL textbooks and classroom activities as their primary learning resources, and they tend to be more focused and attentive during class time. Adult learners are generally more motivated to learn a second language, and they understand the importance of reading and comprehending different types of texts, including those that are multimodal.

On the other hand, younger learners, particularly teenagers and young adults, tend to have a more relaxed and casual approach to language learning. They are often more tech-savvy and spend a significant amount of time on their mobile phones, browsing different

social media platforms, and interacting with various forms of digital content. During language classes, they often finish their tasks quickly and start using their mobile phones for entertainment.

These observations led me to realize the importance of studying L2 learners' reading habits and multimodal reading strategies in different age groups. Recognizing this importance can help language teachers develop effective teaching strategies that cater to the needs and interests of their learners. Therefore, I decided to quantitatively and qualitatively investigate the reading habits and modal preferences of L2 learners of different age groups and the impact of digital technologies, particularly social media platforms on their reading habits.

I believe that this research will contribute to the understanding of how language learners approach reading multimodal texts and provide valuable insights for language teachers and curriculum developers in developing effective teaching strategies. By shedding light on the reading practices of language learners and exploring the intricacies of multimodal texts, this study aims to contribute to the broader goal of enhancing language education. The forthcoming section will present an overview of the precise goals and organization of this dissertation, laying the foundation for a comprehensive exploration of this significant subject matter.

1.4. The goal and the structure of the dissertation

The advent of technology and the internet has transformed our lives, and social media platforms have become ubiquitous. With the proliferation of visual displays in our daily life, it is crucial to examine how second language (L2) learners read and comprehend multimodal texts. Reading multimodal texts involves processing different modes of semiotics, and it requires specific strategies that L2 learners must learn to navigate effectively. Moreover, the constant use of social media has altered our readings habits dramatically. Therefore, this dissertation seeks to investigate the reading habits and modal preferences of second language learners in different age groups and examine how they read multimodal texts. Exploring this topic might enhance our understanding of how L2 learners can develop effective strategies for reading multimodal texts.

The goals of the pilot study are as follows:

1- to investigate the reading habits and modal preferences of Hungarian L2 learners of English

2- to see the effect of age on the reading habits and modal preferences of Hungarian L2 learners of English

3- to explore how Hungarian L2 learners of English read multimodal texts.

4- to find out how intersemitic relations affect modal preferences.

The dissertation has 6 chapters. Chapter 2 provides a comprehensive review of relevant literature, presenting theoretical positions that lay the foundation for the research. It establishes the theoretical framework supporting the research questions, setting the stage for the empirical investigation that follows. In Chapter 3 of the dissertation, the research instruments and methodology are presented. This chapter describes the participants and the research tools utilized, which include the reading habit questionnaire and the multimodal reading test. The subsequent chapter, Chapter 4, details the findings of the study in the order of the research questions. Chapter 5 compares the results of the current study to those of prior research. In Chapter 6, the multimodal reading test's validity and reliability are examined, and guidelines are presented for its future enhancement and development. Finally, Chapter 7 presents the conclusion of the study, including an identification of its limitations, and proposes future directions for research.

CHAPTER 2

REVIEW OF LITERATURE

Visual displays have become increasingly prominent in our daily lives. With the rise of technology and the widespread use of the internet and social media, students are now spending more time on social media platforms, where nearly all texts are presented in a multimodal format. Additionally, compared to a few decades ago, there has been a significant increase in the use of visuals in educational settings, aiming to facilitate and enhance thinking processes and learning outcomes (Schnotz, 2002). As we navigate through this current era of multimodal communication, it is imperative for researchers to delve into the reading and comprehension of multimodal texts. Hence, the purpose of this dissertation is to investigate how age impacts Hungarian L2 learners' reading habits and multimodal reading strategies.

This chapter presents the background literature and the theoretical positions of the dissertation. Section 2.1 provides a general introduction to 'multimodality'. In Section 2.2, the discussion draws upon the social semiotic approach to examine models that highlight the respective roles of text and image within multimodal texts. Taking it a step further, section 2.3 delves into the exploration of intermodal relations between image and text. General ideas on learning from multimodal input and its application in L2 learning are provided in sections 2.4 and 2.5, respectively. Section 2.6 provides a discussion on processing visual and verbal information and how pictures facilitate learning from texts. Section 2.7 explains eye-movement and measurements on image and text. Section 2.8 talks about the effect of age and proficiency on L2 learners' gains from multimodal materials. Section 2.9 discusses how visual and verbal discourses interact on social media to convey the meaning and how social media affect reading habits. Section 2.10 provides a brief summary of background literature and research questions.

2.1. Introduction to multimodality

In this day and age, visual communication lives all around us. Also, in contrast to years ago, visual displays are more present in curriculum designs at schools to develop students' engagement and interpretation during lessons. Comenius (1985) emphasized the significance of visual information for effective learning in his book 'Didacta Magna' during the seventeenth century. This perspective has influenced scholars in the field of educational science, who have incorporated the visualization of information into their curriculum. The rationale behind this approach lies in the fact that visual displays

facilitate and enhance communication, thinking, and learning. However, it is important to note that utilizing visual information effectively necessitates certain prerequisites, including prior knowledge and cognitive skills (Schnotz, 2002).

In the 20th century, the concept of multimodality lacked a precise definition. However, with the rise of technology and the introduction of new presentation methods, multimodality has become the norm in the 21st century, gradually overshadowing monomodality. Multimodality refers to the utilization of multiple modes within a single medium (Lutkewitte, 2013). For instance, a televised weather forecast incorporates spoken language, written text, and weather-specific symbols such as clouds and rain as modes of communication (Multimodality, 2023). Contemporary researchers and theorists have shifted their focus from the predominance of printed words to the visual image and multimodal discourse (Anstey & Bull, 2006; Kress, 2003). These modes are signifiers created within different contexts to express and convey meanings. According to Kress (2010), text often plays a subordinate role to visuals in providing information to readers. Examples of multimodal communication include street signs, billboards, picture books, and websites. Beardsley (1981) suggests:

> *A picture is two things at once: it is a design, and it is a picture of something. In other words, it presents something to the eye for direct inspection, and it represents something that exists, or might exist outside the picture frame. (Beardsley, 1981:267)*

During the last few decades, the discourse of the content area has been popularized as the core of meaning making and learning in any school or academic subjects (Lemke, 1990; Norris & Phillips, 2003; Wellington & Osborne, 2001). Subject discourses have been described focusing on how writing is used in different contexts (e.g., terminology, phrases, and text structure) (Halliday & Martin, 1993; Schleppegrell, 2004). Based on these descriptions, frameworks were developed which explicitly focused on language and content in written texts. Examples are genre-based pedagogies developed in Australia (Gibbons, 2002; Martin & Rose, 2012). Moreover, some frameworks for improving reading comprehension at schools were put forward. The aim of the framework was for students to become active readers and monitor their reading (Beck & McKeown, 2006, Palincsar & Brown, 1984). The problem with these models was that they were biased towards text (written and oral) and did not consider meaning making as multimodal (Kress, 2009; Norris, 2011). Reacting to these models, scholars have asserted the difficulty and challenges students face using and combining different meaning making

resources (Danielsson, 2011; Kress et al., 2001; Tang et al., 2014). Considering chemistry texts as an example, the same scientific content such as 'ionic bond' is presented in different ways (verbals, visuals, chemical symbols), each emphasizing a certain aspect of the ionic bond. However, we do not know if all students are able to figure out why different modes are adopted to represent the same meaning.

A change in knowledge representation in textbooks entails meaning making from multiple semiotic resources. Hence, it is important to investigate various resources readers use and how they move between visual and verbal information to comprehend multimodal texts. In order to consider multimodal texts as a dimension of reading, conceptual frames should go beyond the traditional cognitive models. Multimodal reading combines ideas from different fields as learning psychology and sociosemiotics (Abraham & Farías, 2017). Multimodality introduced a semiotic approach to reading, which consequently revolutionized understanding of meaning-making processes by people of today's world (Abraham & Farías, 2017). Siegel (2012) points out the youth's multimodal literacy capabilities:

> *It is tempting to suggest that this is the time of multimodality: A time when the privileged status of language is being challenged by the ease with which youth can access semiotic resources of all varieties —visual, aural, gestural, and spatial— to assemble meanings. (Siegel, 2012: 671)*

When it comes to language learning, Kress (2000) highlights the importance for language teachers to consider the contextual aspects where language and accompanying resources, such as images in multimodal texts (combining visual and verbal information), contribute to the construction of meaning. Kress (2000) introduces the concepts of modes and affordances to discuss the interplay between images, text, and other semiotic modes, and how users generate their intended meanings. Affordances, as defined by Kress (2010), refer to the social and cultural potentials and limitations that each mode offers, determining what can or cannot be expressed and communicated through different modes of meaning making. Drawing attention to the intersemiotic relations between text and image, Royce (2002) emphasizes the necessity for learners to develop their multimodal communicative competence, enabling them to interpret and construct appropriate meanings by integrating various modes. Multimodality directs readers' focus towards visual information, which plays a role in meaning making (Abraham & Farías, 2017; Afflerbach & Cho, 2010). Thus, it is essential for readers to perceive visuals as an integral

part of the text rather than mere ornamental or aesthetic elements (Abraham & Farías, 2017).

In the realm of meaning making, the interplay between image and text across various media is a subject of research aimed at understanding the semantic relations between these two modes through the creation of models. These models seek to uncover the diverse roles that image can play, ranging from purely decorative purposes to providing substantial information in image-text relations (Kornalijnslijper, 2007). Additionally, there are instances of image-text combinations where the text assumes a secondary role, such as captions beneath images in newspapers (Kornalijnslijper, 2007). As models become more detailed, they enable the identification of specific properties within image-text relations. Consequently, there is a growing demand for models that can effectively describe the relations between these two modes, considering scenarios where text complements image, or both modes are interdependent or independent. The following section will delve into some notable models of multimodality, shedding light on their contributions to this field of study.

2.2. Models of multimodality

The process of multimodal communication is intricate and highlights the significance of examining the utilization of various resources by readers, as well as comprehending the complex interrelationships among these resources. As a result, researchers are driven to develop models that can effectively capture and analyze the dynamics of multimodal discourse. By creating applicable models, researchers can gain valuable insights into the intricate nature of multimodal communication and its impact on meaning-making processes. Table 1 provides an overview of the three models which are discussed in this section as examples.

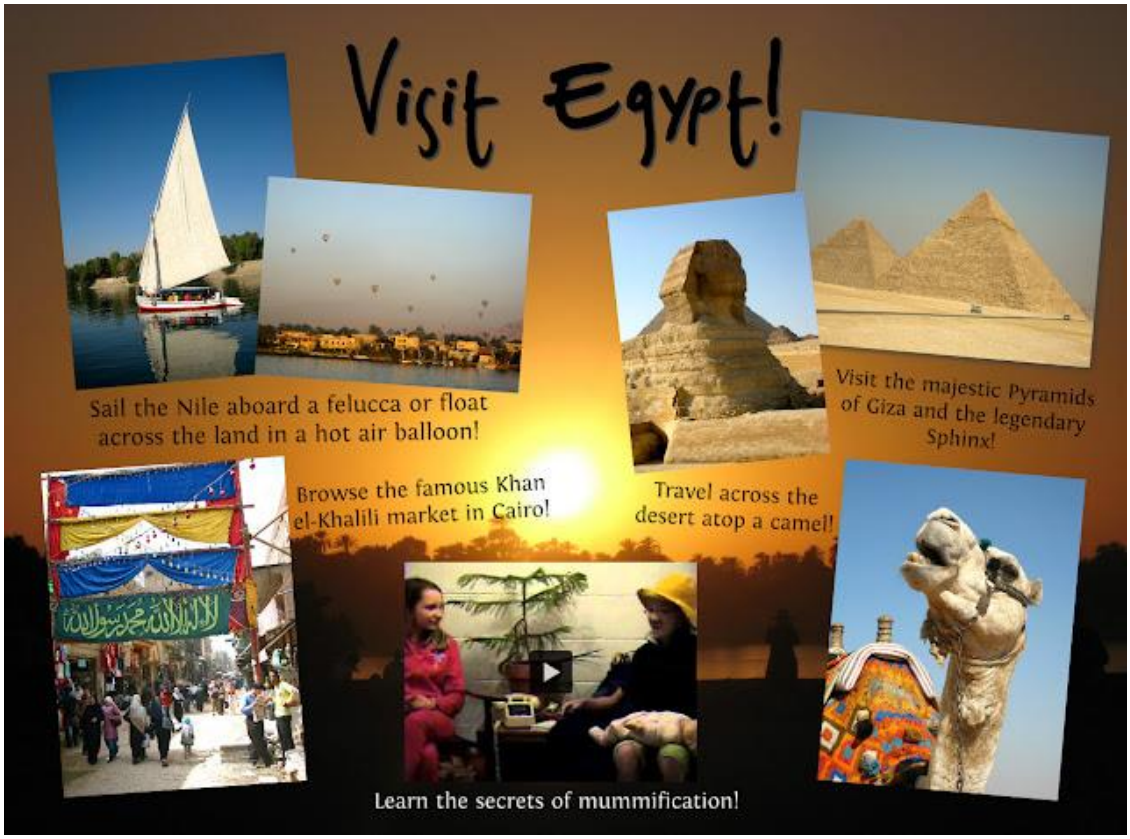
Table 1. Some models of multimodality

Model	Description	Visual/verbal functions
Carney and Levin (2002)	The model does not consider cases where verbal texts serve a function to images or both modes show the same importance.	5 image functions: Decorational Representational Organizational Interpretational Transformational
Marsh and White (2003)	The model only identifies the function of images in text and does not consider cases where text is subordinate to image or where both modes are equally important.	49 image functions in 3 groups: Little relation to text Close relation to text Beyond text
Halliday (1994) Kress and van Leeuwen (2006)	The model can be used to analyze both verbal and visual discourse.	Five functions in visual and verbal discourse: contact distance attitude narrative representation position

Carney and Levin (2002) talk about the role of image in multimodal texts and do not consider cases where verbal text serves a function to image or both modes show the same importance. They identify five different functions of image in multimodal texts: ‘decorational’, ‘representational’, ‘organizational’, ‘interpretational’ and ‘transformational’. Each image-text relationship can be assigned to one single function type.

Image might have a ‘decorational’ function when they only serve text. In other words, it carries little or no further information to text. For instance, in a traveling brochure to Egypt, the picture of the sun is decorational (Kornalijnslijper, 2007) (Figure 1).

Figure 1. Decorational picture



Source: retrieved from https://www.tripadvisor.com/LocationPhotoDirectLink-g294201-d15656120-i368680539-Egypt_Sparrow_Tours-Cairo_Cairo_Governorate.html on February 13, 2023

'Representational' image might also partly or completely depict what text describes. It might even go beyond text and present more than what text describes. An example can be a painting in an art gallery along with a text which describes the painting's content (Kornalijnslijper, 2007) (Figure 2).

Figure 2. Representational picture

Source: retrieved from https://www.pinterest.com/pin/525724956485480959/ on February 13, 2023

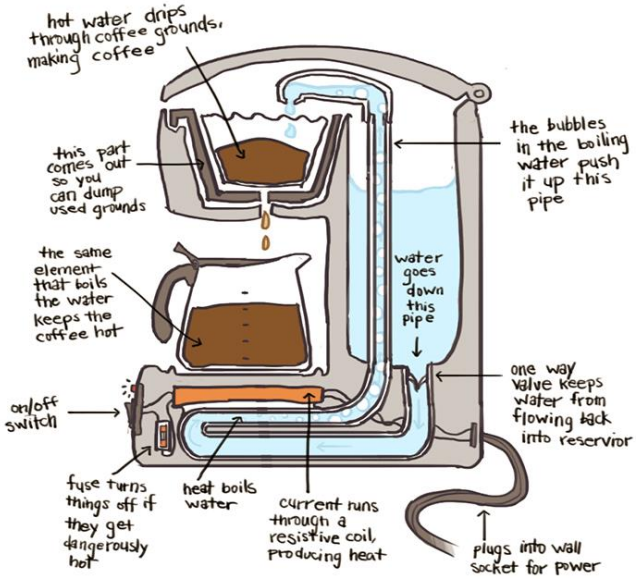
'Organizational' image carries structural information of text. The structural information is usually presented in steps. The example is an illustrated map of a cycling trail, which depicts the steps to find the route (Carney & Levin, 2002) (Figure 3).

Figure 3. Organizational picture

Source: retrieved from https://www.dreamstime.com/bike-route-map-cycling-trip-road-country-path-bike-adventure-tour-vector-map-bike-route-map-cycling-trip-road-country-path-bike-image135971186 on February 13, 2023

'Interpretational' image is used in multimodal texts, which cannot fully explain and communicate the information. Figure 4 shows a picture of working machinery as an example.

Figure 4. Interpretational picture



Source: retrieved from https://www.hipresso.eu/en/blogs/hipresso-blog/hoe-werkt-een-koffiezetapparaat on February 13, 2023

Carney and Levin (2002) mention transformational pictures include systemic mnemonic components, which enhance memory by depicting verbal information in a literal sense. However, the whole picture does not necessarily depict the intended meaning literally (Kornalijnslijper, 2007). Carney and Levin provide an example for this

function of image. In Figure 5, there are elements which are literally translated from the text about the town Belleview, such as a bell in the picture representing the part of the town which is named ‘Bell’. The readers should figure out the associations between the components of image and text to store the mnemonic picture in the memory.

Figure 5. Transformational picture

Source: retrieved from http://mybelleview.org/ on February 13, 2023

The model proposed by Marsh and White in 2003 presents a greater level of complexity compared to Carney and Levin's model from 2002. Marsh and White's model primarily focuses on discerning the role and purpose of image within textual content. They introduce a taxonomy of 49 image functions which are presented in Figure 6.

Figure 6. Taxonomy of functions of images to the text (Marsh & White, 2003: 653)

A Functions expressing little relation to the text	B Functions expressing close relation to the text	C Functions that go beyond the text
A1 Decorate	*B1 Reiterate*	*C1 Interpret*
A1.1 Change pace	B1.1 Concretize	C1.1 Emphasize
A1.2 Match style	B1.1.1 Sample	C1.2 Document
A2 Elicit emotion	B1.1.1.1 Author/Source	*C2 Develop*
A2.1 Alienate	B1.2 Humanize	C2.1 Compare
A2.2 Express poetically	B1.3 Common referent	C2.2 Contrast
A3 Control	B1.4 Describe	*C3 Transform*
A3.1 Engage	B1.5 Graph	C3.1 Alternate progress
A3.2 Motivate	B1.6 Exemplify	C3.2 Model
	B1.7 Translate	C3.2.1 Model cognitive process
	B2 Organize	C3.2.2 Model physical process
	B2.1 Isolate	C3.3 Inspire
	B2.2 Contain	
	B2.3 Locate	
	B2.4 Induce perspective	
	B3 Relate	
	B3.1 Compare	
	B3.2 Contrast	
	B3.3 Parallel	
	B4 Condense	
	B4.1 Concentrate	
	B4.2 Compact	
	B5 Explain	
	B5.1 Define	
	B5.2 Complement	

The following section explains the categories. Group A includes image that decorates, elicits emotion and controls texts. ‘Decorate’ type makes text more attractive without adding to the content of the information or easing its understanding. ‘Elicit emotion’ type shows a content which brings about a certain emotion or feeling. ‘Control’ type directs influence on the reader by holding his attention and suggesting a response from him (Kornalijnslijper, 2007). In group B, image reiterates, organizes, relates, condenses and explains text. ‘Reiterate’ type describes image which present the same content as text. ‘Organize’ type depicts the information in a structured form or in a way which is graphically explained like diagrams, charts, and maps. ‘Relate’ type describes processes and concepts in text. ‘Condense’ type compacts or reduces the information to their necessary details. ‘Explain’ makes the information plain or understandable by following text closely. Considering this type, image identifies essential qualities of information and complements text by transferring the intended meaning (Kornalijnslijper, 2007). In group C image interprets, develops, and transforms text. ‘Interpret’ type clarifies complex textual concepts by emphasizing text or providing factual or substantial support. ‘Develop’ expands information by creating more details and illustrations. ‘Transform’ presents text in another form. The information is recoded, related or organized to expand recall and memory. In this category, image might expand text or even model ideas which

cannot be presented by text. Group C embodies a more intricate model that embraces relationships where text is subordinate to image, or both modes hold equal significance (Kornalijnslijper, 2007).

Halliday (1994) and Kress and van Leeuwen's (2006) framework can also be used to analyze verbal and visual discourse, respectively. Halliday's main focus is on verbal language. The use of social semiotics in reading image is proposed by Kress and van Leeuwen (2006), among others. Kress and van Leeuwen's (2006) model of analyzing image is based on Halliday's grammar. The authors show that reading image and text is similar because they are both representations of social identities, social realities and social relations. Five items of 'contact', 'distance', 'attitude', 'narrative representation' and 'position' are introduced in their frameworks.

'Contact' determines the type of relation between viewers and participants. There are two kinds of contact based on Kress and van Leeuwen's (2006) model. In an image in which the participant is gazing at the viewer directly, it is demanding or simply asking the viewer for a reaction. On the other hand, when the participant gazes away from the viewer, he/she is offered as an object for the viewer's thorough investigation. Kress and Van Leeuwen (2006) relate the 'image contact' to the 'speech act' in language. Based on Halliday (1994), speech acts are divided into 2 basic categories of speech roles ('giving' and 'demanding') and commodities ('information' and 'goods and services') out of which 4 basic speech acts are yielded: statements, questions, commands, and offers.

'Distance' is "the distance from which people, places or things are shown" (Macken-Horarik, 2004: 14). 'Distance' in visual discourse includes very close, medium close, medium, medium long, and long shot. 'Distance' is realized in language through 3 basic formalities of style (Kress & Van Leeuwen, 2006). In 'personal style', language is implicit and context dependent and includes more informal and colloquial expressions. In 'social style', language is more explicit and standard in terms of syntax and lexis through which the business is conducted. 'Public style' of language, more explicit than social style, is articulated more formally and consciously (Kress & Van Leeuwen, 2006).

Participants are also shown from two different dimensions of 'vertical' and 'horizontal'. Vertical dimension (Kress & Van Leeuwen, 2006) indicates power relation and includes 3 basic 'low-angle', 'high-angle', and 'eye-level' shots. In the 'low angle' shot, the participant is shown from below indicating the power of the participant over the viewer. In the 'high angle' shot, the participant is shown from above indicating the power of the viewer over the participant. In 'eye-level' angle, both the viewer and the participant

are set in the same level indicating equality. 'Horizontal' dimension explains the degree of involvement and includes frontal and oblique dimensions. Viewers are considered to be more involved with the participants presented frontally (Kress & Van Leeuwen, 2006). Like visual discourse, there are two dimensions (vertical and horizontal) in verbal discourse. Discursive analysis of attitude was first proposed by the work done within Sydney School of Systemic Functional Grammar in the domain of appraisal system. Within appraisal system attitude is analyzed through the category of value. For value, 3 subcategories of 'affect', 'judgment' and 'appreciation' are under the system of appraisal. Although expressions of emotional impulses and desires bring about intimacy with the reader, statement of one's 'judgment' or 'appreciation' forms hierarchies (Macken-Horarik, 2004). Thus, it is expected that viewers involve more with the participants who express their personal emotional responses or impulses and less with the subjects who concern 'judgment' and 'appreciation'. In order to investigate text vertically, the system of 'theme' and 'focus' is applied (Halliday, 1994). 'Theme' is the point of the author's departure on the message which happens to come at the first position in a declarative clause or, in other words, takes the role of a subject. And 'focus' shows the speaker-oriented prominence and happens at the end of the clause and is called 'new'.

Actions in visual discourse are represented by vectors among elements or represented participants (narrative representation) (Kress & Van Leeuwen, 2006). Depending on the number and kind of participants and vectors, different narrative processes can be distinguished: 'action' process (the one through which the participant emanates a vector), 'reactional' process (the one identified by the vector of a look or glance direction of the reactor), 'mental' process (the one formed by the thought or dialogue balloons emanate from the participant), and 'verbal' process (the one indicating the participant is talking to someone else). Based on Transitivity System of Halliday (1994), four types of 'material', 'mental', 'relational', 'verbal' types of process are investigated in verbal discourse. *I kick the ball* indicates an action done by the participant and extended to the goal or 'the ball'. While 'material' processes are mostly concerned with experiences of material worlds, 'mental' clauses are mostly concerned with emotive (like, love, hate verbs), cognitive (think, know, notice, remember), perceptive (hear, taste, see) and desiderative (want, would like, wish) processes (Halliday, 1994). 'Relational' clauses reflect inner and outer experiences as being. Thus, *She is exercising in the park*, *She really loves parks* and *she is in the park* are 'material', 'mental' and 'relational' processes, respectively (Halliday &

Matthiessen, 2004). ‘Verbal’ clauses are clauses of saying (Halliday & Matthiessen, 2004) like *John said I’m thirsty*.

Kress and van Leeuwen (2006) define ‘composition' as the elements that attract viewer’s attention. Participants in an image can be positioned on the left or right. The vertical arrow divides left (old) and right (new) parts. The left tends to have meanings of what is taken as given, known, or assumed to be the case. The right space is reserved for the new or for the finishing point.

In summary, researchers propose different models regarding the relationship between text and image. While some researchers do not consider the hierarchical or equal status of text and image, others introduce a model that applies to both modes. However, in today's post-modern era, where boundaries in various aspects of life are becoming blurred due to the unrestricted exchange of information, it becomes crucial to acknowledge the relationship between two prominent forms of communication: text and image. This relationship becomes particularly significant considering the rapid advancements in information technology, as text and image progressively merge, giving rise to the emergence of multimodal texts. The following section will delve further into the general system of image-text relationships that can be applied to instances of multimodal discourse where image and text coexist.

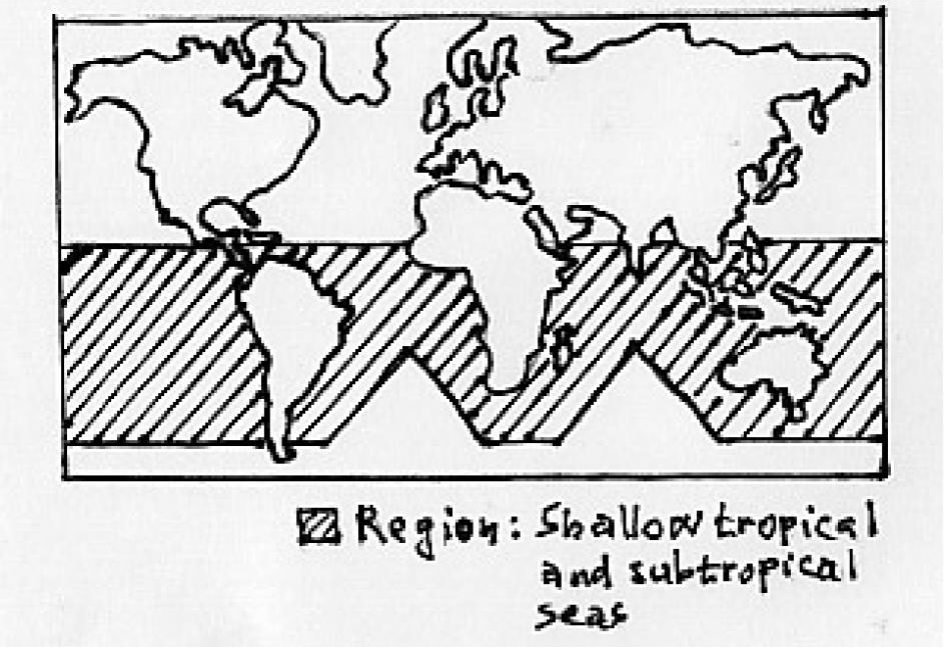
2.3. A system for image-text relations

The main focus of this section is to analyze the connection between image and text, highlighting scenarios where the degree of fusion is not overly intense. Instead, the article explores situations where the two elements maintain their distinctiveness while effectively combining in terms of both meaning and structure. Martinec and Salway (2005) adopt an inclusive approach, not distinguishing between visual and verbal discourse, and instead acknowledge the interdependence between these modes. Martinec and Salway (2005) have not categorized their relational types based on the type of the mode. However, they refer to visual and verbal interactions together in a multimodal relation. This model ‘a generalized system of image-text relations’ can be used to describe relations in which image is subordinate to text or text is subordinate to image or both modes are equally dependent or independent of each other. This model is based on the earlier work of Barthes (1977a, 1977b) and Halliday (1994). Halliday’ model (1994) concentrates on text, which is later combined with Barthes’s work to analyze multimodal texts. Although their idea has previously been used only for specific examples of

multimodal relations, Martinec and Salway (2005) use their model for all image-text relations. They also assert more than one relation can exist between image and text.

Status relation identifies the relative status between text and image, which can be equal or unequal. In an equal relationship the modes are either independent or complementary to each other. In an independent relationship “the whole image is related to the whole text” (Martinec & Salway, 2005: 345) and they can exist independently. An example of such an independent relationship is provided in Figure 7 where the whole map is in accord with the caption and they create their own processes.

Figure 7. Example of independent image-text relationship (Martinec & Salway, 2005: 346)

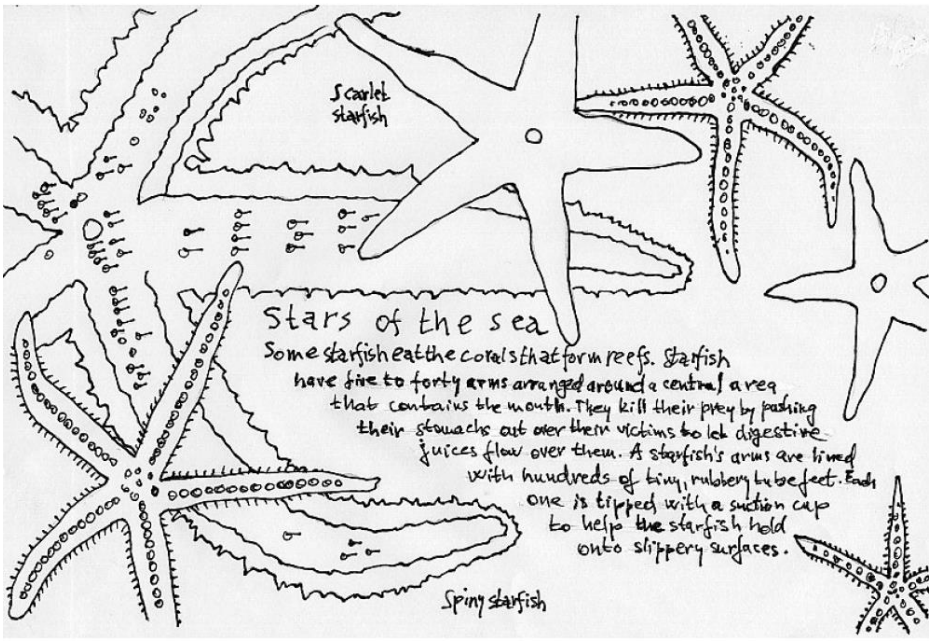


In a complementary relationship, both modes are merged "to form part of a larger syntagm" (Martinec & Salway, 2005: 346). Figure 8 provides an example of a complementary relationship. On one side, two teddy bears embody the characteristic of being "Sweet. But not too sweet." On the other side, these teddies transfer their attributes of being "cute but naughty" to the cereals, which entice the younger audience targeted by the product.

Figure 8. Example of complementary image-text relationship (Martinec & Salway, 2005: 347)

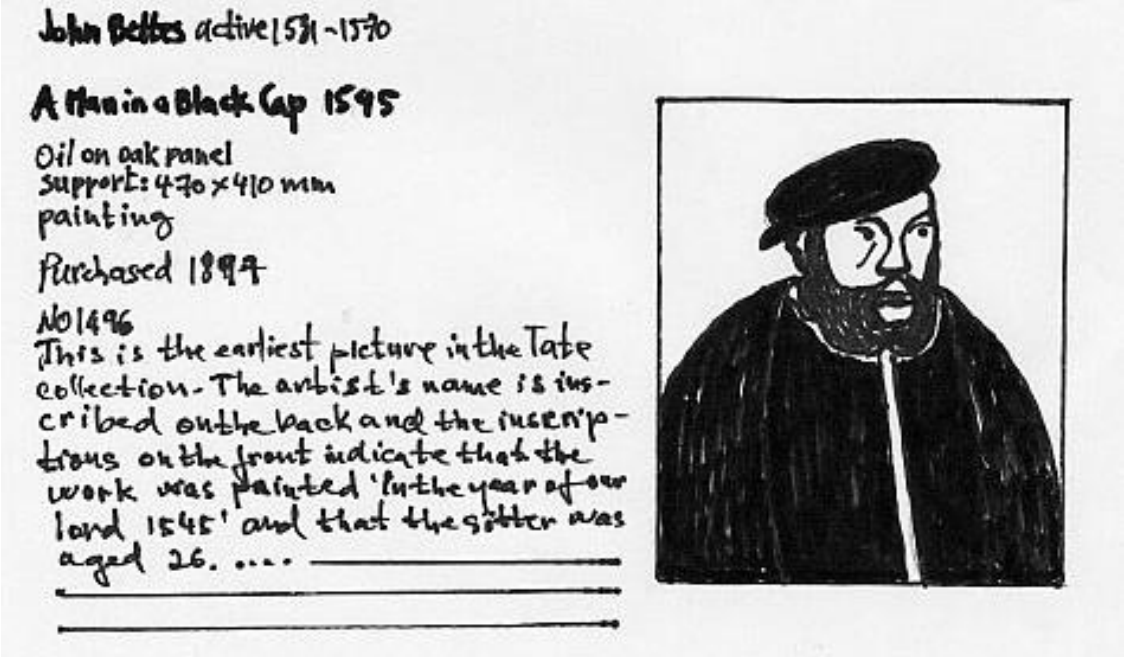
In an unequal relationship one mode is subordinate to and serves the other. Image subordination happens when "the image is related only to a part of the text" (Martinec & Salway, 2005: 346). An example of such a relationship is in Figure 9. The image of the starfish relates to only some parts of the information in the text, for instance '*Starfish have five to forty arms arranged around a central area that contains the mouth*.' However, some verbal information such as '*Some starfish eat the corals that form reefs. They kill their prey by pushing their stomachs out over their victims to let digestive juices flow over them*.' is not illustrated in the image.

Figure 9. Example of image–subordinate-to-text relationship (Martinec & Salway, 2005: 348)



Text is subordinate to image when it only relates to part of the image. The implicit devices that refer the readers to image are indicators of text subordination. An example of text subordination is provided in Figure 10. Presence of numerous textual references (e.g. *this, the work, the sitter,* etc.) subordinates the text to the image.

Figure 10. Example of text–subordinate-to-image relationship. (Martinec & Salway, 2005: 348)



With regard to the Logico-semantic relations, one uses exemplification by expanding the information with more specific examples of the information. Exemplification is subdivided into 'text more general' and 'image more general' groups. In figure 11, the image is more general than the text. The image features a skull and crossbones, a widely recognized symbol that conveys danger and death. In contrast, the accompanying text provides specific details, narrowing down the meaning by explaining how eels have the capability to kill their prey.

Figure 11. Image-more-general-than-text relationship (Martinec & Salway, 2005: 348).

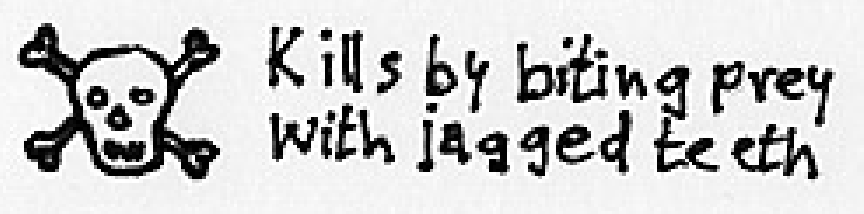


Figure 9 exemplifies a case where the text is more general in nature compared to the image. For instance, the term "Starfish" mentioned in the text encompasses the entire class of sea creatures known as starfish. However, the image showcases various specific types of starfish (such as scarlet and spiny) as examples belonging to that particular class.

When one mode enhances the other, it imparts information regarding the location, time, or reason. For example, in Figure 12, the image offers some details about the specific place where the woman arrived late.

Figure 12. An example of enhancement by place. (Martinec & Salway, 2005: 353)

Another example is shown in Figure 13. The phrase ‘following a nervous breakdown’ provides information about when the painting, which is illustrated on the right side of Figure 13, was painted.

Figure 13. Example of enhancement by time. (Martinec & Salway, 2005: 353)

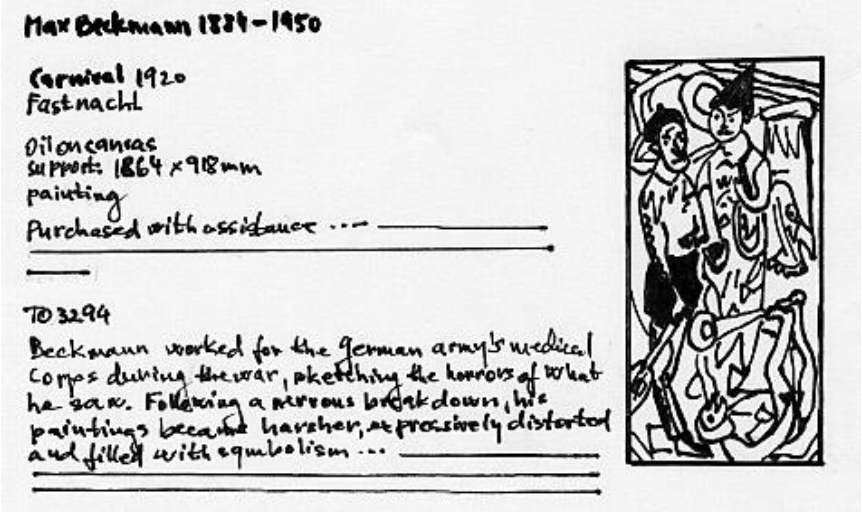


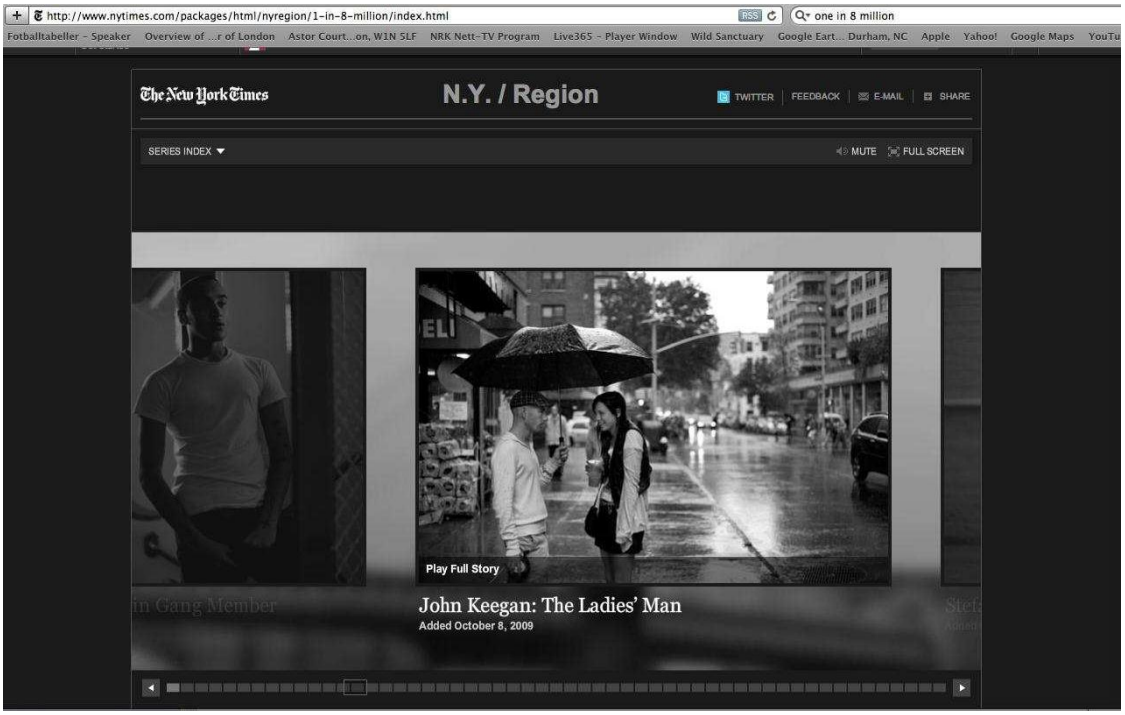
Figure 14 serves as another illustration where the image enriches the text by presenting additional information regarding the aftermath of the fire in the hall, specifically the presence of deceased individuals on the floor.

Figure 14. Example of enhancement by reason. (Martinec & Salway, 2005: 354)

When visual information diverges from verbal information, it creates tension between image and text. Engebretsen (2012) explains that cohesion in multimodal texts forms the basis for comprehension, while tension prompts readers to react, engage, and draw conclusions, thereby facilitating a meaningful learning process. Figure 15 serves as an example of image-text tension, where the image depicts a man seemingly providing shelter to a girl under his umbrella. According to Engebretsen (2012), both cohesion and tension coexist between the text and image in this scenario.

> *It is, however, not a relationship of pure repetition, as the verbal label offers us additional information about one of the person's identities as well as an instruction of how we are to interpret the picture. The label thus reduces the polysemy inherent in the picture: the situation is to be interpreted as one of flirting, and it is the male individual that is central in this context. On the semantic level, we thus recognize a high level of coherence between the modes, although still not a situation of total inter-modal paraphrase. (Engebretsen, 2012: 15)*

Figure 15. Example of image-text tension (Engebretsen, 2012: 13)



The presence of semantic dissimilarity, distance, and discontinuity in multimodal texts plays a crucial role in fostering students' engagement and attention, as noted by Engebretsen (2012). Tension, as emphasized by Engebretsen (2012), serves as a vital mechanism that invites readers to actively interact with and contemplate verbal-visual texts. The balanced interplay between cohesion and tension establishes a dynamic relationship between visual and verbal elements of multimodal texts (Engebretsen, 2012).

To conclude, researchers underscore the significance of providing a diverse range of models to thoroughly examine and understand multimodal discourse. They strongly urge scholars and academics to delve deep into the intricate and dynamic interplay that takes place between various modes of communication. By exploring and unraveling the complexities of how different modes interact and influence each other, researchers can gain profound insights into the multifaceted nature of human communication. This comprehensive analysis not only enriches our understanding of multimodal discourse but also fosters a more holistic approach to studying and interpreting the various ways in which humans convey meaning and construct knowledge. Moreover, scholars acknowledge that students can gain substantial advantages from immersive and interconnected encounters facilitated by multimodal resources. These materials possess the capacity to augment learning results by involving learners in a comprehensive and diverse manner, as explored in the subsequent section.

2.4. Multimodality and learning

Visual displays are increasingly used in education, providing supportive effects on communication, thinking, and learning (Schnotz, 2002). Literacy is no longer confined to language alone, as learners now interact with multimodal texts incorporating images, audios, and videos (Kress, 2003). Moreover, the form of representation shapes meaning-making and learning, necessitating a better understanding of how students engage with multimodal materials (Jewitt, 2009). Informed by the social semiotic approach and a multimodal perspective on meaning making (Bezemer & Kress, 2010; Jewitt, 2005, 2009; Kress & van Leeuwen, 2006; Selander & Kress, 2010), Danielsson and Selander (2016) emphasize, based on their classroom studies, the limited presence of meta-textual discussions in classroom practices within their model. Specifically, there is a scarcity of discussions concerning the multimodal aspects of texts (e.g. Danielsson, 2011; Løvland, 2010). Table 2 draws on multimodality and the social semiotic approach to show how multimodal texts should be worked with in the classrooms. Metaphors and analogies are part of disciplinary discourse, which can be expressed through text and image. Nonetheless, if students take the analogy too far, it might create some learning problems (Haglund, 2013). Taking an example from chemistry classrooms, metaphoric phrase 'electronic shells' is used to explain how electrons move around at a certain distance, which might give the false impression to students that they are actually shells (Danielsson, 2011). Values (implicit and explicit) can also be expressed through different semiotic resources. For instance, image can be used to convey values implicitly (Kress & van Leeuwen, 2006). This model introduces how teachers and students can unfold text and support meaning making from the content. It also develops meta-cognitive skills by helping beginners realize the missing parts and the links between different pieces of information and the interplay between different resources to present knowledge (Danielsson & Selander, 2016). By emphasizing subject content and engaging in meta-textual classroom discussions, students can enhance their understanding and broaden their multimodal literacy skills.

Table 2. Model for working with multimodal texts in education (Danielsson & Selander, 2016: 26)

	Multimodal text focus	**Classroom focus**
General structure-setting	Thematic orientation and sequencing (What is the text about and how is the content arranged?) What do each one of the resources express? (images, diagrams, headings, introductory paragraph, etc.)	Examination of the different parts of the text and what content they provide
Interaction between textual parts	Proximity/closeness and coherence between writing and other semiotic resources Congruence and coherence between concepts, description and explanations	Reflection as to the interaction between different resources and what aspects of the content that appear as central
Figurative language (in writing and images, etc)	What analogies and metaphors are used?	Deconstruction of the figurative language. How far do the analogies reach, and how well do they function for the specific content?
Values	Explicit Implicit (e.g. in metaphors, images, perspectives)	Discussion of aspects such as right/wrong, us/the others, female/male, etc.

Moreover, Paivio's dual coding theory (1986) acknowledges the benefits of multimodal texts for literacy development and explains how image facilitates learning from text (Clark & Paivio, 1991). In his theory, it is proposed that cognition is served by two interconnected yet independent systems. One system processes verbal information, such as text and speech, while the other system encodes visual information, including images and graphics. Meaningful learning involves building connections between visual and verbal representations and their active processing (Mayer & Moreno, 2003). Building on the Dual Coding Theory, Mayer (1997) also develops the Cognitive Theory of Multimedia Learning, according to which visual and verbal information is processed through different cognitive subsystems. Readers actively select pertinent words and images, construct their own mental representations, and develop the information into both visual and verbal mental models. These models are then integrated and mapped onto each other in the cognitive process. Based on his multimedia theory, Mayer (2009) makes a list of principles which defines successful multimedia learning. The *multimedia* principle explains that students learn better from visual and verbal information integrated than from verbal information alone. The *coherence* principle states that students learn better when

irrelevant material is excluded and *redundancy* occurs when the same information is presented in simultaneous multiple forms (Mayer, 2001). The Generative Theory of multimedia learning (Mayer, 1997) is concerned with meaningful learning which involves cognitive activities of selecting, organizing and integrating verbal and pictorial representations. It explains that the amount of information which can be processed by learners via each channel is limited (limited-capacity assumption) (Mayer, 1997). Cognitive Load Theory suggests that "effective instructional material facilitates learning by directing cognitive resources toward activities that are relevant to learning rather than toward preliminaries to learning" (Chandler & Sweller, 1991: 293). Consequently, the process of learning is facilitated by eliminating redundant information from multimodal materials, which helps prevent learners from experiencing unnecessary cognitive burden (Kalyuga et al., 1999).

Despite Paivio and Mayers' principles, which support the use of pictures alongside text in multimodal texts, a great deal of empirical research has been conducted in the context of domain learning, especially science in first language (L1) (Johnson & Mayer, 2012). This emphasis on domain learning has led to a gap in research specifically targeting language learning, where language continues to be regarded as the primary mode of communication. Jakobsen (2019) states that language learning is different from domain learning as language is still considered the main mode of communication, adding that multimodality is about recognising modes in connection with language and is not necessarily concerned with replacing it. Therefore, there is a clear demand for additional exploration and investigation into how multimedia principles can be effectively applied in second language (L2) learning contexts.

2.5. Multimodality and second language learning

Due to the advent of technology, the current teaching and learning environment has become increasingly multimodal and teaching materials, such as textbooks, employ multiple modes of presentation, including images and texts, which will hopefully support reading comprehension and facilitate learning (Pellicer-Sánchez, 2022). This gradual change in knowledge representation, from language as the dominant representational mode to multimodality (Bezemer & Kress, 2008), suggests research on various resources readers use and their movements between visual and verbal information to read multimodal texts. Furthermore, studies must be conducted to respond to the main challenges in using and combining different meaning making resources.

Although most of Mayer's work has focused on L1 and domain learning, due to the absence of a more relevant theoretical account, his principles of multimedia learning have been applied in the interpretation of the findings from L2 multimodal studies (Pellicer-Sánchez, 2022). For instance, while redundant information in multimodal materials is believed to create cognitive overload and inhibit learning by such concurrent presentations (Mayer, 2009), Lee and Mayer (2018) observe a reversed redundancy effect in L2 context. For instance, according to their findings, learners who are exposed to the narrated video and onscreen text learn better than students who only watch the narrated video due to the fact that the print text provides a more long-lasting material, which contributes to their processing of each word. Hence, it is essential to develop efficient models for L2 multimodal and multimedia language learning because its primary goal is to master a new language and not to learn about a specific domain (Hung et al., 2013; Schnotz & Baadte, 2008). In recent research on multimodality in second language (L2) learning, various aspects of the teaching and learning process have been examined. This includes the exploration of ELT textbooks, captioned videos, and multimodal reading texts, which are briefly discussed in the following sections.

2.5.1. ELT textbooks

So far, a lot of studies in multimodality and L2 learning have focused on multimodal discourse analysis of language textbooks. Ajayi (2012) believes English language textbook compilers combine visual and verbal resources (e.g., image, color, layout, typography and font) to communicate information. However, Bezemer and Kress (2010) mention textbook producers sometimes create messages through visual discourse, which might not be conveyed by verbal text.

Three major studies on EFL textbooks have been done in China. Tan (2012) examines *Advance with English* for senior high school students, and concludes that relevant layout and image ease literacy learning and increase students' participation in learning practice. He also asserts text and image interact with each other to make meaning, although text content is mainly repeated in image. Li (2011) compares two coursebooks of listening/speaking and reading/writing named *New Horizon College English* and concludes mode selection highly relies on different tasks and cognitive characteristics of different classes. Liu and Qu (2014) conduct pioneering research to explore and compare visual and verbal modes of two representative EFL textbook series (*Experiencing English* (*EE*) and *New Century College English* (*NCCE*)) for Chinese college students. The study

attempts to examine how to develop good multimodal EFL textbooks through multimodal discourse analysis. The findings suggest that *EE* and *NCC* incorporate visually and verbally coherent multimodal texts. The presence of interpersonal intersemiotic complementarity and multimodality within these texts enhances understanding. The study highlights the significance of organizing different modes effectively on the page. Furthermore, it emphasizes the importance of tailoring textbooks to learners' English proficiency and individual needs, as well as including a diverse range of multimodal resource to facilitate comprehensive comprehension.

Salbego et al (2015) examine how multimodality in three English textbooks of *Interchange* (Richards, 1999), *Interchange* (Richards et al., 2001) and *Cutting Edge Elementary* (Cunningham et al., 2005) may contribute to learning through visual texts. The analysis of images using Kress and van Leeuwen's visual grammar, which considers the metafunctions of representational, interactional, and compositional aspects, reveals that visual information aids beginners in comprehending the activity's content. The findings further establish that images not only complement the overall meaning of texts but also contribute to students' understanding of the activity, leading to enhanced learning outcomes. These results underscore the crucial role of multimodality in language learning environments and emphasize the importance of incorporating visual elements to support students' engagement and comprehension.

Marefat and Marzban (2014) adopt a multimodal analysis from a social semiotic perspective to investigate how image and text interact to represent gender identity in a national (*Iran Language Institute*) ELT textbook. They use Halliday's (1994) framework and Kress and Van Leeuwen's (2006) reading images model to examine the visual and verbal aspects of a dialogue. To do so, 4 items of contact, social distance, attitude, and narrative representation are examined in the image and the text. The findings show that although there is high interaction and support of image and text to depict gender identity, some complementary interactions are detected (Marefat & Marzban, 2014). In their study, they attempt to direct attentions to multimodal issues and increase awareness to take critical views towards underlying represented ideologies.

2.5.2. Captioned videos

Research on multimodal L2 input has also investigated the effects of captioned/subtitled TV viewing on comprehension. Montero Perez et al (2013) investigate how captioned videos improve L2 listening comprehension and vocabulary. The research also attempts

to examine if there is a relationship between learning from captioning and test type and proficiency level. The findings show captions improve listening comprehension and vocabulary acquisition and the test type moderates the effect sizes of listening comprehension. Rodgers and Webb (2017) investigate the comprehension of video among second language (L2) learners by examining the use of captions. The study involves a group of Japanese university students (N=372) who are instructed to watch ten episodes of an American television program, both with and without captions. Subsequently, comprehension tests are administered. The findings suggest that the episodes with captions moderately enhance L2 learners' comprehension scores across all episodes. The researchers conclude that captions can provide support for comprehension, particularly during challenging sections, especially at the initial stages of viewing. Therefore, it is recommended to use captions for L2 learners when they begin watching a series or encounter difficulties with comprehension in specific episodes. L2 learners may also become aware that they are still able to understand television programs, particularly series, even without the presence of captions.

The effects of captioned/subtitled TV viewing on L2 vocabulary learning have also recently attracted the attention of SLA researchers. For instance, Puimège et al (2021) investigate how textual enhancement affects L2 learners' learning (N=28) of multiword units while watching enhanced and unenhanced captioned videos. Eye-tracking technology is used to collect L2 learners' eye-measurements while processing the multiword units in the captions. The results of the recall pre- and posttests indicate that the learners' have acquired the target items. The findings also show that enhanced items enjoy more visual attention, longer reading times and more rereading. Furthermore, longer fixations on the items result in faster recall in the immediate posttest. The researchers acknowledge the visual attention improves the recall of multiword units appeared in captioned television and conclude that the difficulty of the item and the amount of attention are better predicators of learning, in comparison with textual enhancement.

Rodgers and Webb (2019) investigate the effects of watching full-length TV programs on incidental vocabulary learning. One hundred and eighty seven Japanese university students are asked to watch ten episodes of an American drama and complete two vocabulary pre- and posttests which assess their knowledge of word-families. The findings indicate that watching TV significantly enhances vocabulary knowledge among

learners, with a notable correlation between the frequency of vocabulary occurrences in the videos and improved learning outcomes.

Winke et al (2010) also study the effects of captions during video-based listening activities. Three short videos with and without captions are played for second- and fourth-year learners of Arabic, Chinese, Spanish, and Russian. Following the viewing of the videos, the learners are assessed through comprehension and vocabulary tests. Additionally, post-experiment interviews are conducted to gather their overall opinions regarding the captioned and non-captioned videos. The results reveal that participants perceive captioning as more effective than non-captioning. Moreover, captioning during the initial viewing of the videos demonstrates greater efficacy in enhancing participants' performance on aural vocabulary tests. Spanish and Russian L2 learners prefer captioning first while Arabic and Chinese prefer to see the captions second. The findings from the interviews also reveal that captions focus L2 learner's attention and allow them to explore the language.

Moreover, the effect of multimodal input (captioned/subtitled TV viewing) on grammar learning has been slightly studied among SLA researchers. As an example, Lee and Révész (2018) conduct a study to examine the impact of textually unenhanced and enhanced captions on learners' attention and acquisition of the present perfect versus the past simple in news reporting. The study involves 72 second language (L2) Korean learners who complete various treatment tasks while watching news clips. Subsequently, pretests, posttests, and delayed posttests are administered to assess their learning gains. The L2 learners are divided into three group of enhanced, unenhanced and no captions groups. Eye-movement measurements are examined for attentional allocations during the treatment. The findings show that both captioning and textual enhancement increase the L2 learners' learning and use of the target construction.

2.5.3. Multimodal reading texts

Although watching captioned videos involve multimodal input, text has only a supporting role (Pellicer-Sánchez, 2022). However, in the multimodal reading condition, text is the main source of information (Pellicer-Sánchez, 2022). Despite their prevalence in everyday life, the effect of static pictures in multimodal reading conditions have not been discussed extensively in L2 conditions (Seburn, 2017). Some studies have shown that static pictures in reading improves reading comprehension.

Omaggio (1979) conducts a study to investigate the impact of pictorial contexts on the reading comprehension of 664 individuals studying French at Ohio State University. The study involves the design of six pictorial and three textual contexts. In addition to the reading task, a writing task and a test of recognition knowledge are employed to assess comprehension of the text. The findings of the study reveal that the inclusion of illustrations enhances second language learning and facilitates the hypothesis-testing process.

Plass et al (1998) ask English-speaking college students in a German course to read German language story on computer. In order to understand the meaning of the key words in the story, students can see a translation on the screen in English (verbal annotation) or view a picture/video clip about the word (visual annotation), or both. The findings indicate that the students who have selected both visual and verbal annotations remember words translation better than the people who have only selected one or no annotation. Accordingly, the students who have selected their preferred mode of annotation can comprehend the story better. The findings also acknowledge a generative theory of multimedia learning, which believes learners select relevant verbal and visual information and later integrate and form them into a coherent mental representation.

Center et al. (1999) investigate the effects of visual aids on the reading and listening comprehension abilities of a specific group of year 2 students who face difficulties in their listening comprehension skills. The results indicate that in comparison with the control group, the experimental group improve significantly on the listening and reading comprehension. The researchers encourage the visual imagery training for listening comprehension from the early stages of language learning.

Bisson et al (2015) investigate the effect of multimodal stimuli on the attention during an incidental learning situation. Foreign language auditory words on their own with written native language translations, or with both written native language translations and pictures are shown to the participants. The participants are also asked to recognize translation equivalents the following day and a week later to evaluate their incidental acquisition of foreign language words. The findings show the participants receive higher scores for foreign language words presented with written native language translations, or with both written native language translations and pictures than for foreign language words presented in audio format only. However, the foreign words which have been presented with a picture during the incidental learning are recalled better after a week. Moreover, the time spent looking at the pictures during incidental learning is closely

related to the recall scores one week later. The findings emphasize the significant effect of visual information in incidental vocabulary acquisition.

Tang (1992) investigates the effect of graphic knowledge structures on intermediate L2 learners' comprehension of content knowledge and second language learning for academic purposes. Findings from the pre- and post-tests indicate that graphs develop comprehension and immediate recall. Moreover, L2 learners believe that graphic representation of knowledge fosters learning.

Serrano and Pellicer-Sánchez (2022) use eye-tracking technology to investigate young L2 learners' processing of the visual and verbal information in a graded reader in reading-only and reading-while-listening conditions. The findings indicate that although the readers receive similar comprehension scores in the two conditions, in the reading only condition the readers spend longer time on the text and in the reading-while-listening condition they process the images more often. Moreover, longer time spent on the text result in lower comprehension scores in both conditions.

In conclusion, extensive research demonstrates the positive influence of multimodality on language acquisition. The integration of various modalities creates an immersive learning environment that engages learners on multiple levels, leading to enhanced understanding and improved information retention. In the next section, we will explore the concept of multimodal cognitive processing in order to gain further insights into this topic.

2.6. Multimodality and cognitive processing

We acquire the knowledge in our everyday life through different verbal and nonverbal modes of communication. For instance, newspapers often use both image and text to communicate their news. Print advertisements also depict product's attributes and benefits using words and images. And it seems we are able to receive and understand the information from different stimulus modalities and form them in a way that allows us to remember the information and make decisions and judgements (Wyer et al., 2008). Previous studies on multimodal texts emphasize the mnemonic function of pictures in texts, in other words, text information is remembered better when it comes with an image than when there is no illustration (Boerma et al, 2016; Brody & Legenza, 1981; Levie & Lentz, 1982; Levin et al., 1987; Mayer et al., 2008; Nyström & Ögren, 2012; Seel, 1986; Shepard, 1967; Wolfe, 2018; Wright, 2010).

Pictures aid learning from texts by engaging separate cognitive subsystems for verbal and visual information processing (Paivio, 1986; Clark & Paivio, 1991). Dual coding enhances memory by encoding visuals in both imagery and verbal systems, providing an advantage over single coding when processing pictures in texts (Molitor et al., 1989; Florax & Ploetzner, 2010). Verdi and Kulhavy (2002) also mention verbal and visual information can be processed and kept simultaneously in working memory. Hence, it is easier for the reader to move between two codes (visual and verbal) and make connections and later retrieve the information. Mayer (1997) presents a multimedia learning model that integrates visual and verbal information processing. Separate mental models are created for words and images, which are later interconnected. This facilitates simultaneous access to both visual and verbal information in working memory (Baddeley, 1992).

Although Mayor's model (1997) explains visual and verbal processing to some extent, the parallelism of visual and verbal processing turns out to be problematic as visual and verbal texts are produced by different sign systems. Completing Mayor's model, Schnotz and Bannert (1999) put forward an integrative model of multimodal comprehension with an emphasis on the representational principle. This model is composed of descriptive and depictive representations. The descriptive model is made up of the text, the mental representation of the text surface structure, and propositional representation of its semantic content, whose interactions are based on symbol processing (Schnotz, 1994). The visual model comprises the picture itself, the perception of the picture and the mental model of the topic depicted in the picture (van Dijk & Kintsch, 1983). These processes are constructed through bottom-up and top-down interactions of cognitive schemata.

Graesser et al (1997) also introduce five types of representation while reading a text: a surface representation, a propositional text base, a mental representation, a communication level, and a genre level. Surface representation is the linguistic information of text such as syntax and lexical structures. The propositional text base reveals the semantic content of text. The mental model shows the referential content limited by text and world knowledge. The communication level represents the pragmatic content in communication between the author and the reader. And the genre level is about the class and function of the text. Multiple mental representations also occur in picture comprehension: surface representation, a mental representation, a propositional representation, a communication level and a genre level. The surface representation is the visual concept which is constructed in mind. The mental representation is the topic or

subject matter of image. The propositional representation is the information read and encoded from image. The communication level is the pragmatic content between the image maker and the reader. Genre level shows the class and function of image (Schnotz, 2002).

In the light of the review of literature, another reading model is suggested for multimodal texts: bottom-up, top-down, and interactive models. The bottom-up considers linguistic code at the lexical and syntactical levels (Abraham & Farías, 2017; Bruce & Tsotsos, 2006; Kress, 2010). Visual dimension is expanded in its relation to textual references, based on which the role of image and other visual elements such as layout and color are identified. Unsworth (2006, 2008) and Unsworth and Chan (2009) believe this metalinguistic awareness is vital to identify visual and verbal relations in educational materials for school literacy. In the top-down view, it is believed visual and verbal modes bring cultural concepts that add background to text (Abraham & Farías, 2017; Marefat & Marzban, 2014; Tinio, 2013; Wang & Pomplun, 2012). Hence, readers bring their background and previous knowledge to their reading. The activation of visual and verbal representations creates ground for inferences and text integration (Eitel & Scheiter, 2015; Hardiess & Weissert. 2021; Holsanova, 2014). Siegel (2012) mentions attention should be paid to cultural functions of different modes. Considering the interactive dimension of reading multimodal texts, the interplay between bottom-up and top-down processes are important.

Moreover, Serafini (2012) proposes his four roles of the reader-viewer as 1) navigator, 2) interpreter, 3) designer, and 4) interrogator. Readers as navigators need to decode written language and look into the design of the written text and realize the role of graphs, visual images, and charts they encounter in multimodal texts. Although written language has a temporal nature, there is not a predetermined way for readers to read multimodal texts (Kress, 2003) and particular features of visual grammar and compositional structures might lead their eyes in certain directions. Reader as an interpreter is involved in the act of interpretation which is the process of constructing and creating possible meanings and responses while interacting with multimodal texts. Readers should not thus be viewed as passive individuals who discover only a single main idea (Faust, 1994). To reach their final understanding, they draw on their prior knowledge of culture, language, and life experiences (Smagorinsky, 2001). It should be considered that images like languages are not neutral or objective representation of reality (Gombrich, 1961). And viewers' society have trained them to see things in specific ways (Berger, 1972). Emphasizing the fact that

images like written texts are not created in a vacuum, Hartman (1992) mentions "image is not a unitary, seamless whole, but can be fragmented and atomized into constituent elements, and these elements represent a life of their own and that of others" (Hartman, 1992: 297). During the process of interpretation, readers draw upon a wide range of their life experiences. Hence, people of the same culture and society are expected to reach a relatively same interpretation of the text and image. Reader as designer creates meaning from multimodal texts (Spivey, 1997) and decides and designs how to read and attend the text and constructs their own experience from it. In other words, the text to be read is not ready-made for the reader, but its interpretations are designed through the act of reading (Serafini, 2012). Kress (2010) highlights the sequential process involved in text creation. Initially, the designer conceptualizes the task, followed by drawing upon their semiotic knowledge to guide their choices. Subsequently, by considering the social context, they analyze the text to shape the intended meaning. Consequently, when engaging with multimodal texts, readers consciously consider them within their respective contexts. However, since frames create both temporal and spatial factors and separate text from its context, the reader is able to decide what part of text or context should be considered while reading (Kress, 2010). Reader as interrogator considers cultural part of the meaning (Smagorinsky, 2001). Based on cultural theories of meaning, reading a text has both personal and public meanings. Considering this view, reading is a social practice which is constructed in a social context through power relations in the context and the readers' identity. Reader as interrogator should consider multimodal texts and their intended audience while constructing meanings through reading (Serafini, 2012).

Moreover, building upon multimodal mental models, Mayer and Moreno (2003) suggest some effective features of multimodal texts which ease cognitive processing and enhance learning. First and foremost, verbal and visual information should be coherent with some semantic overlap. Second, visual and verbal information both should enter the working memory at the same time to develop interactions between them. Third, before semantic processing of visual and verbal information, relevant knowledge needs to be activated (Mayer & Moreno, 2003). In order to activate prior knowledge, meaningful associations to their referents should be used. Fourth, in order not to overload cognitive system, visual and verbal information should not use the visual channel to enter the working memory. Fifth, visual information should be designed in a way that can be easily recognized. And finally, visual information should be prepared according to the purpose of learning and teaching. Verdi and Kulhavy (2002) mention in case visual and verbal

information cannot be presented at the same time, it is better that the picture is presented first because processing pictures takes little space in the working memory. However, if the text occurs first, the capacity of working memory is almost used, which leaves little space for processing the picture (Verdi & Kulhavy, 2002). Moreover, a text never gives a detailed description of a subject to allow for a specific envisioning of a visual image. Hence, the mental model constructed only by text is different from following image and might interferes with the constructed visual mental model. To avoid this interference image is better to be presented first (Verdi & Kulhavy, 2002).

In conclusion, the role of visual displays in fostering learning through multimodal texts has been acknowledged (Pellicer-Sánchez, et al., 2020). Multimodality emphasizes that learning occurs when various semiotic modes interact to create meaning. However, to gain a comprehensive understanding, it is crucial to analyze eye movement measures during reading. This analysis provides valuable insights into how individuals interact with multimodal texts, including their attention allocation, cognitive processes, and reading strategies. By studying eye movement patterns, researchers can identify the modalities that capture learners' attention more effectively and facilitate deeper engagement with the material. This knowledge enables the development of more effective and engaging multimodal learning materials and strategies, ultimately optimizing language learning experiences. Further exploration of this topic will be discussed in the next section.

2.7. Multimodality and eye measurements

Comprehension results from the development and coordination of different reading competencies such as word recognition, reading fluency and syntactic processing (Rayner et al., 2006). One way to observe and inspect reading comprehension is monitoring eye movement measures during reading, which provides useful information about moment-to-moment comprehension processes (Rayner et al., 2006). This section starts with reviewing some basic characteristics of eye movements including regressions and number and duration of fixations and continues with some discussion on comprehension processes. According to Castelhano and Rayner (2008), eye movements are related to the amount of cognitive and perceptual effort to process the input and indicate the difficulty and complexity of what people are reading.

Regressions are inseparable parts of all fixations. Around 15% of the words in text are fixated again due to resting in a wrong place in the word and happen immediately to the left of the last fixated word (O'regan & Levy-Schoen, 1987; Rayner & Pollatsek, 1989;

Vitu & McConkie, 2000). Regressions help the eyes to get to the optimal viewing location and may also indicate difficulty in text and comprehension. Frazier and Rayner (1982) assert that when readers face a word which goes against or ambiguates their prior information, they often make regressions and skip fewer words (Castelhano & Rayner, 2008). Therefore, failures in comprehension are likely to cause more regressions (Frazier & Rayner, 1982).

New information is encoded during fixations, typically lasting around 200–250 milliseconds (Rayner, et al., 2006). Moreover, when readers see difficult words (e.g., low-frequency words) or syntactically complex sentences, they fixate longer (Rayner, et al., 2006). It is assumed that as text gets more difficult, readers have more and longer fixations and more regressions (Rayner, et al., 2006). Regarding the multimodal texts, on the one hand, fixation durations tend to be longer on the pictures due to the fairly wide field of view in the images (Rayner, 2009). On the other hand, there are more fixations on the text because the information is presented densely and eyes need to recognize letter information (Rayner, 1998).

In order to integrate the information coming from different sources in the process of meaning-making, the readers usually start with early and brief looks at text which are followed immediately by some gazes on image (Carroll et al., 1992; Hegarty, 1992a, 1992b; Rayner et al., 2001; Hardiess & Weissert, 2021). This scan phase is done by readers to get the 'gist' of a scene very early in the viewing process and find their entry point (Holsanova, 2014). Moreover, theoretical analyses and empirical findings indicate that different segments of a text are not processed equally (Hy¨on¨a & Lorch, 2004). A text segment which includes a new topic receives more attention than a segment which continues the same topic. Accordingly, reading and processing introductory sentences are cognitively more demanding because readers need to identify the general idea of the text (Hy¨on¨a & Lorch, 2004). During the first-pass fixation time, when the initial processing of materials happens, reading is more intentional because the reader attempts to process, combine and comprehend the information (Hy¨on¨a & Nurminen, 2006). The frequency of movements between the visual and verbal information indicates that readers attempt to integrate the information and build a representation of the text (Schwonke et al., 2009). Later, readers use the semantics of text to guide their eye movements and make fixations on the relevant elements of image (Hayhoe & Ballard, 2005; Hardiess & Weissert. 2021). Final fixations wrap up the reading process during which the number and duration of fixations decrease (Pieters et al., 1996).

In summary, eye-movement measurements provide valuable insights into readers' cognitive engagement and the integration of visual and verbal information (Rayner et al., 2008; Rayner & Castelhano, 2008; Rayner, 2009). This understanding sets the stage for the subsequent chapter, which delves into how age influences multimodal literacy by considering cognitive development and perceptual abilities. The connection between multimodality and age holds significant implications for education and the promotion of learning experiences across different age groups.

2.8. Multimodality and age

Studies indicate that attention to text becomes more intensive with age and higher proficiency (Roy-Charland et al., 2007). In order to investigate the viability of these results across reading development, Roy-Charland et al (2007) monitor eye movements of children from kindergarten to Grade 4 in shared book reading experiences with three difficulty levels. The results show that as grade level increases, children spend longer time and have more reading-like saccades on texts. The study further establishes a correlation between the difficulty level of the material and attention to text, revealing that readers dedicate more time to texts that align with their reading abilities rather than those that surpass them.

In another study, including five native French speakers in 48- 61 months' age range, Evans and Saint-Aubin (2005) conduct research using an EyeLink headband to investigate where in storybooks young children fixate during reading. The findings suggest that children spend little time reading the text. The length of the accompanying text also positively affects the fixations on the illustrations, however, the fixations on the text are not correlated with the length of the written text (Evans & Saint-Aubin, 2005).

Tragant and Pellicer-Sánchez (2019) examine how two groups of young EFL learners in 10-11 age range use and integrate different sources of input in an illustrated storybook with audio support and a video with foreign language captions. The findings from the eye-tracking experiment show that the learners spend most of their time processing the text than the image in both conditions. Moreover, storybooks encourage regular reading behaviors while the video results in various reading patterns. The findings also demonstrate that storybooks and subtitled videos foster students' engagement with reading.

Both young and adult L2 learners rely on written text in processing information more frequently than image (e.g., Serrano & Pellicer-Sánchez, 2022; Tragant & Pellicer-

Sánchez, 2019). Similarly, proficient adult readers allocate most of their attention to text (Johnson & Mayer, 2012). Chang and Choi (2014) examine the effect of emotionally engaging text segments and attention-grabbing pictures on adult L2 learners' reading comprehension and on their information retrieval from the text. The results reveal that more attention to emotionally engaging text segments leads to worse comprehension. Moreover, attention to the pictures facilitates neither recall nor comprehension. Similar findings have emerged from studies involving young L2 learners (Serrano & Pellicer-Sánchez, 2022), indicating a negative correlation between comprehension and the time taken for text processing. Additionally, research suggests that the presence of auditory input enhances the processing of images in multimodal texts, both for young and adult L2 learners (Serrano & Pellicer-Sánchez, 2022; Pellicer-Sánchez et al., 2021).

Muñoz (2017) uses an eye-tracking methodology to explore the effects of age and proficiency on the reading behaviour of foreign language learners in three age groups (children, adolescents and adults) when watching audiovisual material with L1 and L2 subtitles. The results indicate that while no differences are found between two older groups, children make more fixations and spend longer time on subtitles and skip them much less than adolescents and adults. One reason could be the children's slower reading speed (Häikiö et al., 2009) and smaller perceptual span which result in more fixations on the subtitles (Rayner et al., 2010). Second, children's higher number of fixations might indicate that the task is cognitively more demanding and difficult and they have developed a limited proficiency. Children spend more time on subtitles in L2 than in L1 which might explain that reading in L2 is more challenging for them while adults skip more L1 than L2 subtitles. The findings may indicate that adults have a higher proficiency and find L1 subtitles redundant.

Moreover, An et al (2017) use an eye-tracker to investigate children and adults' attention to Chinese print and pictures. The results show although the participants spend the same amount of time looking at the pictures, the adults pay more attention to Chinese print and look at the print sooner than the children do.

In sum, as L2 readers age and develop their language proficiency, their attention to text rather than image and their reading speed increase (Calabrèse et al., 2016; Chen et al., 2019; İyigün et al., 2018; Solan et al., 1995). However, most of the studies investigating multimodal reading processes have been conducted with preschoolers, young adults and adults. Therefore, more research is required to compare and examine how L2 learners develop their multimodal reading processes in their middle adulthood,

when cognitive function might start to decline (Cullum et al. 2000; Eckert et al. 2010; Grady & Craik, 2000; Park et al. 2002). The subchapters on multimodality sets the stage for the exploration of social media inside and outside of school. By understanding the interplay of various modes of communication, we gain insights into how social media platforms impact communication patterns, social interactions and reading habits. Connecting multimodality with social media provides valuable insights for educational practices and digital literacy.

2.9. Multimodality and social media

2.9.1. Social media out of classroom

In today's modern era, social media has become an essential part of our daily routines, serving as a crucial platform for online communication. Users actively engage with social media to interact, share content, stay informed, and make informed choices. The widespread adoption of social networking sites like Facebook, Instagram, and Twitter has made them integral to the majority of society. These platforms provide a rich source of multimodal data, allowing users to express themselves and create content using diverse textual and multimodal elements. As social media continues to permeate our lives, extensive research has been conducted to explore the reading strategies employed by individuals within these dynamic digital environments.

Chen et al (2015) investigate the multimodal content of image tweets and the interactions between image and text which have been widely used around the world. Visual-Emotional LDA is used to investigate the image-text correlation from different perspectives in English and Chinese tweets. The findings indicate the significant role of both visual and emotional elements in image-text relations.

Morin et al (2019) investigate how image and text interact to produce discourse on a humanitarian crisis, Ebola. The data from 182 tweets which include image are quantitatively and qualitatively analyzed. The results emphasize the importance of image in building the discourse on social media. The interdependence of text and image on social media and the important supporting role of image in making meanings help their users express their intent and emotions.

Instagram is an online image sharing social media in which image and text interact to create the content so it is important to investigate how visual and verbal information integrates to create the final meaning. To do so, Pillai (2019) studies how National Geographic conveys the overall meaning through image and text narratives on their

Instagram posts. First, photo elicitation interviews of four participants are carried out to collect data after they are shown some photographs of National Geographic without the caption. Second, Mieke Bal's (2017) framework of narratology is used to analyze the captions of the photographs quantitatively. Third, the combined effect of image and text is analyzed to answer the research question. The results indicate that image arouses emotions and creativity while text creates information for the reader. Pillai (2019) recognizes the significance of the symbiotic relationship between image and text on Instagram in constructing meaning.

Kruk et al (2019) argue that in order to understand the author's message from Instagram posts, we need to realize the interplay between text and image. An example of this can be observed in the phenomenon known as "meaning multiplication," where a caption on Instagram deliberately creates an ironic contrast with the accompanying image, amplifying the overall meaning. According to Bateman (2014), image and text can create a new meaning which is more than their literal meanings. Hence, multimodal platforms such as Instagram and Facebook provide the ground for their creators to adopt visual and textual modalities in a variety of ways. The widespread use of multimodal texts on social media makes interpreting the author's message an important task for the readers. Existing studies have examined the reading of images accompanied by text captions (Chen et al., 2015; Faghri et al., 2018), often treating the relationship between the two as asymmetric, where one mode is considered subordinate to the primary mode (Bateman, 2014). However, scholars in semiotics and computer science have mentioned that image-text relation is not that simple such as mere addition of one mode to another, but it is more of meaning multiplication, which helps meaning making process through image and text interactions (Bateman, 2014). Kruk et al (2019) conclude that both image and text are helpful to use on Instagram but the degree of its helpfulness differs based on the type of image-text relations. In the semiotic category when image and text are different, there is the greatest gain, followed by the condition in which image and text are additive. And when image and text are independent and provide no new information, they help the least (Kruk et al, 2019).

Ståhl and Kaihovirta (2019) do a case study with one student in upper secondary school in Finland to investigate the visual communication when the student interacts with longer-lasting images on social media. The data include 41 images that the participant has shared and liked on Tumblr and Instagram. The data analysis from the interviews indicates that the participant interacts with the image as a persona by activating visual and technical

competencies and knowledge of social norms and self. These different competencies need to be taken into consideration in education to harness the skills they might need in the future world.

Vergara et al (2021) investigate the participants' incidental consumption of news on social media. Eye-tracking technology is employed to study individuals' visual engagement with posts on Facebook. The analysis focuses on various aspects of visual interaction, including the duration of content viewing, eye movements on the screen, as well as the frequency of likes, comments, and shares of news. The findings indicate the incidental consumption of news is widespread among participants. Participants use different eye-movements to read the news, which mostly involve a superficial processing of multimodal stimuli.

In conclusion, as the use of smartphones, Wi-Fi networks, and social media platforms like Instagram and Facebook continues to grow, schools have recognized the limitations of traditional phone-off policies. Acknowledging that students are digital natives, educators are now embracing the integration of social media into the classroom to enhance student engagement and participation. This fresh approach to education has prompted researchers to delve into the learner's behaviors and habits on social media, which will be thoroughly examined in the following section.

2.9.2. The effect of social media on academic reading habits

Reading, a powerful and indispensable way of learning, ensures academic and professional success (Hassan et al., 2021) and plays a key role in second or foreign language education (Grabe, 1991). Additionally, developed L1 is a pre-requisite to all kinds of learning, including foreign language learning, and it has positive correlations with L2 achievements (Ardasheva, 2016; Bernhardt, 2011; European Commission, 2015). It is thus essential for second language learners to form good reading habits which support life-long learning experience and activate their thinking power. Moreover, books specially fictions help readers look for and build their identity and experience and live the culture (Erikson, 1974; Sarland, 1991). Reading is greatly appreciated as the best source of knowledge and many studies have thus been carried out to investigate reading habits (Fusco, 1986; Celik, 2019; Platt, 1986; Scales & Rhee, 2001). In light of unprecedented technological advancements, traditional reading habits are being progressively replaced by modern technological devices. The internet's stimulating environment has led to decreased engagement with books and other reading materials, as students are constantly

engrossed in social media networking. The widespread availability of the internet, smartphones, tablet computers, and other mobile devices, along with on-campus Wi-Fi, has created a generation that is consistently "plugged in" (Baird & Fisher, 2006) and is often "powering down", upon entering the classroom (Prensky, 2001). Unfortunately, this shift has resulted in a decline in reading habits among children, teenagers, and adults (Hopper, 2005; Johnson-Smaragdi & Jönsson, 2006; Marsh, 2005), limiting their reading activities to comments, posts, and the sharing or submitting of daily events. Thus, it is essential for language teachers, educators, and researchers of linguistics to highlight the positive effects of social media on second language learning and investigate its negative impacts. The valuable findings can help teachers teach students, on the one hand, how social media can be used to achieve learning and, on the other hand, attempt to increase their reading book sensitivity. Although social media have advantages such as allowing individuals to use more than one mode to express understanding (Jewitt, 2009**;** Oldakowski, 2014) and sharing educational materials and subscriptions to educational pages and channels, their domination led to a decrease in second language reading among instructors and students and affected their reading habits both inside and outside educational settings (Qenaway, 2019).

A lot of research has acknowledged the profound effect of social media on the reading habits of second language learners. For instance, Mirza et al (2021) explore the application of technology and social networking sites by the students of Pakistan. The findings reveal that the majority of students have prior access to smartphones than desktops, laptops and tablets. Most of the students prefer electronic books for academic reading and use social networking sites for fun.

Qenaway (2019) investigates the effect of social media on second language reading habits among 30 Arab students at secondary stage. The results validate that students demonstrate a willingness to engage in reading books or stories that incorporate pictures outside of the school environment, typically once or twice a week. Among students, Instagram and Facebook emerge as the most preferred applications. The author endeavors to increase students' awareness regarding the use of social media and recognizes the substantial impact that reading habits have on their second language proficiency and academic accomplishments.

Arumugam et al (2019) investigate the undergraduates' perceptions towards WeChat application and study whether it can improve critical thinking and comprehension skills of English as a second language. The results indicate that ease of use and user-friendliness

of WeChat improve students' critical thinking, communication and reading and writing skills.

Rafiq et al (2019) conduct a quantitative study to investigate the effects of social media on the reading habits of 430 students in Pakistan. The results indicate that although social media positively affect reading habits, they distract students during the time of study. The results also show that Facebook, WhatsApp, Google and YouTube are used daily by the students. The participants also state that social media have provided an environment for them to learn collaboratively. The researchers suggest that the educational settings should launch mobile applications which contribute to students' reading and academic activities and hold workshops about social media usage to introduce the positive aspects of social media and develop reading habits among the students.

George (2018) highlights the numerous learning opportunities that technology and online resources provide for language learners to enhance receptive skills, reading and listening. Social media foster language acquisition through meaningful engagement and outlet (Chartrand, 2012). A survey is conducted among a few college students and their teachers in Kerala to investigate the impact of social media like Facebook and WhatApp on listening and reading skills. In conclusion, the researcher emphasizes the importance of teacher guidance in improving students' receptive skills. Access to materials on social media platforms enables students to become independent learners, allowing them to learn at their own pace. However, while social media offers numerous benefits, issues such as information overload and the need for teachers to possess technical and digital competencies to effectively integrate various mobile applications into language learning should be taken into account by schools that incorporate technology into their curriculum. Additionally, providing appropriate feedback is crucial in maximizing the educational potential of these platforms.

Anankulladetch (2017) investigates English language learners' experience and performance in an educational setting where social media are used as a learning tool. The study includes a cohort of 37 ESL students who are currently enrolled at an adult school in Central California. Their feelings about the effect of social media on their academic performance are measured through the Student Perception Toward Social Media in Education survey. The findings reveal that the students in the treatment group enjoy online blogging discussion, greater engagement and learning performance and higher scores than the students in the control group. The students mention that social media help

them build conversations, share ideas and increase the engagement through the comments written by their friends.

Owusu-Acheaw (2016) employs a questionnaire to examine the influence of social media on the reading habits of the students at the University of Koforidua Polytechnic. The findings reveal that most of the respondents visit social media sites every day spending between 30 minutes to 3 hours. The participants also indicate that the use of social media negatively affects their reading habits. Hence, faculty staff need to encourage students to spend more of their time and energy on educational purposes.

Rajab and Ali-Sadi (2015) also conduct a quantitative study to investigate the reading habits of Saudi EFL students in their first and second language at a university-level preparatory year program. The results indicate that the students are not interested in 'academic reading' in both L1 and L2, however, they are actively engaged with reading on social media.

Derakhshan and Hasanabbasi (2015) highlight social media as an important tool with authentic materials to improve second language learning outside the classroom. With this aim in view, the researchers attempt to describe the role of social networks in students' learning English out of classroom. Social networks provide a ground for students to interact and share their ideas and use online tools to develop their learning skills, activate their passive English knowledge and develop communication skills.

Wang and Vasquez (2014) investigate the effect of target language use in social media on Intermediate-level Chinese learner's writing performance. The research involves two groups participating for one semester: an experimental group that writes weekly updates and comments on a Facebook group page, and a control group that does not post on the Facebook page. The findings suggest a significant difference in the amount of writing produced by the two groups. While there is no notable distinction in the initial writing task, the experimental group generate a significantly higher number of Chinese characters compared to the control group in subsequent writing tasks. However, when it comes to the quality of writing, there is no significant differences observed between the two groups across all three writing tasks.

Moreover, Noor (2011) examines the reading habits and preferences of a group of EFL graduate students in Malaysia through a questionnaire and quantitative analysis. The results indicate that the students' top preference for reading material is internet-based content. Furthermore, the EFL learners engage in reading materials for both pleasure and

academic study, with materials in their first language being favored for leisure reading and materials in their second language being favored for academic purposes.

In educational settings, Facebook, an online social platform, can improve interaction between teachers and their students (Sturgeon & Walker, 2009), encourage students to do their assignments (Kitsis, 2008), foster second language learning (Abidin et al., 2010), develop learning language skills (Depew & Skerrett, 2011), improve students' interaction among teenage readers and publishers (Hamilton, 2009), encourage reading literature among them (Walker, 2010), develop competition among L2 learners to learn incidental vocabularies (Blattner & Fiori, 2009), lift up communication barriers between learners (Schaffhauser, 2009) and enhance students' participation in learning process (Mazman & Usluel, 2010).

Computer media can also foster L2 learning (Pennington, 1989), facilitate L2 learners' communicative competence (Lowry et al., 1994), improve their writing skills (Hertel, 2003), boost the confidence of shy students and help them express themselves (Kroonenberg, 1995), encourage the students who do not like to speak in the class to discuss the topics electronically (Keln, 1992), eliminate barriers set up by students' age and economic problems by studying at home via computer media (Selfe, 1990).

In sum, here has been a decline in reading books and most of students read books to prepare for the examination (Chauhan & Lal., 2012). In addition, the mere utilization of social media alone does not lead to meaningful learning experiences. Students require guidance from teachers on how to effectively use these applications for educational purposes. Teachers need to use the present EFL/ESL reference resources to plan and design collaborative activities on social media. Education institutions should also consider the latest social media in education for curriculum design to offer more engagement and enriched learning experience to deal with students with different needs and preferences. Moreover, assessment tools should be developed to grasp students' broader understanding of multimodal texts (Burgess, 2016; Kress & Selander, 2012; Svärdemo Åberg & Åkerfeldt, 2017), which could consequently determine what happens in classrooms (Alderson & Wall, 1993) and foster multimodal learning. The following section aims to consolidate and provide a comprehensive overview of the topics that have been discussed thus far. By synthesizing the information presented earlier, a foundation can be established to formulate a compelling research question and delve deeper into the subject matter.

2.10. Summary and research questions

Reading has long been multimodal including various modes such as font size and letter size to hyperlinks, audios and images in recent years (Abraham & Farías, 2017). The advent of technology in the digital age allows for creating texts of varied formats to present language and has changed "the how and the what of reading" (Abraham & Farías, 2017: 59). This shift to multimodal texts as the primary source of communication in the digital age adapts and transforms traditional reading models and incorporates the semiotic approach to examine reading multimodal texts and develop literacy practice (Abraham & Farías, 2017).

L2 are also actively involved in multimodal texts (written or spoken texts and images), which are believed to improve reading comprehension and facilitate learning (Pellicer-Sánchez, 2022). Despite the extensive use of multimodal texts in second language learning, little research has been done to investigate how L2 readers create meaning from multiple sources of semiotics in learning materials and how these affect the readers' attentional demands and cognitive processes (Pellicer-Sánchez, 2022). Thus, reception studies should be carried out which help the researchers understand how readers exploit different semiotic resources to reach their final understanding and whether all the modes equally affect and help learning and reading comprehension.

Moreover, different age groups demonstrate varying preferences and benefits when it comes to social media and multimodality. Adults generally favor EFL/ESL textbooks and traditional classroom activities, emphasizing the significance of reading and comprehending diverse texts, including those with multiple modes of communication. On the other hand, younger learners, such as teenagers and young adults, tend to take a more casual approach to language learning. They possess strong technological skills and devote substantial time to mobile phone usage, actively participating in social media and consuming digital content. Therefore, it is essential to examine the multimodal reading approaches employed by L2 learners across various age groups in the era of technology. Additionally, it is important to explore their cognitive processes involved in reading and comprehending multimodal texts.

This research seeks, on the one hand, to investigate the effect of age on second language learners' reading habits and modal preferences and, on the other hand, to explore how they read multimodal texts. There are four main questions that the present study attempts to answer. The questions are the following.

1- What are the prevalent reading habits and modal preferences among Hungarian L2 learners of English?

   We will look at the participants' responses to the online reading habits questionnaire, multimodal reading test and follow-up questions.

2- How does age affect the reading habits and modal preferences of Hungarian L2 learners of English?

   Again, we will look at the participants' responses to the online reading habits questionnaire, multimodal reading test and follow-up questions in different age groups.

3- What is the typical sequence in which Hungarian L2 learners of English approach the reading of multimodal texts?

   We will see the results from the eye-tracking experiment to investigate the participants' eye measurements.

4- How do intersemiotic relations in the multimodal texts affect modal preferences?

   We will integrate the results from the online and eye-tracking experiments.

In the next chapter, the systematic approach employed to address the research objectives and answer the research questions will be explored.

CHAPTER 3

RESEARCH METHODOLOGY

The current chapter presents the methodological considerations of the pilot research. The online experiment investigates the effect of age on L2 learners' reading habits and modal preferences. The eye-tracking experiment attempts to explore how L2 learners read multimodal texts. This chapter deals with all the methods, procedures and analyses of the separate studies. In section 3.1, the research design is presented very briefly. Section 3.2 discusses the demographics of the participants. In section 3.3, the research instruments and methods adopted in the online and the eye-tracking experiments are introduced and discussed in details. In Section 3.4, English language tests and participant selection and sampling are examined. Section 3.5 provides an explanation of the data collection procedures. Finally, in Section 2.6, the methods of data analysis are discussed.

3.1. Research design

The current pilot study takes in essence a mixed method approach. It adopts various data collection techniques to extensively explore second language learners' practice of multimodality. The present research is composed of two experiments (Table 3) which attempt to qualitatively and quantitatively investigate L2 learners' multimodal reading.

Table 3. Research design

Purpose	Data collection instruments	Data analysis instrument
Online experiment The effect of age on L2 learners' general reading habits and modal preferences	Online reading habit questionnaire Online multimodal reading test Online follow-up questions	SPSS 25 Microsoft Excel 2013
Eye-tracking experiment L2 learners' multimodal reading strategies	Eye tracking system Multimodal reading test	SPSS 25 Microsoft Excel 2013 OGAMA 5.1

Each study is discussed in details in the following sections.

3.2. Participants

In this section, the demographic information of the participants in the online and eye-tracking experiments are discussed.

3.2.1. Online experiment

Seventy participants attended the online experiment. They were divided into two age groups of 13-30 (Group A) and 31-46 (Group B) participants. There were 30 and 40 participants in Group A (M=21.53 SE=1.13, SD=6.19) and B (M=39.76, SE=.94, SD =5.15), respectively.

Figure 16 and 17 show that in Group A the participants were in 13-17 (14 participants) and 24-29 (16 participants) age range and their age was mainly distributed around 16 and 27-29, respectively. However, Group B included a wider age range between 31 and 46 and the participants were mainly between 31 and 37 (15 participants) with a peak at 32 and between 38 and 46 (25 participants) with a peak at 46. Group A consisted of participants who were more likely to be in learning environments such as school, college, or university, or had recently graduated from them without a significant gap in their educational experiences. On the other hand, Group B included participants who were more likely to have embarked on completing their education, leading a more settled life, and choosing a career path.

Figure 16. Normal curve on histogram for the age range in Group A

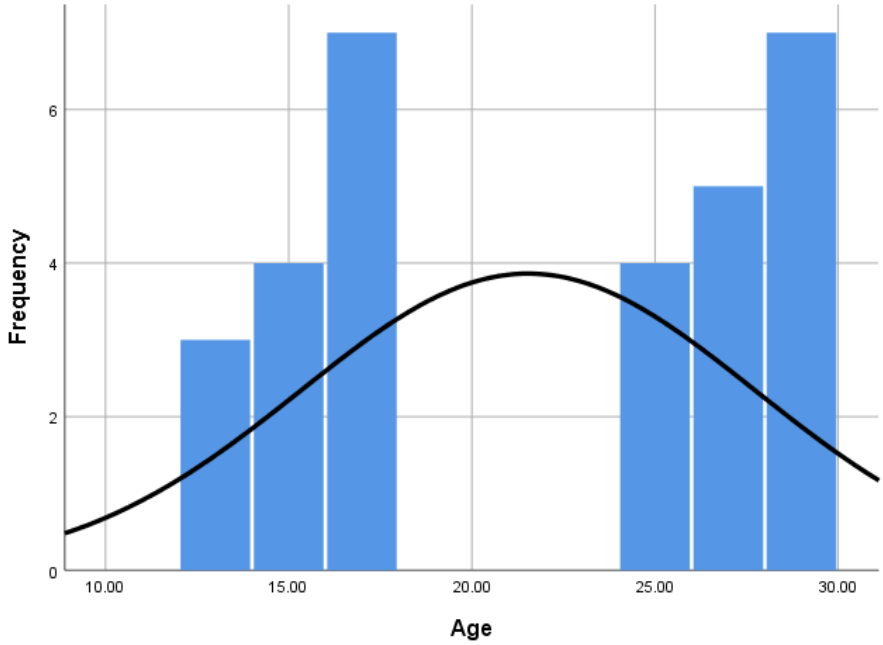
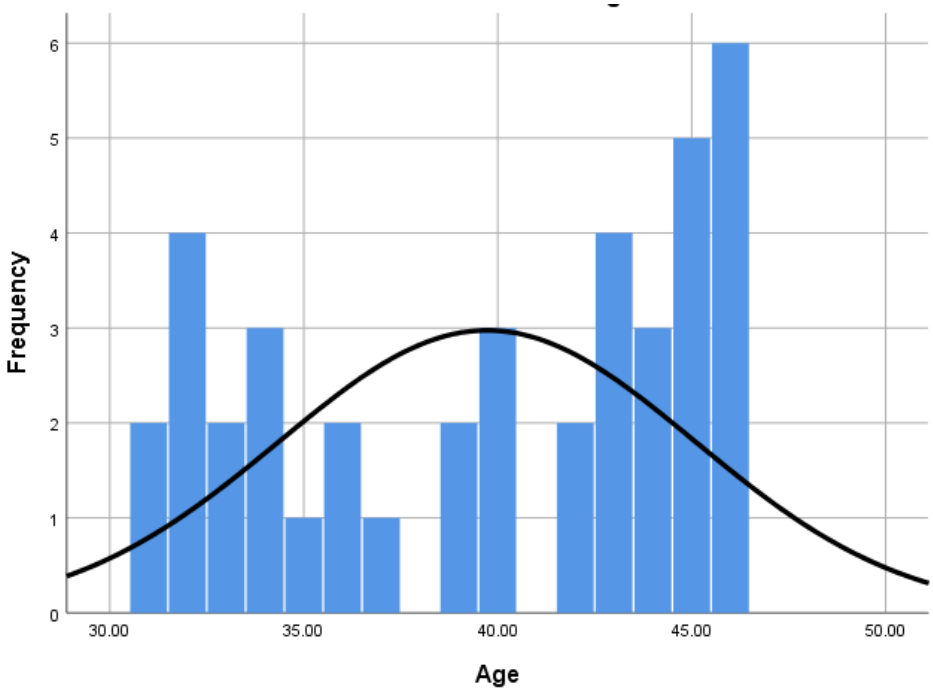


Figure 17. Normal curve on histogram for the age range in Group B



Considering the gender of the participants, 38 males (54.3%) and 32 females (45.7%) created the research group. Table 4 also shows that there are more male participants in Group A (63.3%) than Group B (47.5%).

Table 4. Number and percentage of male and female participants in Group A and B

		Male		Female	
Group	Age range	Number	Percentage	Number	Percentage
A	13-30	19	63.3	11	36.7
B	31-46	19	47.5	21	52.5

Within Group A, the male participants constituted the majority at 63.3%, surpassing their female counterparts who accounted for 36.7% of the group. In contrast, in Group B, the distribution was almost evenly split, with males comprising 47.5% and females making up 52.5% of the participants. All the participants demonstrated the same level of proficiency (B1) at the time of the research. None of the participants studied English as their university major.

3.2.2. Eye-tracking experiment

Nine Hungarian participants completed the eye-tracking experiment. Seven participants were in 19-22 years age range studying at university. Two Hungarian participants, who were studying at a language school, were 38 and 42 years old. They were working at a company at the time of the research. There were 5 males and 4 female participants in the

eye-tracking experiment in total. All the participants demonstrated the same level of proficiency (B1) at the time of the research. None of the participants studied English as their university major.

3.3. Methods and instruments of data collection

In this section the tools used to collect, measure, and analyze the data related to the research interests are discussed.

3.3.1. Reading habit questionnaire

The questionnaire employed in this research is an original test that was specifically developed for this study to investigate the L2 learner's general reading habits (see Appendix A). According to Shen (2006) how often, how much, and what students read create their reading habits. Therefore, the questionnaire includes 13 multiple choice questions which ask about the frequency and amount of time spent on reading print newspapers, magazines, books and comics (4 questions), their electronic versions (4 questions) and on networking, looking at pictures, reading texts and watching videos on social media (5 questions). The 13-item questionnaire on a 4-point Likert scale (never, sometimes, often and regularly) offers valuable insights into the manners with which students may or may not approach reading. The questionnaire was developed in Hungarian in order to avoid any misunderstanding of the language.

3.3.2. Multimodal reading test

In this section, the framework and the process of developing the multimodal reading test are discussed in details. The development of the multimodal reading test aimed to delve into the modal preferences and reading processes of L2 learners by incorporating various modes (image and text).

3.3.2.1 First draft of the multimodal reading test

Adopting the frameworks of Martinec and Salway (2005) and Engebretsen (2012) (See section 2.3 for more details), 10 types of visual-verbal relation were selected to develop the first draft of the multimodal reading test instrument:

- Image-text-independent (ITI)
- Image-text-complementary (ITC)
- Text-more-general-than-image (TGI)
- Image-more-general-than-text (IGT)
- Image-subordinate-to-text (IST)
- Text-subordinate-to-image (TSI)
- Image-enhancement-place (IEP)
- Image-enhancement-reason (IER)
- Text-enhancement-time (TET)

- Image-text-tension (ITT)

English File textbooks are renowned for their captivating and authentic materials that mirror real-life language usage. They include a wide range of subjects, contexts, and language styles, ensuring the practicality and relevance of learning. The textbooks' user-friendly format, clear instructions, and appealing visuals enhance comprehension and establish an enjoyable learning environment. Through interactive activities and multimedia elements, they encourage student engagement and foster communication skills. The widespread recognition and adoption of English File textbooks in language learning institutions influenced my choice to include this series in my research.

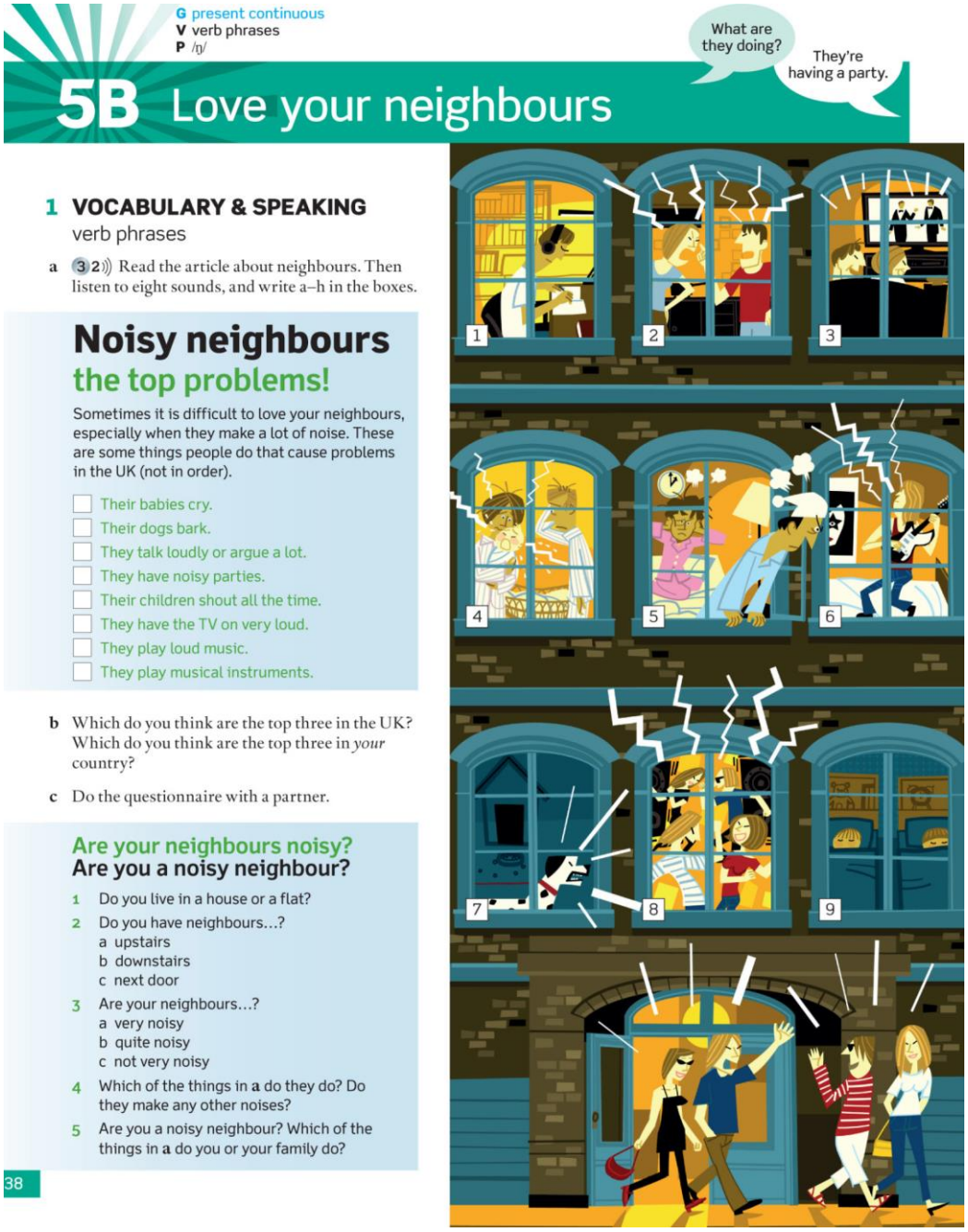
To develop the initial version of the multimodal reading test, I examined the visual and verbal aspects of multimodal texts in English File textbooks (3rd Edition), using frameworks proposed by Martinec and Salway (2005) and Engebretsen (2012). The multimodal texts which carried one of the 10 image-text relations were selected to create the first draft of the test (see Appendix B). Thirty multimodal texts, which carried at least one of the image-text relations, were collected during the first browse of the textbooks. An example of image-text-tension was not discovered in English File textbooks. The multimodal texts were later labeled and classified according to their type of image-text relation. Considering the number of image-text relations a multimodal text embodied, it might have received different labels. From each group, one multimodal text, which represented the type of image-text relation the best, was selected and in total 9 multimodal texts created the first draft of the multimodal reading test. One question with two answers was also added to each text. One answer included verbal (verbal responses) and the other one embodied visual (visual responses) representations of the information in the multimodal text.

Creating the first draft, I realized that the multimodal texts did not carry the same form of information. For instance, some multimodal texts carried more than 2 image-text relations. Figure 18 shows a multimodal text in which the text is subordinate to the image because the word ‘it’ is only expanded in the image (e.g. snake, flight, spider, etc). On the other hand, the image exemplifies the text by including a group of phobias. The text also includes the phrase ‘for years’, which enhances the image temporally.

Figure 18. Fear and phobias (English File (3rd ed), pre-intermediate: 70)

Another example could be the image-text proportion. In Figure 19, the text on the top left 'Noisy neighbors' carries a relatively independent relation with the image. The green sentences describe the events in the apartments. However, the image is much bigger than the text.

Figure 19. Noisy neighbors (English File (3rd ed), elementary: 38)



G present continuous
V verb phrases
P /ŋ/

5B Love your neighbours

1 VOCABULARY & SPEAKING
verb phrases

a 3.2 Read the article about neighbours. Then listen to eight sounds, and write a–h in the boxes.

Noisy neighbours
the top problems!

Sometimes it is difficult to love your neighbours, especially when they make a lot of noise. These are some things people do that cause problems in the UK (not in order).

- ☐ Their babies cry.
- ☐ Their dogs bark.
- ☐ They talk loudly or argue a lot.
- ☐ They have noisy parties.
- ☐ Their children shout all the time.
- ☐ They have the TV on very loud.
- ☐ They play loud music.
- ☐ They play musical instruments.

b Which do you think are the top three in the UK? Which do you think are the top three in *your* country?

c Do the questionnaire with a partner.

Are your neighbours noisy?
Are you a noisy neighbour?

1 Do you live in a house or a flat?
2 Do you have neighbours...?
  a upstairs
  b downstairs
  c next door
3 Are your neighbours...?
  a very noisy
  b quite noisy
  c not very noisy
4 Which of the things in **a** do they do? Do they make any other noises?
5 Are you a noisy neighbour? Which of the things in **a** do you or your family do?

38

What are they doing?
They're having a party.

1 2 3 4 5 6 7 8 9

Moreover, multimodal texts carried different layouts and designs. For instance, in Figure 18, the text is above the image while in Figure 19 the selected text is on the left. Their font size and their background color also differ.

Besides, it was difficult to design questions and answers for some multimodal texts. For instance, Figure 20 includes information about the actress who is depicted in the image. While the text contains ample information about the actress, the accompanying image merely depicts a woman gazing at the viewers, thereby making it challenging for them to discern her profession.

Figure 20. The actress (English File (3rd ed), elementary: 53)

Helena Bonham Carter, the actress, was born in London in 1966. Her mother is half Spanish and her father, who died in 2004, was the grandson of Herbert Asquith, the British Prime Minister from 1908 to 1916. Her first big role was as Lucy Honeychurch in *A Room with a View*, and her other roles include The Red Queen in *Alice in Wonderland*, Bellatrix Lestrange in the Harry Potter films, and Queen Elizabeth in *The King's Speech*. She has two children with her partner Tim Burton, a film director.

To overcome these challenges, a comprehensive explanation of the original multimodal reading test will be provided in the subsequent section. This test is entirely novel, as there are no predecessors or prior versions of it in existence.

3.3.2.2. Multimodal reading test

Building upon the first draft, I decided to design and develop a multimodal reading test including multimodal texts which could carry rather the same form of information. To do so, first, two topics (e.g., everyday routine) along with their scenarios (e.g., A does three activities every day.) were created for each type of image-text relation, proposed by Martinec and Salway (2005) and Engebretsen (2012) (See section 3.3.2.1 for more details on 10 types of image-text relations used in the multimodal reading test). Second, the topics were researched on Google Images, and the pictures that closely aligned with and were relevant to the topics were chosen. In cases where suitable pictures couldn't be found online, the topics and scenarios were modified. Subsequently, texts were added to the selected images. To ensure that the texts remained within the B1 level, the list of grammar structures and vocabulary topics outlined in the table of contents of the intermediate English File textbook were taken into consideration

Twenty multimodal texts finally formed the multimodal reading test (see Appendix C). The initials and topics were used to name the multimodal reading texts. For instance, in Figure 21, the image enhances the text circumstantially by providing information for the place of stay which is a 'hotel', so it is called 'the IEP-Hotel'. The image-text relations along with their names are provided in Appendix C.

Figure 21. IEP-Hotel

The family is in London for a trip. They are going to stay here for a week.

14. Where are they going to stay for a week?

a. They are going to stay in London

b. They are going to stay in a hotel.

Source: retrieved from https://www.visitroanokeva.com/hotels-lodging/ on August 20th, 2021

Following each multimodal text, a question with two possible answers was included. The first and the second answer included verbal (verbal responses) and visual (visual responses) representations, respectively, so there were no wrong or right answer. In all the questions, the verbal response was presented in the first and the visual response in the second sentence. For instance, Figure 22 shows a tension between the image and text in the ITT-Address. While the text asks the reader to turn left at Hillside, the image illustrates a right turn. The question asks the readers to choose a left (first sentence, verbal response) or right (second sentence, visual response) turn at Hillside Road.

Figure 22. ITT-Address

Go straight on North Street. Then turn left at Hillside Road.
The chemist is next to the cinema.

20. What direction should you go after North Street?

a. You should take a left turn.

b. You should take a right turn.

source: https://quizlet.com/393594617/giving-directions-beginner-diagram/

In three multimodal texts, the visual and verbal responses are defined differently. In the ITI-Rain (Figure 23), where text and image present the same information and both responses refer to text and image, the verbal and visual responses indicate the order of the

information presented in the text. The verbal response includes ‘rivers’ which is mentioned first after the water falls from the clouds. The information which is presented later ‘sea’ is inserted in the visual response.

Figure 23. ITI-Rain

Water falls from the clouds as rain.
The rain runs into rivers.
The rain runs from the rivers into the sea.

1. Where does the rain go?

a. It goes to the rivers.

b. It goes to the sea.

Source: retrieved from https://www.facebook.com/photo/?fbid=2989935507699404&set=pcb.2989935711032717 on August 20th, 2021

In the ITC-Jungles and the ITC-Earth, the visual responses stand for the information which is not written but implied in the text. The ITC-Jungles is provided below as an example (Figure 24). While the verbal response includes the information provided in the text (lungs of the earth), the visual response (Jungles make oxygen.) represents the meaning implied by the word and the image of ‘lung’.

Figure 24. ITC-Jungles

Jungles are lungs of the earth.
Jungles are very important.

4. Why should we protect jungles?

a. Because they are lungs of the earth.

b. Because they make oxygen for the earth.

Source: retrieved from https://twitter.com/earthposts/status/582214800937218048 on 20 August, 2021

The multimodal text was presented on the first and the question along with the answers on the second page. The back button was disabled so that participants could not go back to the multimodal text while answering the questions. It was done to investigate what mode participants preferred and recalled first while answering the questions. The participants were required to read the images and texts thoroughly and answer the questions. They were also informed that there were no right or wrong answers to the questions. The participants were not provided with information regarding the disabling of the back button, nor were they instructed on where to begin reading within the multimodal text or which devices to use for the tests. Furthermore, no specific distance between the screen and the participants' eyes was specified. The participants had one hour to complete the test. The font of the text was 15 Times New Roman. The questions and the answers were in bold so it would be big enough for the participants to read. The images varied in size, ranging from 4 to 10.50 cm in length and 8 to 10 cm in width. The sizes of the images in the multimodal texts discussed in chapter 4 are indicated in parentheses next to their respective names. The first and second number show the height and width of the picture, respectively. For each type of the image-text relation, 2 sets were designed so 20 multimodal texts were developed in total. The texts included different word counts and frequencies and character counts.

In this research, 'text' and 'verbal information/discourse/stimulus' are used interchangeably to refer to the written text in the multimodal text. Similarly, 'image' and 'visual information/discourse/stimulus' refer to the picture in the multimodal text.

3.3.3. Follow-up questions

At the end of the online multimodal reading test, the students answered two follow-up questions (see Appendix D):

1) Did the pictures help you read and understand the texts better?
Answers: Yes/No

2) Did you mostly rely on the picture or the text to answer the questions?
Answers: Picture/Text/Both.

The follow-up questions were designed to see what mode (visual, verbal, or both) they preferred to read and respond to the multimodal reading texts. The follow-up questions were developed in Hungarian in order to avoid any linguistic misunderstanding. Nine secondary school students did not answer the follow-up questions because they were added to the online multimodal reading test after the data was collected from them.

3.3.4. Eye-tracking system and OGAMA

Tobii x120 eye tracker manufactured by Tobii Technology Inc was adopted in the research to record the participants' eye movements during a range of activities. OGAMA (OpenGazeAndMouseAnalyzer) software version 5.1 (30.03.2021) was also used to study eye movement registration and export the data. Ogama, developed by Dr. Adrian Voßkühler at Freie Universität Berlin in Germany, is available for free on ogama.net website. OGAMA is an open-source software designed to analyze eye and mouse movements in slideshow study designs.

OGAMA does not provide calibration data. Ten modules for data analysis and recordings are available in the application. The descriptions provided below are extracted from OGAMA website.

1. **Replay Module**: This module plays back the recorded data from the experiment.
2. **Areas of Interest Module**. This module defines and displays areas of interest (AOI) on the given stimulus images. The defined AOI can be later edited and copied.
3. **Fixations Module**. This module calculates and displays the fixations made by the subjects.
4. **Attention Map Module**. This module calculates gaussian distributions of the fixational data and map them onto the original stimulus image, so that you can see fixated and unfixated locations on the stimulus.
5. **Saliency Module**. It calculates the salient locations on the stimulus images.
6. **Database Module**. This module views and edits the raw sampling data and includes the database tables such as subjects, subject parameters, trials, trial events, fixations and areas of interest.
7. **Statistic Module**. This module calculates a lot of parameters useful for further analysis. This data grid can be exported to SPSS.
8. **Scanpath Module**. This module visualizes groups and compares different scanpaths of the subjects.
9. **Slide Design Module**. The design section provides tools for creating different slide shows such as instructional, pictures and graphical elements, which are suitable for OGAMA.

10. **Recording Module**. This module receives and stores the sampling data from eye-tracker interfaces such as Tobii into OGAMA database, which is immediately accessible for replay and analysis.

In this research, the following eye-tracking measures and information are considered for data analysis:

1- **AOI (Area of interest):** Area of interest refers to a spatial region, defined by the researcher, where the eye-movement data will be extracted from the eye-movement recording for further analysis (Godfroid, 2020). The present study aims to investigate how readers move between the visual and verbal stimuli to read the multimodal texts so it moves beyond the word-based interest areas and designs interest areas around text paragraphs. Larger text-based interest areas help the researcher examine readers' more global processing behaviors in relation to a given task and task instructions (Godfroid, 2020). Additionally, pictorial interest areas explore learners' overall processing of language tasks, encompassing language assessment tasks that are not specifically intended to address inquiries related to the acquisition of particular linguistic features. Instead, they focus on the aspects of task design (Godfroid, 2020). An example of text-based and pictorial interest areas is provided in Figure 25. As the purpose of the study is not where exactly on the visual/verbal stimuli readers look at, the pictorial interest area outlines the edge of the image and not around the visual objects and the text-based interest areas are roughly drawn about 1 cm around the paragraph. The white space within the interest area functions as a buffer to account for recording or calibration errors, as well as errors in the human visual system. Fixations that occur outside of this area will be excluded from the data analysis (Godfroid, 2020).

Figure 25. Areas of interest in the ITT-Address

2- **Regression:** Regressions are eye movements that transport the eyes opposite to the reading direction, for instance right-to-left eye movements in English (Godfroid, 2020). Regressions show a participant's need to further process a specific area on the screen after the eyes have left it. As a result, regressions are primarily regarded as an indicator of reanalysis and have been associated with various forms of processing difficulty (lexical, syntactic, and discourse), or as an effort to rectify an oculomotor error after the reader has passed a target (Godfroid, 2020). In the present research only regressions on the verbal stimuli are considered for the data analysis.

3- **Time until first fixation:** It is the time between the onset of a stimulus and the first fixation within an area of interest (AOI).

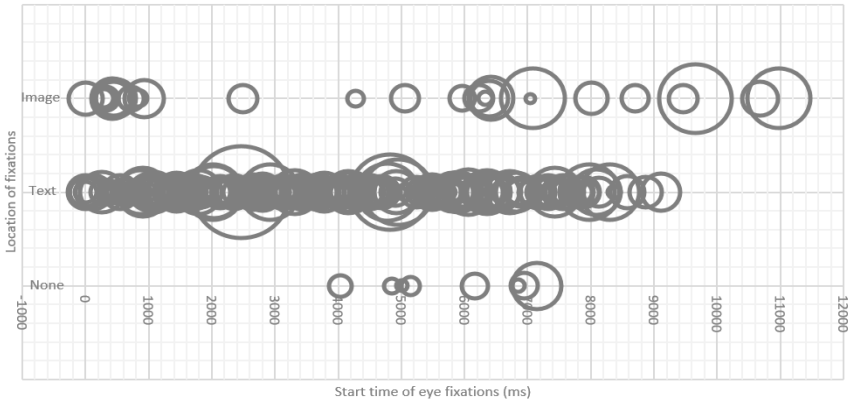
4- **Complete fixation time:** This is the sum of all fixation durations for the selected subjects on the given AOI.

5- **Number of fixations:** This is the count of all fixations for the selected subjects on the given AOI.

6- **Average fixation duration:** This is the sum of all fixation durations divided by the number of all fixations for the selected subjects on the given AOI.

7- **Attentional allocation and scanpath graph:** Attentional allocation and scanpath graph shows the number of fixations and fixation durations on the visual and verbal stimuli from the start to the end of the reading task (see Figure 26). In order to create the graph, the start times of the fixations, the fixation durations, and scanpath strings for each participant were measured and they were later mapped on to each other to find the cognitive processing patterns during the multimodal reading. The center of the circles shows start time of eye fixation and the diameter of the circles represents fixation duration.

Figure 26. Attentional allocation and scanpath graph of the IST-Granny



Forty slides (20 multimodal reading texts slides and 20 question/response slides) were created on OGAMA. The slides designed on OGAMA are exactly the same as the ones used in the online multimodal reading test. However, there are two differences explained in Table 5.

Table 5. Differences in the multimodal reading test between the online and eye-tracking experiments

Text ID	Online experiment	Eye-tracking experiment
The IGT-Lake	George Inn	Craigdarroch Inn
The TSI-Coffee	He is putting it on the desk.	He is putting something on the desk.

Unfortunately, OGAMA did not collect data on the IER-Football so the data from 19 text slides are only discussed. In the subsequent section, a comprehensive explanation of the research samplings and procedures utilized in both the online and eye-tracking experiments is provided.

3.4. Sampling: English language tests and participant selection

I opted for B1 language learners as my participants primarily because according to the Common European Framework of Reference (CEFR), at this intermediate proficiency level, L2 learners demonstrate the ability to understand the primary concepts of straightforward and standard information pertaining to familiar subjects encountered in everyday contexts like work, school, and leisure. They also possess the competence to navigate most situations while traveling in areas where the target language is spoken. Additionally, they are capable of producing coherent texts on topics that are familiar or personally engaging. They are able to speak about experiences, events, and dreams, and provide concise explanations of their opinions and plans. Although at B1 level individuals may encounter challenges when dealing with more complex or abstract subjects, they generally showcase effective communication skills in the language. Consequently, developing a test for B1 L2 learners would be relatively easier. Secondly, upon analyzing the textbooks, their language, and accompanying visuals, I observed that the B1 level offered a broader range of topics and vocabulary that could be effectively represented visually. However, certain topics were too abstract to be depicted through pictures, in the Elementary English File (e.g., dates and time), or in the Upper-Intermediate level (e.g., personality). Thirdly, for some other topics, suitable images were not readily available online, for example, town and cities in the Elementary English File, and illnesses and

treatments in the Upper-Intermediate English File. Finally, given that a majority of my students at the language school were at the B1 level, I presumed that there was a higher likelihood for L2 learners to fall into this category. In fact, the number of B1 students significantly outweighed the number of students at other proficiency levels.

My motivation for studying language learners originates from Anna Pellicer-Sánchez's (2022) article, titled "Multimodal Reading and Second Language Reading," which was published in the International Journal of Applied Linguistics. Pellicer-Sánchez highlights the limited number of studies that delve into the processing of text and images in multimodal L2 reading, despite the prevalent use of visuals in L2 texts. She emphasizes the importance of researchers exploring the connection between cognitive processes, learning outcomes, and multimodal texts that incorporate various intersemiotic relations. To contribute to a more comprehensive understanding of L2 multimodal and multimedia learning, I made the decision to investigate multimodal reading within the context of second language learning. In this section, the selection of the participants in the online and eye-tracking experiments are discussed in details.

3.4.1. Online experiment

The online experiment attempts to investigate the effect of the L2 learners' age on their reading habits and their multimodal reading. The final sample of the participants in our research was established through the application of two types of proficiency tests administered in two distinct groups of language learners. The main aim of the testing procedure was to identify the participants of the research with a proficiency level of B1.

On the one hand, some students from a secondary school in Veszprém, Hungary completed a set of test tasks from B1 Euroexam practice test book, which was selected and presented on the online platform, Flexiquiz. I chose the Euroexam test as it is a well-regarded language assessment that maintains a consistent structure and scoring system across all test administrations. This ensures fair evaluation of test takers and facilitates comparability of scores across different test sessions. Additionally, Euroexam tests are designed to mirror real-life language usage, incorporating tasks that align with everyday situations. This enhances the test's relevance and meaning for the test takers. Furthermore, Euroexam offers tests at various proficiency levels, ranging from A1 to C1 on the Common European Framework of Reference for Languages (CEFR). This allowed me to select a test that matched the participants' current language proficiency level.

The selected test included 3 reading tasks and 20 questions in total. The tasks are explained in details in Table 6. No images were included in the texts. The students had 60 minutes to complete the test.

Table 6. Type of task and number of questions in Euroexam test

Type of task		Number of questions
Task 1	Match the headings with the paragraphs.	8
Task 2	Read the texts and decide if the information is in text A, B, C or D.	7
Task 3	Read the text and answer the multiple-choice questions about it.	5

According to the Euroexam website (euroexam.com) candidates need to score 60% of the total marks with a minimum of 40% in each skill (reading, writing and listening, and speaking skills) to reach B1 level. Because only the reading section of the Euroexam test was used to evaluate the participants' English language proficiency, 60% of the total marks was considered as the pass mark in our research.

Initially, 20 students in a classroom at a secondary school completed the B1 Euroexam test. Based on their results, 9 students who answered 60% (or above) of the questions correctly were selected for the research. Their mean result was 82.11% with a standard deviation of 12.03%, which shows that their exam scores were within the minimum of 71% and maximum of 100% (Table 7).

Table 7. Descriptive statistics of Euroexam test results

	Minimum	Maximum	Mean	Std. Deviation
Euroexam test results	71.00%	100.00%	82.11%	12.03%

As the research was conducted during the COVID-19 lockdown and school closures period, in-person meeting with the students was not possible, so the data collection was achieved completely online. To collect data from the secondary school students, first I acquired the permission from the school principal to run the research. I was not allowed to contact the students so the course teacher was informed about the aim and procedure of the research via email. The teacher was asked to explain the aims and procedures of the research to the students during her online lessons with the students. She also emphasized that participation in the research was completely voluntarily. The link to the Euroexam test was sent to the teacher, which she forwarded to 20 students. After finalizing the Euroexam test results, the teacher was informed about the students' achievements. Only 9 participants received pass marks.

On the other hand, 61 B1 second language learners from a language school in Veszprém, Hungary attended the research. The language school students had been required to complete a placement test run by the language school before enrolling for their courses. The written placement test included 60 grammar, vocabulary and language function questions in total from A1 to C1 level and the students were ranked based on the number of the questions they could answer. The language learners were later interviewed by the head teacher to evaluate their speaking skill.

It is important to note that due to the utilization of different tests for selecting secondary school students and other language learners, there is a possibility that the sample lacks homogeneity in terms of language competence.

3.4.2. Eye-tracking experiment

The eye-tracking experiment adopts an eye tracking technology to investigate how L2 learners read multimodal texts. All the participants in the eye-tracking experiment completed the set of test tasks from B1 Euroexam practice test book and received pass marks (see Table 6 for more information on the Euroexam test). The mean result is 87.77% with the standard deviation of 10.03%, which shows that the exam scores are spread out with the minimum of 75% and maximum of 100% (Table 8).

Table 8. Descriptive Statistics of English language proficiency test

	Minimum	Maximum	Mean	Std. Deviation
Euroexam test results	75.00%	100.00%	87.77%	10.03%

Due to COVID-19 lockdown, when in-person and on-campus activities at university and the access to the laboratory equipment were limited, for the eye-tracking experiment non-probability sampling methods, namely convenience sampling and voluntary response sampling, was used where individuals were selected based on non-random criteria. In order to collect data, several calls for participation in the research were posted on Facebook pages of the university, which invited students to attend the research. The posts included the aim and procedure of the study and the required level of English language proficiency (B1) for the participants. In total, 15 people volunteered to attend the research. I contacted them on Facebook Messenger explaining the research aims and procedures in details. Afterwards, I sent them the link to the Euroexam test. Ten participants received pass marks (60% or above), however, the results of 7 participants are only provided in this research as either the eye-tracking calibrations or recordings failed for 3 participants. Moreover, 2 students who were my students at the language school at the time of the

research voluntarily took part in the eye-tracking experiment. They did the Euroexam test, which was emailed to them, and received pass marks.

3.5. Research procedure

In this section, the steps and methods employed to gather data are outlined.

3.5.1. Online experiment

Nine students who passed the exam at B1 level later received the second link to the online reading habit questionnaire and multimodal reading test sent by the teacher.

Additionally, in order to collect data from the language school students, I discussed the details of my research, its aim and procedures with my B1 level students during the online lessons and asked students to contribute to my research by filling in the online reading habit questionnaire and multimodal reading test. Later, I sent the link of the online tests to 72 students, which were completed by 61 persons.

By clicking on the link, all the participants were directed to Felixquiz website, where they could first read the consent form consisting of a description of the study, the research process and the statement of confidentiality of the data collected during the study. The participants were required to tick the consent box and click the 'submit button'. The consent form was provided in English as the aims and procedures of the research and the confidentially of the data had already been discussed with the students in details and all their possible questions about the research and procedures were answered (see Appendix E). Next, they were asked to fill in the reading habit questionnaire. After filling in the reading habit questionnaire, the participants were automatically directed to the second task, the multimodal reading test. At the end of the multimodal reading test, the participants answered the follow-up questions. Participants were required to register before each test entering their gender and age. After each registration, they were directed to the instruction page. Before completing the reading habit questionnaire, the participants were instructed to read the sentences about their reading habits and choose the option which best defined them (see Appendix A). Before doing the multimodal reading test, the instruction page asked the participants to read the images and texts thoroughly and answer the relevant questions. It was also mentioned that there were no right or wrong answers to the questions (see Appendix C). At the end of the online multimodal reading test, the students were required to fill in the follow-up questions (see Appendix D). The register and the instruction for both tests were in Hungarian in order to avoid any linguistic misunderstanding. The data were later exported to Microsoft Excel

2013 (the Microsoft Inc., Redmond, WA, USA) and IBM SPSS Statistics 25 (the IBM Inc., Armonk, NY, USA) for further analysis.

3.5.2. Eye-tracking experiment

In order to conduct the eye-tracking experiment, I went to the room in the informatics department, where the eye-tracker was available, 30 minutes earlier to set up the eye-tracker and test the camera calibration. On the arrival of the participants, the aim and the procedure of the research were explained to them, all the participants' possible questions were answered and they were asked to read and sign the consent form, which consisted of a description of the study, the research process, and the statement of confidentiality of the data collected during the study (Appendix F). The consent form was developed in English as I had already discussed the research aims and procedures and the confidentiality of the data with the participants in details on Messenger, online lessons and in the in-person meetings before the main trial.

Later, the following instructions were given to the participants:

- Please stay still during the whole experiment.
- Please don't move your head.
- Please have a constant look at the screen and don't look away.
- The text includes 20 multimodal texts. A multimodal text includes an image and a text. Please read them thoroughly and answer the questions. There is no right or wrong answer. Please choose the answer you feel like first.
- Please, press Space to go to the next page.
- Please left click on the box next to your selected answer.

Afterwards, their age and gender were added to OGAMA and the students were asked to sit on the chair in front of the screen as they were comfortable. The Space button and the left click on the mouse were also shown to them. The eye-tracker was calibrated according to each participants' height and chair position. Unfortunately, chin rest, which holds participant's head stable during eye tracking experiments, was not available so it took several trials for the eye-tracker to be calibrated. The recordings were later saved and exported to Microsoft Excel 2013 and SPSS 25.

3.6. Methods of data analysis

Altogether, there are two experiments which shape this dissertation. On the one hand, the online experiment aims to explore L2 learners' reading habits and their modal preferences. With this aim in view, descriptive statistics (e.g., mean, frequency and standard deviation) were used to examine all the L2 learners' responses to the online reading habit questionnaire, multimodal reading test and follow-up questions. The

reading and response durations of the online multimodal reading test were also calculated. Paired samples t-tests were conducted to compare the means of visual and verbal responses and reading and response durations. Besides, a Pearson correlation coefficient was computed to assess the linear relationship between the reading duration and the type of responses (visual or verbal). The first question in the follow-up test (Did the pictures help you read and understand the texts better?) includes two responses (Yes and No), A Chi-square test was thus done to examine the relationship between 'Yes' and 'No' responses. Moreover, an independent samples t-test was run to see if the means of visual/verbal responses and reading/response durations were significantly different in 'Yes' and 'No' responses. The second question (Did you mostly rely on the pictures or the texts to answer the questions?) includes three responses (Picture, Text, Both). In order to compare the distribution of the number of participants who selected 'Pictures', 'Text' or 'Both', a Chi-square goodness-of-fit test was run. Additionally, a one-way ANOVA was performed to compare the effect of 'Picture', 'Text' and 'Both' answers on the verbal and visual responses, and reading and response time. To address the uneven sample sizes, a Gabriel post-hoc test was conducted to identify the specific pairwise comparisons of means that significantly contributed to the overall difference.

The online experiment also attempts to investigate L2 learners' reading habits and multimodal reading in Group A (13-30) and B (31-46). Descriptive statistics were used to examine the L2 learners' responses to the online reading habit questionnaire, multimodal reading test and follow-up questions in Group A and B. Furthermore, an independent-samples Mann-Whitney U test was run to compare the reading habits in Group A and B Furthermore. A Pearson correlation coefficient was run to assess the linear relationship between the age and the reading and response durations in Group A and B. An independent samples t-test was also used to see if the means of visual/verbal responses and reading/response durations are significantly different between Group A and B. Moreover, an independent samples t-test was conducted to compare the means of 'Yes' and 'No' responses to the first follow-up question in Group A and B. A Chi-square test was done to investigate the relationship between 'Pictures', 'Text' and 'Both' responses in the second follow-up question in Group A and B.

In order to locate the specific differences among the age groups, the population was divided into 4 age groups of adolescents (13-17 years old) who were school students, young adults (24-29 years old) who had recently left school and were likely to study at university, adults (31-37 years old) who were likely in the process of completing their

education and establishing a stable lifestyle, and the middle-aged (39-46 years old) who were reaching or were already in their middle age. First, a one-way ANOVA was performed to see if age had a measurable effect on the reading and response durations and visual/verbal responses between at least two age groups. Later, due to the unequal sample sizes, a Gabriel post-hoc test was run to determine which pairwise comparison of means contributed to the overall significant difference.

The eye-tracking experiment aims to examine the L2 learners' cognitive processes while reading multimodal texts. Descriptive statistics and Pearson correlation coefficient were run to report on eye measurements and their possible linear associations. Paired samples t-tests were run to compare the means of eye-measurements in the image and text.

Finally, in order to integrate the results from both experiments, the descriptive statistics and attentional allocations and scanpath graphs of 6 multimodal texts were used to discuss how the intersemiotic interactions encourage visual or verbal responses. Two multimodal texts elicited strong visual responses, while two other multimodal texts evoked substantial verbal responses. The participants rather relied on their background knowledge to provide answers for two other sets of multimodal texts. Paired samples t-tests were run to compare the means of eye-measurements in the image and text of these 3 sets of multimodal texts. The Chi-Square test was also used to examine if the visual and verbal responses to the multimodal texts in the online experiment were related.

In the following chapter, the results of the online and eye-tracking experiments will be presented, providing insights into the collected data and addressing the research questions and objectives.

CHAPTER 4

RESULTS

This chapter presents the results of the analyses as elaborated in the previous chapter. Section

4.1 presents the general results of the online reading habit questionnaire, multimodal reading test and follow-up questions. Section 4.2 discusses the results of the online reading habit questionnaire, multimodal reading test and follow-up questions in different age groups. Section 4.3 elaborates on the results of the eye-tracking experiment. And section 4.4 integrates the results of the online and eye-tracking experiments to investigate how students read and responded to the multimodal texts. Section 5.5 summarizes the main findings of the study.

4.1. Reading habits and multimodal reading

This section addresses the first research question and explores the L2 learners' responses to the online reading habit questionnaire, multimodal reading test and follow-up questions.

4.1.1. The reading habit questionnaire

In this section, the results of the reading habit questionnaire are discussed. Table 9 reveals that in general the participants were mostly involved in reading social media than print and electronic media.

Table 9. Frequency of reading habits (in percentage)

	Never %	Sometimes %	Often %	Regularly %
Print newspaper	35.70	60.00	2.90	1.40
Print magazine	30.00	62.90	5.70	1.40
Print book	4.30	35.70	34.30	25.70
Print comic	78.60	17.10	4.30	0.00
Electronic newspaper	7.10	34.30	32.90	25.70
Electronic magazine	21.40	44.30	21.40	12.90
Electronic book	24.30	52.90	17.10	5.70
Electronic comic	75.70	20.00	2.90	1.40
Social media use	2.90	15.70	15.70	65.70
Keep in touch	4.30	18.60	20.00	57.10
Read text	7.10	20.00	25.70	47.10
Read picture	10.00	20.00	22.90	47.10
Watch video	2.90	21.40	24.30	51.40

Considering print media, the participants were mainly involved in reading books. They often (34.30%) or regularly (25.70%) read books. The participants sometimes spent time reading newspapers (60%) and magazines (62.90%) and seldom read comics (78.60%).

As for electronic media, the participants often read newspapers (34.30%) and magazines (44.30%). They sometimes (52.90%) read books and almost never (75.70%) read comics. The participants regularly used social media (65.70%) to keep in touch with the people (57.10%), watch videos (51.4%) and read texts and images (47.10%).

4.1.2. The online multimodal reading test

This section discusses the results from the online multimodal reading test. In general, the participants preferred the verbal (M=11.10, SD=3.62) over the visual responses (M=8.90, SD=3.62) to answer the questions. The results of the paired samples t-test show that there was a statistically significant difference between the means of visual and verbal responses $t(68)=2.54$, $p= .01$. It also took longer for the participants to read than respond to the multimodal texts. The paired samples t-test results also indicate that the means of reading (M=.13, SD=.07) and response durations (M=.10, SD=.05) were significantly different, $t(68)=4.50$, $p=.00$. Moreover, a Pearson correlation coefficient was computed to assess the linear relationship between the reading duration and the type of responses. The results indicate that the reading duration and the verbal responses were moderately negatively correlated, $r(68) = -.276$, $p = .021$. Accordingly, the reading duration and the visual responses were moderately positively correlated $r(68)=.276$, $p=021$.

In sum, the participants spent longer time reading than responding to the text and they preferred the verbal over the visual responses. Longer reading duration was positively correlated with the number of visual responses.

4.1.3. The follow-up questions

At the end of the online multimodal test, the participants answered two follow-up questions:

3) Did the pictures help you read and understand the texts better? Yes/No.
4) Did you mostly rely on the pictures or the texts to answer the questions? Pictures/Texts/Both.

The results show 65.60% of the participants acknowledged that the pictures helped their reading comprehension (‘Yes’ condition) while 34% of the participants did not find the pictures helpful (‘No’ condition). According to the Chi-square test results, the

relationship between the two variables was also significant (X^2 (1, N = 61) = 61.00, p= .00).

Table 10 also shows that the participants who acknowledged the contribution of the images to their reading comprehension had selected more visual responses and enjoyed longer reading and response durations. However, the independent-samples t-tests reveal that the difference in means was only significant between the visual (t(58)=2.72, p=.00) and verbal (t(58)=-2.72, p=.00) responses in 'Yes' and 'no' conditions. The mean scores of the reading (t(58)=1.24, p=.21) and response duration (t(58)=1.67, p=09) in 'Yes' and 'no' conditions were not significantly different. The Pearson correlation coefficient test results also show a moderately positive relationship between the number of 'Yes' responses and the number of visual responses (r= .33, p=.00).

Table 10. The results of the independent sample t-test for the first follow-up question

	Yes or no	Mean	Std. Deviation	Sig. (2-tailed)
Verbal responses	Yes	9.85	3.47	.00
	No	12.33	3.18	
Visual responses	Yes	10.15	3.47	.00
	No	7.66	3.18	
Reading duration	Yes	.14	.087	.21
	No	.12	.05	
Responding duration	Yes	.11	.05	.09
	No	.09	.03	

As seen in Table 11, the participants believed that the visual-verbal interaction (63.90%) facilitated their response process. They also found the verbal information (29.50%) more helpful than the images (6.60%) to answer the questions. In order to compare the distribution of the number of participants who selected 'Picture', 'Text' or 'Both', a Chi-square goodness-of-fit test was run. A p-value of .00 <.05 indicates statistical significance.

Table 11. The frequency of 'Pictures', 'Texts' and 'Both' answers to the second follow-up question (in percentage)

	Valid Percent
Pictures	6.60
Texts	29.50
Both	63.90
Total	100

Table 12 shows that the participants who selected ‘Text’ enjoyed higher verbal responses (M=13.72, SD=3.10). Similarly, the participants who selected ‘Picture’ (M=13.75, SD=2.21) and ‘Both’ (M=10.23, SD=2.82) enjoyed higher visual responses. The ‘Both’ condition increased the reading duration the most (M=.15, SD=.08), followed by the ‘Text’ condition (M=.11, SD=.04). The ‘Picture’ condition encouraged the shortest reading duration (M=.08, SD=.04). The ‘Both’ (M=.12, SD=.05), ‘Text’ (M=.09, SD=.03) and ‘Picture’ (M=.08, SD=.03) conditions affected the response duration, in descending order.

A one-way ANOVA was performed to compare the effect of ‘Picture’, ‘Text’ and ‘Both’ conditions on the verbal and visual responses, and reading and response time (Table 12). The one-way ANOVA reveals the means of the verbal and visual responses were significantly different in 3 conditions (F(2, 58) = [16.71], *p*=.00). Similarly, the means of the reading duration were significantly different in 3 conditions (F(2, 58) = [3.33], *p*=.04). However, the three conditions did not affect the means of the response duration (F(2, 58) = [2.79], *p*=.06).

Table 12. The results of the one-way Anova for the second follow-up question

	Responses	Mean	Std. Deviation	*Sig.*
Verbal responses	Picture	6.25	2.21	.00
	Text	13.72	3.10	
	Both	9.76	2.82	
	Total	10.70	3.55	
Visual responses	Picture	13.75	2.21	.00
	Text	6.27	3.10	
	Both	10.23	2.82	
	Total	9.29	3.55	
Reading duration	Picture	.08	.04	.04
	Text	.11	.04	
	Both	.15	.08	
	Total	.13	.07	
Responding duration	Picture	.08	.03	.06
	Text	.09	.03	
	Both	.12	.05	
	Total	.11	.05	

The post hoc analysis using Gabriel's method indicated notable variations in verbal and visual responses across conditions (picture vs. text *p*=0.00< 0.05; picture vs. both *p*= 0.03< 0.05; text vs. both *p*=0.00 < 0.05). Conversely, there was no specific pairwise comparisons among the conditions (picture, text, both) with regard to the reading durations.

After obtaining significant differences in the reading duration means among the three conditions through the one-way ANOVA test, the post hoc test did not yield any significant pairwise comparisons. Consequently, the non-parametric Kruskal-Wallis test and pairwise comparisons were also conducted, resulting in consistent outcomes that indicated no significant differences among the means of the reading durations across the three conditions. These findings might suggest a marginally significant result, typically defined as ($0.05 \leq P < 0.10$), which implies the presence of a potential real effect despite the lack of statistical significance.

In Figure 27, it is evident that the participants who perceived 'text' as beneficial for answering the questions predominantly chose the verbal responses. Conversely, participants who believed 'image' was advantageous for answering the questions tended to select more visual responses (see Figure 28). Those who indicated a preference for 'both' options tended to select the visual over the verbal responses.

Figure 27. Means plot for the means of verbal responses

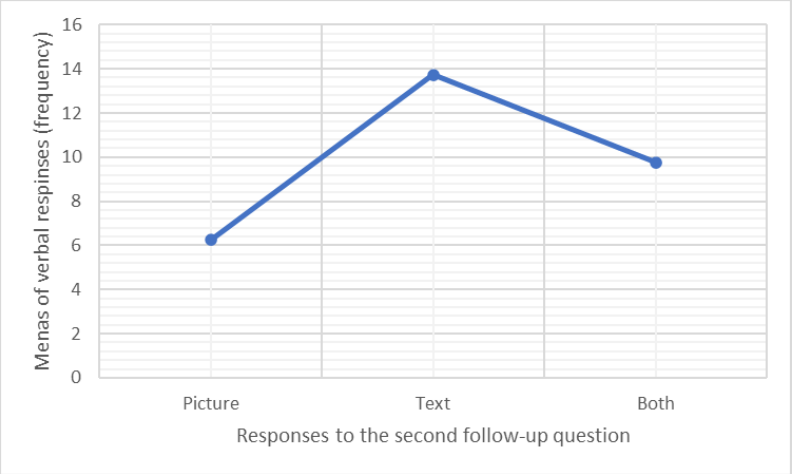


Figure 28. Means plot for the means of visual responses

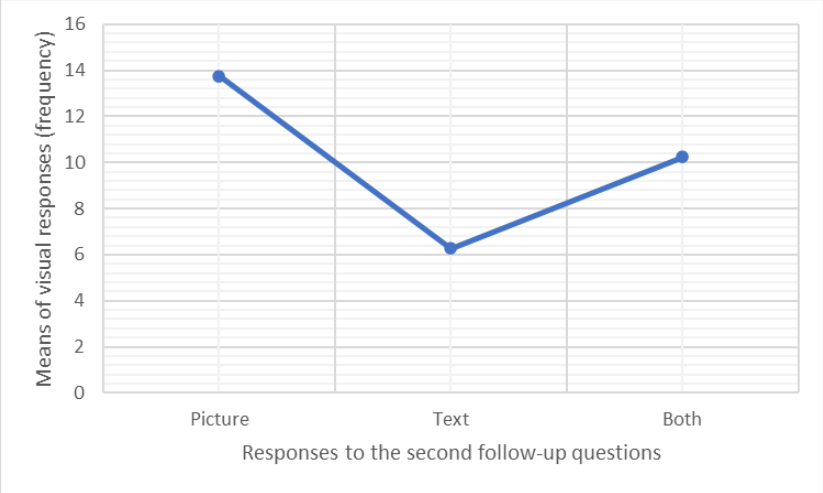


Figure 29 and 30 demonstrate a tendency where the participants who indicated 'picture' and 'both' in the follow-up questions allocated the least and greatest time, respectively, to read and respond to the multimodal texts.

Figure 29. Means plot for the means of reading duration

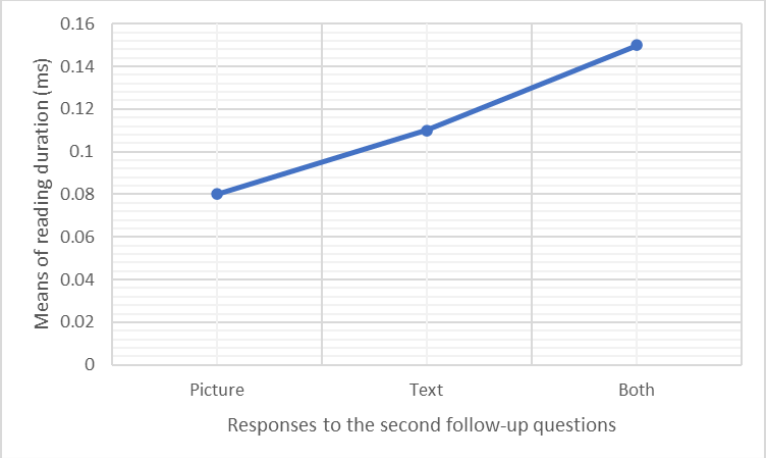
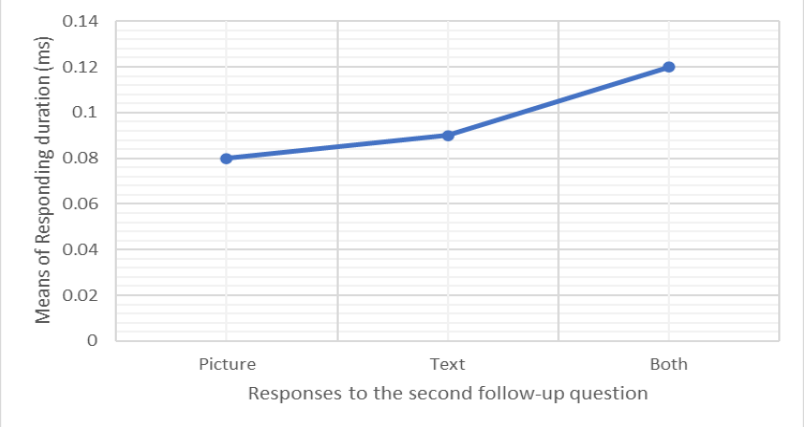


Figure 30. Means plot for the means of response duration



To conclude, most of the participants believed that images helped their reading comprehension, which resulted in increased visual responses. Moreover, the participants found the interaction of visual and verbal discourse more helpful than each mode singly, which increased the reading duration. Moreover, participants who viewed 'Text' as helpful for answering the questions predominantly opted for verbal responses, whereas those who considered 'Picture' as beneficial tended to choose more visual responses.

4.2. Age, reading habits and multimodal reading

In the following section, the effect of age on the reading habits and multimodal reading, the aim of the second research question, is discussed.

4.2.1. Group A (13-30) and B (31-46)

4.2.1.1. Reading habits questionnaire

This section seeks to investigate the reading habits in different age ranges. Table 13 investigates the reading habits of 2 different age groups, A (13-30) and B (31-46). With regard to print media, Group A (33.30%) and B (20%) were regularly involved in reading books, however, the percentage for Group A was much higher. Group A often or sometimes read comics (30%) while Group B often or regularly read print newspapers (7.50%) and magazines (10%).

As to electronic media, the participants in Group B more regularly spent time on electronic newspapers (27.50%) and magazines (15%) and sometimes read electronic comics (27.50%). Group A regularly read electronic books (10%) and often read comics (10%).

Group A (76.70%) went through social media more regularly than Group B (57.50%). While Group A regularly used social media to watch videos and keep in touch (66.70%), read pictures (63.30%) and texts (60%), Group B used social media to keep in touch (50%), watch videos (40%), read texts (37.50%) and pictures (35%), in descending order. An independent-samples Mann-Whitney U test was run to compare the reading habits in Group A and B. The results showed only a significant difference between age groups and reading pictures (p=.00≤.05) and watching videos on social media (*p*=.01≤.05).

Table 13. Frequency of reading habits in Group A and B (in percentage)

	13-30 (Group A)				31-45 (Group B)			
	never %	sometimes %	often %	regularly %	never %	sometimes %	often %	regularly %
Print newspaper	40.00	60.00	0.00	0.00	32.50	60.00	5.00	2.50
Print magazine	36.70	60.00	3.30	0.00	25.00	65.00	7.50	2.50
Print book	0.00	30.00	36.70	33.30	7.50	40.00	32.50	20.00
Print comic	70.00	20.00	10.00	0.00	85.00	15.00	0.00	0.00
Electronic newspaper	10.00	33.30	33.30	23.30	5.00	35.00	32.50	27.50
Electronic magazine	30.00	36.70	23.30	10.00	15.00	50.00	20.00	15.00
Electronic book	16.70	56.70	16.70	10.00	30.00	50.00	17.50	2.50
Electronic comic	80.00	10.00	6.70	3.30	72.50	27.50	0.00	0.00
Social media use	0.00	6.70	16.70	76.70	5.00	22.50	15.00	57.50
Keep in touch	0.00	20.00	13.30	66.70	7.50	17.50	25.00	50.00
Read text	0.00	20.00	20.00	60.00	12.50	20.00	30.00	37.50
Read picture	0.00	13.30	23.30	63.30	17.50	25.00	22.50	35.00
Watch video	0.00	10.00	23.30	66.70	5.00	30.00	25.00	40.00

In sum, Group A read print and electronic books and comics more frequently than Group B. Group B spent time on print and electronic newspapers and magazines more

regularly. While Group A used social media more regularly to watch videos and keep in touch, read pictures and texts, Group B used social media to keep in touch, watch videos, read texts and pictures, in descending order. However, according to the Mann-Whitney U test, there was a noteworthy distinction observed solely in the activities of reading pictures and watching videos on social media when comparing Group A and Group B. In the following section, the results of the online multimodal reading test in group A and B are discussed.

4.2.1.2. Online multimodal reading test

This section discusses how age affects the L2 learner's modal preferences. The Pearson correlation coefficient was run to assess the linear relationship between the age and the reading and response durations in Group A and B. The results indicate that there was a moderately positive correlation between the age of the participants in Group B and reading and response durations, that is the older the participants got, the longer it took for them to read (r(38)= .37, *p*=.01) and respond (r(38)= .35, *p*=.02) to the multimodal texts. There was no significant relationship between the age of participants in group A and their reading (r(28)= -.15, p= .42) and response duration (r(28)= -.06, p= .72).

Table 14 shows that it took longer for Group B (M= .15, SD=.08) to read the multimodal texts than group A (M= .11, SD= .05) and according to the independent sample t-test results, the means were significantly different, *t*(68) = -1.98, *p* = .05. The mean of the response duration for Group B (M= .12, SD= .05) was also higher and significantly different from Group A (M= .08, SD= .03), *t*(68) =-2.98 , *p* =.00. While Group A had a higher number of verbal responses (M=12.53, SD= 3.21), *t*(68) =3.03 , *p* = .00, Group B selected visual responses more frequently (M= 9.97, SD= 3.56), *t*(68) =-3.03, *p* = .00.

Table 14. The descriptive statistics and the results of independent sample t-tests for Group A and B

	Age group	Mean	Std. Deviation	Sig.
Reading duration	Group A	.11	.05	.05
	Group B	.15	.08	
Response duration	Group A	.08	.03	.00
	Group B	.12	.05	
Verbal responses	Group A	12.53	3.21	.00
	Group B	10.02	3.56	
Visual responses	Group A	7.46	3.21	.00
	Group B	9.97	3.56	

In sum, as the participants got older, they spent more time reading and responding to the multimodal texts. It took longer for Group B to read and respond to the multimodal reading test. Moreover, they preferred the visual over verbal responses. In contrast, Group A read and responded to the multimodal texts more quickly. They also selected verbal over visual responses. In the following section, the results of the follow-up questions in Group A and B are discussed.

4.2.1.3. Follow-up questions

In this section, the results of the follow-up questions in Group A and B are discussed. In general, both groups believed that the pictures helped their reading comprehension (Table 15), however, the percentage for Group B (70%) was much higher than Group A (57.10%). In contrast, Group A (42.90%) found the verbal discourse more helpful for their reading comprehension than Group B (30%) did. However, the independent sample t-test results show that the difference between the means of 'Yes' condition in Group A and B was not significant, t(58)=.99, *p*=.32).

Table 15. The percentages of responses to the first follow-up question in Group A and B

Age group	Yes	No
Group A	57.10	42.90
Group B	70.00	30.00

According to Table 16, Group A (57.10%) mostly relied on the text to answer the questions while group B (80%) highly considered the verbal-visual interactions to respond to the multimodal texts.

According to the Chi-Square test of independence, the association between these variables (picture, text and both) in Group A and B was strong X^2 (2, N = 61) = 13.40, p= .00).

Table 16. The percentage of the responses to the second follow-up question in Group A and B

Age group	Picture	Text	Both
Group A	9.50	57.10	33.30
Group B	5.00	15.00	80.00

In sum, Group B believed that the images and their interactions with the verbal discourse improved their reading comprehension while Group A found verbal information more helpful.

4.2.2. Post hoc test

In order to locate the specific differences among the age groups, the population was divided into 4 age groups of adolescents (They are school students.), young adults (They

are likely to have left school recently and might be studying at university.), adults (It is probable that they have started completing their studies and established a stable lifestyle.) and middle-aged (They are reaching or are already in their middle age.) (Table 17).

Table 17. Number of the participants in 4 age groups

Age range	N
Adolescents 13-17	14
Young adults 24-29	16
Adults 31-37	15
Middle-aged 39-46	25

First, a one-way ANOVA was performed to compare the effect of age on the reading and response duration of the participants. The results revealed that there was not a statistically significant difference between the means of at least two groups regarding the reading duration (p=.052>.050). However, the one-way ANOVA showed there was a statistically significant difference between the means of at least two groups regarding the response duration (p=.004<.050). The Gabriel post-hoc test showed significant difference between the adolescent and the middle-aged (p=.03<.05) and the young adults and the middle-aged participants (p=.008<.050) while the rest of groups did not differ significantly. In order to see the tendency for the reading and response durations, their means plots are provided in Figure 31 and 32. The means plots show that the reading and response speed increased from the age of 13 to 29, started to decrease around 30 and reached its lowest speed at the age of 46.

Figure 31. Means plot for the mean of the reading duration

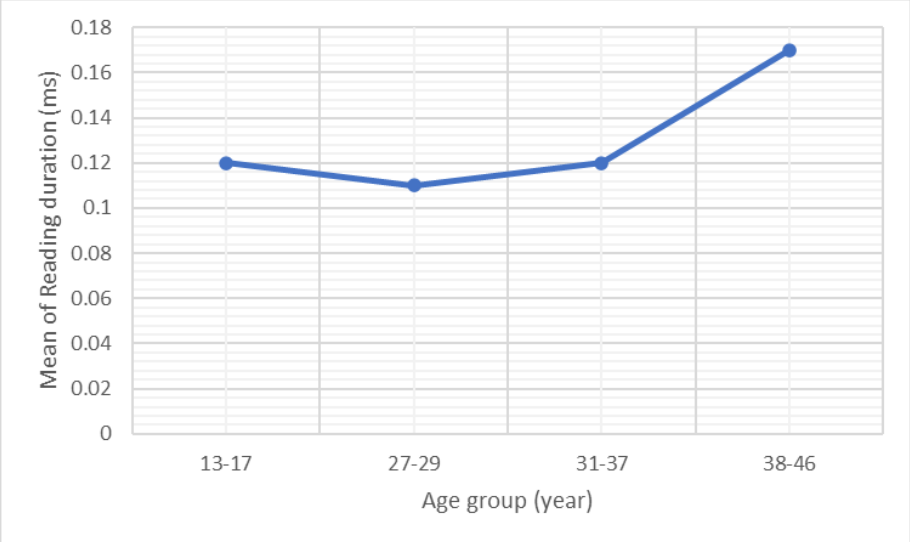
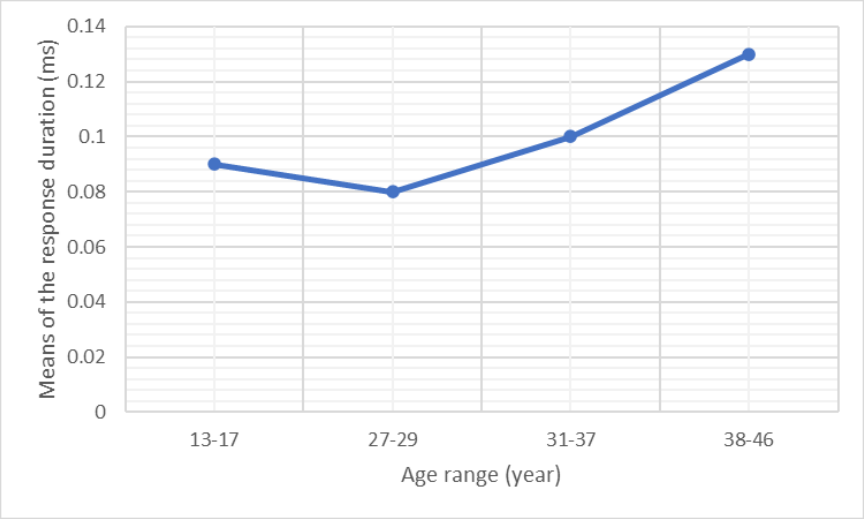
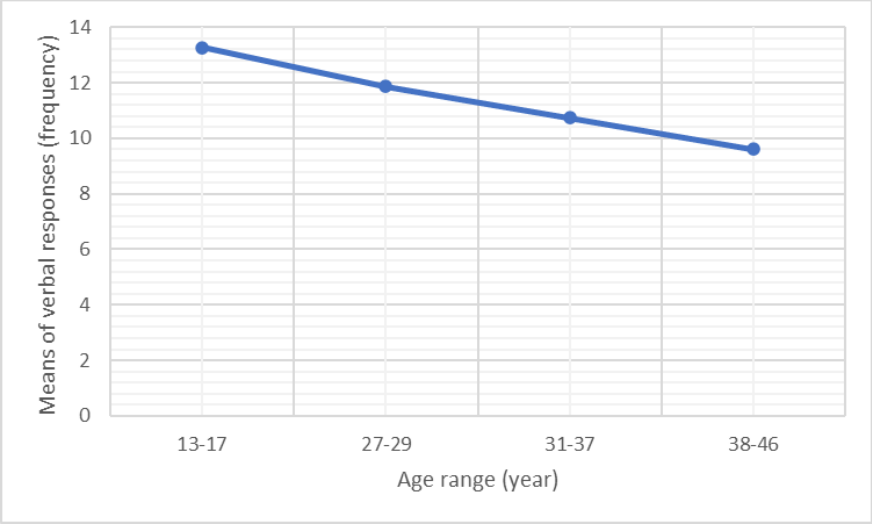


Figure 32. Means plot for the mean of the response duration



A one-way ANOVA was also performed to compare the effect of age on the visual and verbal responses of the L2 learners. The results of the one-way ANOVA revealed that there was a statistically significant difference between at least two groups. (p=.01). The Gabriel post-hoc test showed significant difference between the adolescent and the middle-aged participants (p=.01), while the rest of groups did not differ significantly. Figure 33 shows that the verbal responses experienced a downturn from age 13 to age 46 while there was an increase in the number of visual responses as the participants got older, accordingly.

Figure 33. Means plot for the mean of verbal responses



In sum, the reading and response durations tended to decrease from adolescence to young adulthood, however, it started to increase towards the middle age. The visual responses also increased as the participants aged.

4.3. The eye-tracking experiment

This section provides the descriptive statistics of eye-measurements on text and image to address the third research question and summarizes the main findings of the eye-tracking experiment. The eye-tracking experiment was conducted in order to examine eye-measurements on image and text. Paired samples t-tests were run to see if there was a difference between the means of eye-measurements on text and image (Table 18). The results from the eye tracking experiment reveals that the complete fixation time (M=44203.53, SD=15152.78) on the text was much higher that the complete fixation time (M=10581.37, SD=5201.51) on the image, p=.000<.05. Similarly, the number of fixations (M=185.47, SD=61.09) on the verbal discourse was much higher than the visual discourse (M=43.16, SD=22.64), p=.00<.05. Although the fixation durations on the images (M=250.05, SD= 36.59) were longer than the fixation durations on the text (M=238.00, SD=21.30), the means were not statistically significant, p=.16>.05. Moreover, time until first fixation on the text (M=200.86ms, SD=106.39) was much shorter than the image (M=1918.47, SD=1017.40) (p=.00<.05), which implies that the participants mostly started with a quick skim of the verbal information which was followed by fixations on the image.

Table 18. The descriptive statistics and results of the paired-samples t-test for the eye-measurements on the image and text

	Mean	Std. Deviation	Sig.
Complete fixation time on text (ms)	44203.53	15152.78	.00
Complete fixation time on image (ms)	10581.37	5201.51	
Number of fixations on text	185.47	61.09	.00
Number of fixations on image	43.16	22.64	
Fixation duration on text (ms)	238.00	21.30	.16
Fixation duration on image (ms)	250.05	36.59	
Time until first fixation on text (ms)	200.86	106.39	.00
Time until first fixation on image (ms)	1918.47	1017.40	

Bivariate correlations were also conducted to find out the relationships between word counts and eye-measurements. Table 19 shows that the word counts were strongly positively correlated with the complete fixation time (r=.88, p=00), the number of fixations (r=.89, p=00) and the number of regressions (r= .83, p= .00) on the text and time until first fixation on the image (r= .49, p= .03), that is as the number of words in a text increased, the participants enjoyed longer complete fixation time, more number of

fixations and regressions on the text and a delay in the first fixation on the image. It is likely that long texts were cognitively more demanding and in order to process and keep large information in the short memory, the readers re-read the previously-fixated words to cue their memory for what they previously read. The results did not find a correlation between the word counts and complete fixation time, number of fixations on the image and the average fixation duration on the visual and verbal stimuli.

A Bivariate correlation was also run to see the relationship between the eye-measurements on the image and the number of regressions on the text. The results reveal that as the number of regressions increased on the text, there were relatively a greater number of fixations on the image (r= 472, *p*= .04).

Table 19. The results of the correlations between word count and complete fixation time, number of fixations and fixation duration on visual and verbal discourse

	Pearson correlation	Sig. (2-tailed)
Complete fixation time on text (ms)	.88	.00
Complete fixation time on image (ms)	.34	.14
Number of fixations on text	.89	.00
Number of fixations on image	.40	.08
Average fixation duration on text (ms)	.13	.59
Average fixation duration on image (ms)	-.31	.18
Number of regressions	.83	.00

In sum, while the complete fixation time and the number of fixations were much higher on the text, the findings show that the fixation durations tended to be longer on the images than the text. The results also reveal that the length of the text (in terms of the word counts) affected the complete fixation time and the number of fixations and regressions on the text. To put it differently, an increased word count in a text led to longer complete fixation time, a higher frequency of fixations and regressions, and an extended duration of the first fixation on the image. It might imply that the participants preferred to have a very quick skim of the whole text before moving on to the image. However, the number of the words did not necessarily encourage visual engagements. Regressions along with the fixations on the images were likely to contribute to holding the information in the short-term memory or resolving the difficulty of the text.

4.4. The online multimodal reading test and the eye-tracking experiment

In this section, the results from the online multimodal reading test and the eye-tracking experiment are integrated to address the fourth research question and investigate how students read and responded to the multimodal texts.

4.4.1. Visual responses

According to Martinec and Salway's (2005) framework, when image is subordinate to text, image is related to only a part of text, and when text is subordinate to image, text may well be related to only a part of image. The presence of implicit devices is a reliable indication of text subordination which is decoded by reference to an image. In the TSI-Vegetables (7.3cm×7.73cm), the text is subordinate to the image by including an imperative sentence (Look at Oliver.) and an object pronoun (He doesn't like them), which refer the readers to the image (Figure 34). In the IST-Student (5.5cm×8.20cm), there is more information available in the text than the image. While the text says William is a hardworking and kind student and he helps his friends a lot, the image only reinforces the concept of a hardworking student (Figure 35).

The question in the TSI-Vegetables asks the readers for the reason behind Oliver's sadness. The participants were required to select either the pronoun 'them', which is written in the text (verbal response), or 'vegetables', which is illustrated in the image (visual response). The participants also needed to conclude what makes William a good boy by selecting either his generous help to his friends, which is only written in the text (verbal response), or his hardworking personality, which is the only information illustrated in the picture (visual response) and repeated in the text.

The results of the online experiment revealed that in both texts the participants selected the visual over the verbal responses (70% to 30% in the TSI-Vegetables and 74.3% to 25.70% in the IST-Student). According to the chi-square test there is no significant association between the variables in the TSI-Vegetables and the IST-Student. (p=.40≥.05).

Figure 34. The TSI-Vegetables

Look at Oliver.
He is sad.
He does not like them.

11. Why is Oliver unhappy?

a. He does not want to eat them.

b. He does not want to eat vegetables.

Source: retrieved from https://www.northshore.org/healthy-you/the-tough-stuff-when-eating-and-sleeping-dont-come-easily-for-your-child/ on August 20th, 2021

Figure 35. The IST-Student

William was a hardworking student.
He was also very kind.
He helped friends a lot.

10. What made William a good boy?

a. He helped his friends a lot.

b. He was a hardworking student.

Source: retrieved from https://philnews.ph/2019/09/20/students-powerful-motivational-quotes-study-harder/ on August 20th, 2021

Presenting the eye measurements, Table 20 shows eye-movement measurements in the TSI-Vegetables and the IST-Student. To normalize the measures, certain text-related quantities such as complete fixation time and number of fixations on text were divided by the respective word counts, considering that the texts contain varying numbers of words. Paired-samples t-tests were also conducted to compare the means of eye-movement measurements in the TSI-Vegetables and the IST-Student and p-values below .05 were considered insignificant.

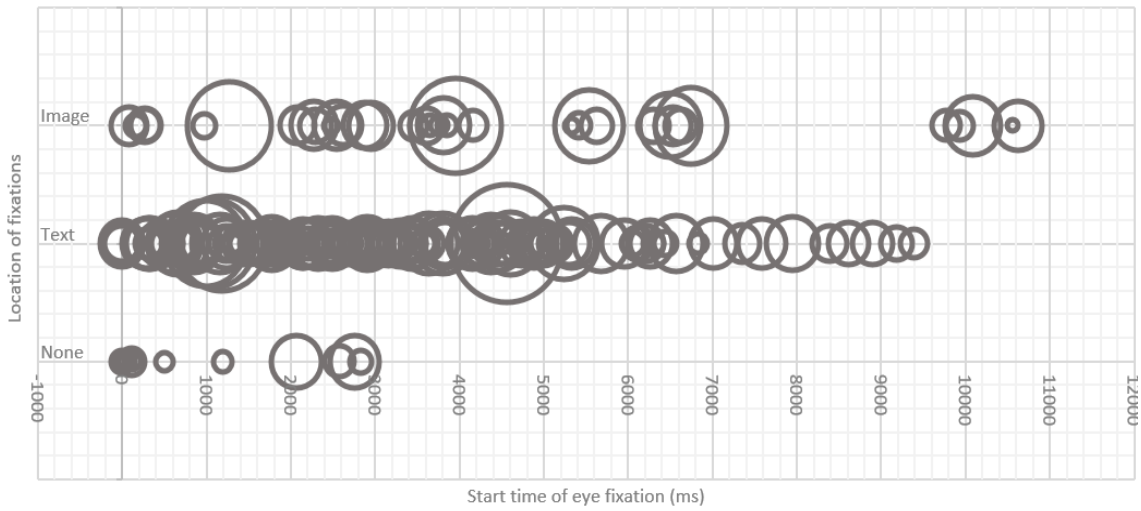
According to Table 20, the average complete fixation time on the image in the TSI-Vegetables surpasses the IST-Student (p=.02≤.05). The image in the TSI-Vegetables enjoys higher number of fixations compared to the IST-Student (p=05≤.05). The average fixation durations on the image (p=.05≤.05) and the text (p=.01≤.05) in the TSI-Vegetables are longer than the IST-Student. Normalized complete fixation time, number of fixations and regressions on the text and time until first fixation on the image do not differ significantly between the TSI-Vegetables and the IST-Student.

Table 20. Means of eye-movement measurements and paired-samples t-test results in the TSI-Vegetables and IST-Student

	TSI-Vegetables		IST-Student		Sig. (2-tailed)	t
	Mean	SD	Mean	SD		
Complete fixation time on image	1112.13	745.26	471.75	755.86	0.02	2.87
Fixation duration on text	285.26	97.11	255.72	106.76	0.01	2.98
Fixation duration on image	250.77	122.82	175.70	90.18	0.05	2.35
Number of fixations on image	4.00	2.00	2.00	1.69	0.05	2.36
Normalized complete fixation time on text	365.90	169.91	396.92	119.38	0.60	-0.54
Normalized number of fixations on text	1.33	0.68	1.65	0.49	0.09	-1.88
Normalized number of regressions on the text	0.28	0.18	0.40	0.14	0.09	-1.87
Normalized time until first fixation on image	165.01	180.03	170.82	162.04	0.94	-0.06

Figure 36 shows that the participants start with skimming the text and the image in the TSI-Vegetable (0ms to 400ms). The number and duration of fixations on the text peak between 400ms and 1600ms. Between 1600ms and 3600ms, intermodal interactions continue when the fixation durations on the text drop. The visual engagement stops shortly at around 4200ms. This period is followed by a slight increase in the number and duration of fixations on the text (between 4200ms and 5000ms). Later on, the fixation counts and durations on the text decrease towards the end of the task during which there is a period of visual engagement (between 5200ms and 7000ms).

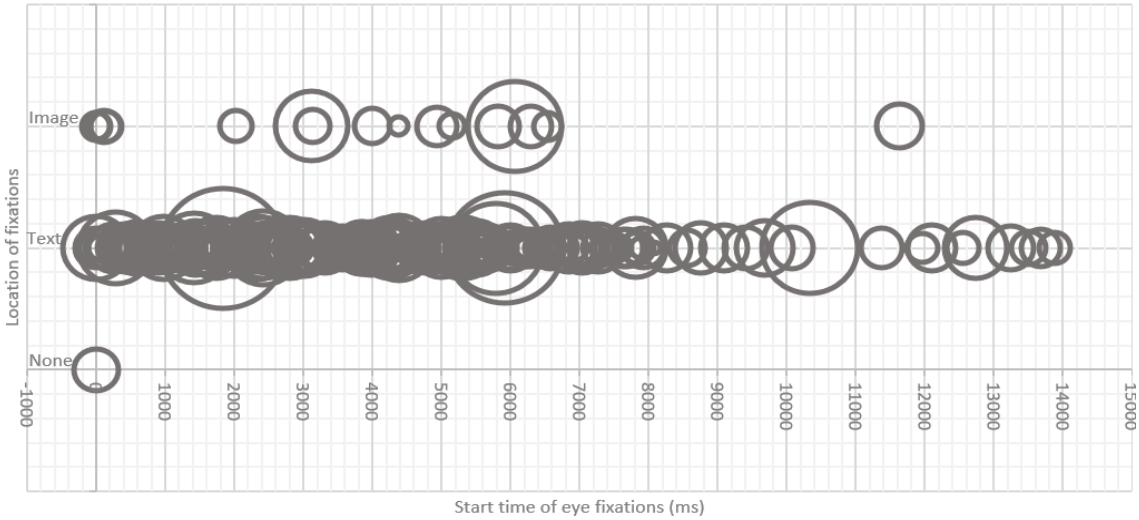
Figure 36. Attentional allocations and scanpath for the TSI-Vegetables (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)



In contrast with Figure 36, Figure 37 does not show a harmonous relation between the image and the text. After a quick skim of the text and the image (0ms to 400ms), there is a long period of processing the text, where the number and duration of fixations increase (400ms to 2800ms). As the visual engagement starts and continues between

2800ms and 6800ms, the numebr and duration of fixations on the text start to drop (2600ms to 3800ms). The number and duration of fixations on the text start to increase again between 3800ms and 5600ms and fall at around 6000ms. The eye fixations and durations on the text gardually start to drop and the reading process stops at 14000ms.

Figure 37. Attentional allocations and scanpath for the IST-Student (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)



In conclusion, the results of the study indicate that the TSI-Vegetables promote increased intermodal interactions, longer fixation durations on both the image and text, as well as a higher number of fixations on the image.

4.4.2. Verbal responses

In this section, the findings of the IEP-Hotel (5.87cm×.10cm) (Figure 38) and the TET-Reading (6.93cm×9.23cm) (Figure 39) are presented. In the IEP-Hotel the image enhances the text by place by illustrating a family who is in a hotel. The word ‘here’ could refer to either ‘London’ or ‘hotel’. In the TET-Reading, the text enhances the image temporally by the word ‘midnight’ while the girl is depicted in bed reading a book before sleeping. She could go to sleep either at midnight or after reading a book. Both texts receive high percentages of verbal response (81.40%). The results of the chi-square test reveals that there is a significant association between the two categorical variables in the IEP-reading and the IEP-Hotel ($p = .04 \leq .05$).

Figure 38. IEP-Hotel

The family is in London for a trip. They are going to stay here for a week.

14. Where are they going to stay for a week?

a. They are going to stay in London

b. They are going to stay in a hotel.

Source: retrieved from https://www.baamboozle.com/game/290508 on August 20th, 2021

Figure 39. TET-Reading

Anne always reads a book in bed at night.
She sleeps around midnight.

17. When does she sleep?

a. She sleeps around midnight.

b. She sleeps after reading a book.

Source: retrieved from https://www.wikihow.com/Have-a-Weekend-Morning-Routine-(Girls) on August 29th, 2021

Presenting the eye measurements, Table 21 shows eye-movement measurements in the IEP-Hotel and the TET-Reading. In order to normalize the eye-measurements, text-related variables such as total fixation time and number of fixations on text were divided by the corresponding word counts, taking into account the variability in the number of words across texts. Based on Table 21, the complete fixation time on the image in the IEP-Hotel is greater than in the TET-Reading; however, this distinction does not yield a statistically significant difference (p=.15≥.05). Conversely, while the fixation duration on the text in the IEP-Hotel is considerably longer than in the TET-Reading (p= .02 ≤ .05), the fixation duration on the image in the TET-Reading surpasses that of the IEP-Hotel. Nevertheless, this contrast lacks statistical significance (p= .17 ≥ .05). Moreover, the number of fixations on the image in the IEP-Hotel exceeds that in the TET-Reading (p = .04 ≤ .05). Regarding normalized measures, the complete fixation time and number of fixations on the text in the TET-Reading are higher than in the IEP-Hotel; however, their means do not demonstrate statistical significance. Additionally, although the number of fixations on the text and the time until the first fixation on the image in the IEP-Hotel surpass those in the TET-Reading, the disparity lacks statistical significance.

Table 21. Means of eye-movement measurements and paired-samples t-test results in the IEP-Hotel and TET-Reading

	IEP-Hotel		TET-Reading		Sig. (2-tailed)	t
	Mean	SD	Mean	SD		
Complete fixation time on image	1414.33	1484.83	858.11	637.91	0.15	1.55
Fixation duration on text	232.41	62.65	209.68	55.43	0.03	2.54
Fixation duration on image	212.69	113.73	261.43	128.77	0.17	-1.47
Number of fixations on image	5.55	3.87	3.33	1.87	0.04	2.40
Normalized complete fixation time on text	288.50	108.35	314.95	128.28	0.47	-.74
Normalized number of fixations on text	1.22	0.30	1.49	0.51	0.10	-1.80
Normalized number of regressions on the text	0.22	0.11	0.27	0.13	0.35	-0.98
Normalized time until first fixation on image	149.82	147.99	95.42	136.10	0.43	0.82

The scan path of the IEP-Hotel in Figure 40 shows that there is a brief skim of the text (0ms to 400ms) which is followed by a period of few and relatively long fixations on the image (0ms to 1000ms). From 400ms to 2600ms, the number and duration of fixations on the text reach a peak, when there are no fixations on the image, and gradually drop from 2600ms to 8000ms during which visual engagement happens. From 3000ms to 11000ms, there is a constant interaction between the image and text while the durations of fixations on the image increases between 8000ms and 11000ms.

Figure 40. Attentional allocations and AOI sequence chart for the IEP-Hotel (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)

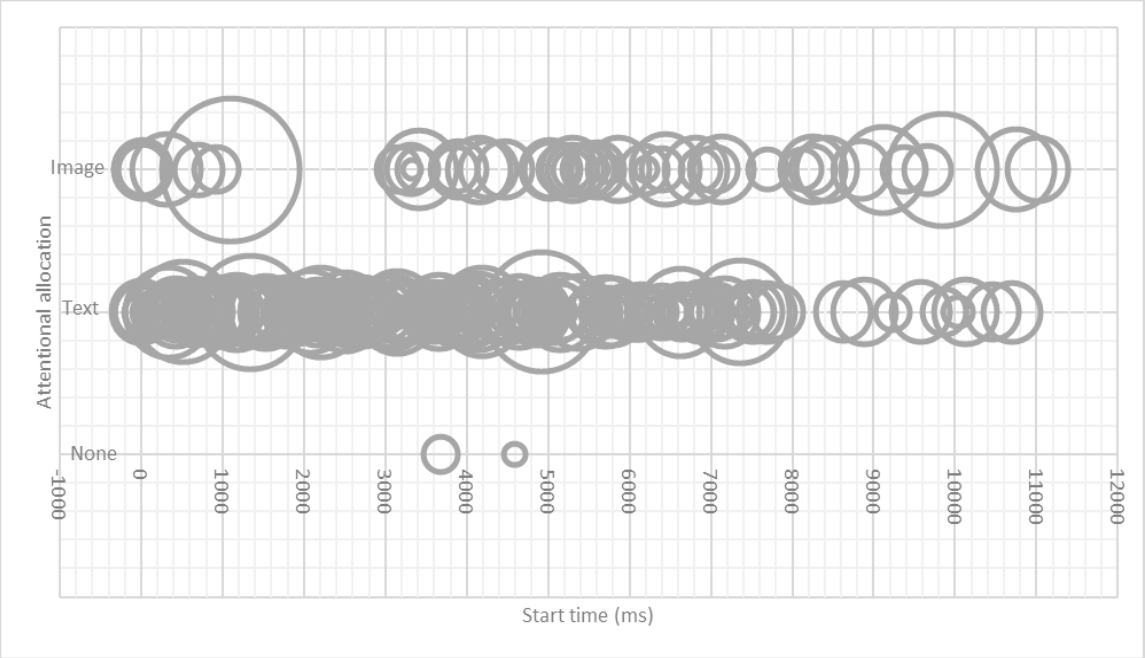
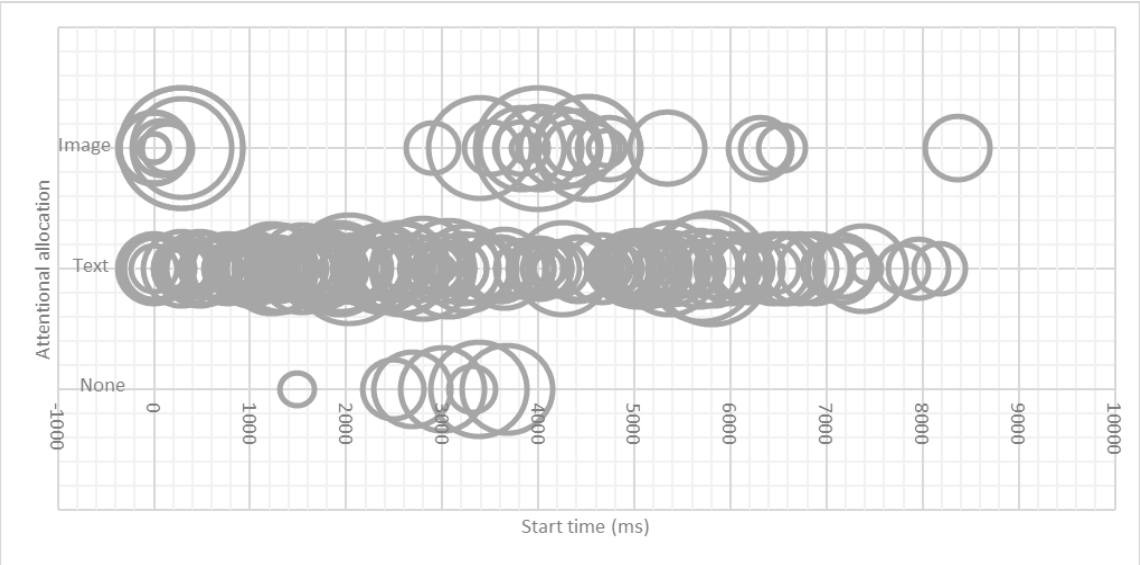


Figure 41 shows that there is a brief skim of the text and the image (from 0ms to 600ms) in the TET-Reading. From 600ms the number and duration of fixations on the

text start to increase, reach a peak around 2000ms and drop around 3600ms when the long fixations on the image start and continue till 5200ms. As the fixation durations on the image drop, the fixations and durations on the text increase again (between 4600ms and 6000ms). There are few fixations on the image at around 6000ms and later the number and duration of fixations start to drop till the end of the reading process (8000ms).

Figure 41. Attentional allocations and AOI sequence chart for the TET-Reading (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)



To summarize, the results indicate that the IEP-Hotel exhibits more significant intermodal interactions compared to the TET-Reading. Additionally, the participants displayed longer fixations on the text and a higher number of fixations on the image in the IEP-Hotel. However, both texts received a high percentage of verbal responses in the online experiment.

4.4.3. Readers' experience and knowledge of the world

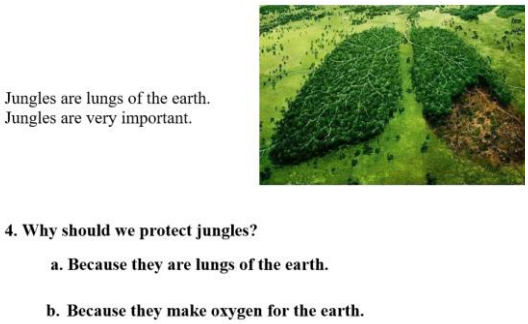
In the ITC-Earth (7.27cm×7cm) and the ITC-Jungle (5.51cm×8.52), the image attempts to complete the words and provide a schema into which the text is integrated (See Figure 42 and 43). This image-text complementary relation creates a third semantic notion which is larger than some of its part. For instance, while the text in the ITC-Earth (Figure 42) says that the earth is melting, the melting ice-cream in the picture reminds the readers of the heat. The text in the ITC-Jungle (Figure 43) also talks about jungles as the lungs of the earth. The image shows an effected part of the lungs (jungles) which implies global warming and environmental threats. The results of the online multimodal reading test reveal that the number of visual exceeded the verbal responses (62.90% for the ITC-Jungle and 57% for the ITC-Earth). Visual responses in these two texts stand for the

information which is not written in the text. The chi-square test results indicate that there is a significant association between the two categorical variables (visual and verbal responses) in the ITC-Earth and the ITC-Jungles (p =.01 ≤ .05)

Figure 42. ITC-Earth

Source: retrieved from https://www.foodnavigator.com/Article/2016/12/01/Carbon-price-would-result-in-food-costs-rising-3 on August 20th, 2021

Figure 43. ITC-Jungle



Source: retrieved from https://www.change.org/t/deforestation on August 20th, 2021

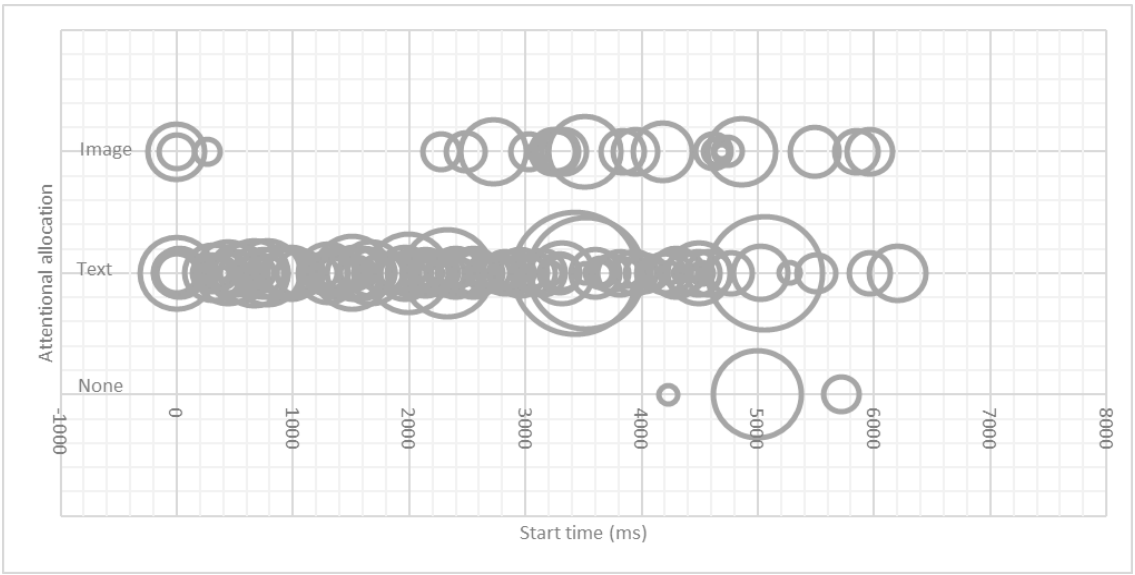
Table 22 displays eye movement data from two multimodal texts, namely the ITC-Jungles and the ITC-Earth. To ensure consistency in the eye measurements, certain variables related to the text were adjusted by dividing them with the corresponding word counts, accounting for variations in the number of words across the texts. According to Table 22, the complete fixation time on the image in the ITC-Earth is longer compared to the ITC-Jungle. However, there is no statistically significant difference in their means (p = .71 ≥ .05). Conversely, the durations of fixations on both the image and text in the ITC-Jungle are longer than those in the ITC-Earth. However, only the mean fixation duration on the text differs significantly between the ITC-Jungle and the ITC-Earth (p= .05 ≤ .05). The number of fixations on the image in the ITC-Jungle and the ITC-Earth are the same. Regarding the normalized measurements, the complete fixation time, number of fixations and regressions on the text, and time until the first fixation on the image in the ITC-Earth are higher than those in the ITC-Jungles. However, their means are not statistically different.

Table 22. Means of eye-movement measurements and paired-samples t-test results in the ITC-Jungle and the ITC-Earth

	ITC-Jungle		ITC-Earth		Sig. (2-tailed)	t
	Mean	SD	Mean	SD		
Complete fixation time on image	929.16	788.21	1007.97	979.90	0.71	0.37
Fixation duration on text	291.62	124.85	225.42	70.21	0.05	2.35
Fixation duration on image	257.17	117.09	247.65	83.38	0.85	0.19
Number of fixations on image	3.66	2.16	3.66	2.65	1.00	0.00
Normalized complete fixation time on text	399.33	116.87	412.04	230.73	0.88	-0.15
Normalized number of fixations on text	1.45	0.35	1.81	0.90	0.32	-1.05
Normalized number of regressions on the text	0.27	0.14	0.40	0.26	0.34	-1.00
Normalized time until first fixation on image	174.41	183.42	306.35	221.09	0.26	-1.26

Figure 44 depicts that the readers start the ITC-Jungles with a few brief fixations on the text and the image in order to get the gist of the message (0ms to 400ms). From 400ms to 2400ms the number of fixations on the text remains at peak. At 2400ms intermodal interactions start with a couple of intermittent fixations on the image and continue till 6000ms during which the number of fixations on the text and the image start to drop.

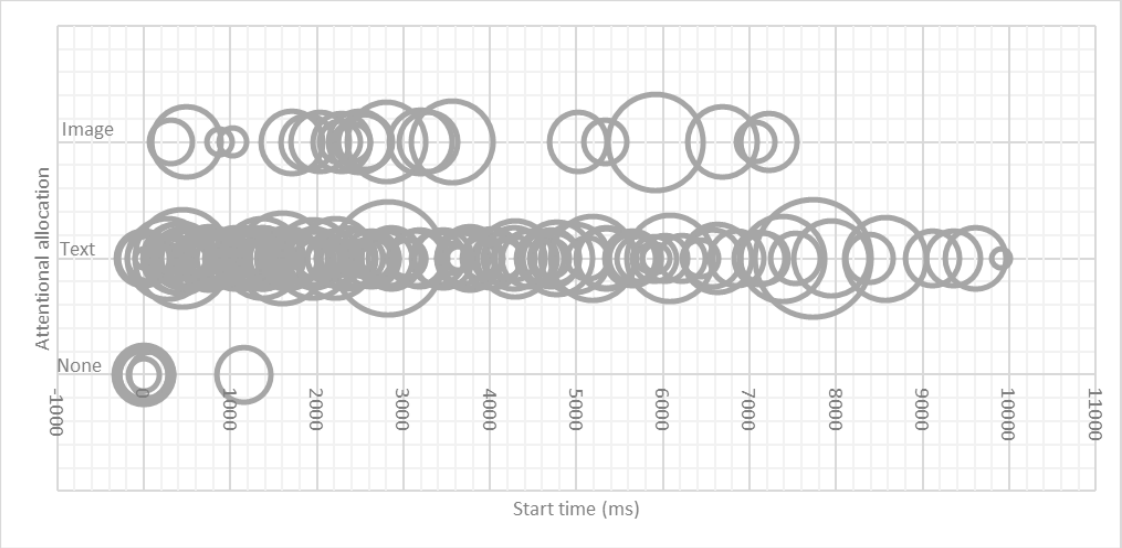
Figure 44. Attentional allocations and AOI sequence chart for the ITC-Jungles (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)



Regarding the ITC-Earth (Figure 45), the participants initially have brief fixations on the text lasting from 0ms to 200ms. The first fixations on the image occur relatively late, between 200ms and 1000ms. Subsequently, the number of fixations on the text reaches its peak within the timeframe of 200ms to 1600ms. Two periods of casual visual engagement can be observed. In the first engagement period (1400ms to 4000ms), the number and duration of fixations on the text slightly decrease. During the subsequent

pause in visual processing, there is an increase in the number and duration of fixations on the text (4000ms to 5000ms). In the second period of visual engagement (4800ms to 7400ms), there is a slight decrease in the number and duration of fixations. When fixations on the image cease, there is a noticeable increase in the duration of fixations on the text (7000ms to 8000ms), and the reading task concludes with a few fixations on the text (8000ms and 1000ms).

Figure 45. Attentional allocations and AOI sequence chart for the ITC-Earth (center of the circles: start time of eye fixation, diameter of the circles: fixation duration)



Overall, despite the fact that the duration of fixations on the text was longer in the ITC-Jungles compared to the ITC-Earth, both multimodal texts elicited fairly similar eye-movement measurements. The findings from the online experiment might further demonstrate that the participants incorporated their own experiences and worldly knowledge into the reading process.

4.5. Summary

The results of the online reading habit questionnaire show that the participants were mostly involved in reading social media than print and electronic media. The findings from the online multimodal reading test also reveal that the readers preferred verbal over visual responses. And the longer reading duration caused higher visual responses. With regard to the follow-up questions, the majority of the participants perceived images as beneficial for their reading comprehension, leading to an increase in visual responses. Additionally, the participants found the combination of visual and verbal discourse more advantageous than each mode individually, resulting in longer reading durations. Furthermore, the participants who regarded 'text' as helpful predominantly selected verbal

responses, while those who considered 'image' as beneficial tended to choose more visual responses.

Considering the age and the reading habits, Group A were more involved in print and electronic books and comics while Group B read print and electronic newspapers and magazines more frequently. Group A used social media more regularly to watch videos and keep in touch, read pictures, and texts, however, Group B used the social media to keep in touch, watch the video, read text and pictures, in descending order. However, the Mann-Whitney U test findings indicated a statistically significant distinction solely in reading pictures and watching videos on social media when comparing Group A and Group B. With regard to the age and multimodal reading, Group B spent longer time reading and responding to the multimodal texts and preferred visual over verbal responses. Conversely, Group A selected verbal responses more frequently. Moreover, there is a tendency that the reading and response speed increased from adolescence to young adulthood, however, it started to decrease towards the middle age. The visual responses also increased as the participants aged. As for the follow-up questions, Group B believed that images and their interactions with the verbal discourse improved the reading comprehension while Group A found verbal information more helpful.

The findings from the eye-tracking experiment show that the participants spent most of their time reading the verbal stimulus. The texts with higher word counts enjoyed longer complete fixation time, a greater number of fixations and regressions on the text, and a delay in the first fixation on the image, which might indicate that the participants preferred to have a holistic overview of the text before fixating on the image. Visual stimuli along with regressions on the text might help the readers to hold and store the information in their short-term memory or overcome the difficulty of the text. Moreover, visual-verbal relations (e.g., text-subordinate-to-image and image-subordinate-to-text, image-text-complementary) might determine the degree of the participants' involvement with the image.

In the upcoming section, the research findings will be connected back to the literature review.

CHAPTER 5

DISCUSSION

The aim of this dissertation was to find out the reading habits and modality preferences of Hungarian L2 learners in two different age groups (13-30 and 31-46). Moreover, this research attempted to investigate how Hungarian L2 learners read multimodal texts. The results indicate that two groups expressed different reading and modality preferences and multimodal reading was mainly text directed. This chapter elaborates on the results of the study and relates them back to the literature. Section 5.1 discusses the results of the online reading habits questionnaire, multimodal reading test and follow-up questions. The L2 learners' reading and modality preferences in Group A and B are reviewed in section 5.2. In section 5.3, the results of the quantitative data analysis from the eye-tracking experiment are discussed. Section 5. 4 integrates the results of the online and eye-tracking experiments to discuss the findings of six multimodal texts and look into different phases of multimodal reading.

5.1. L2 learners' reading habits and modal preferences

The present research attempts to investigate the reading habits and interests of 70 L2 learners with B1 English language proficiency in Veszprém, Hungary. The results from the reading habit questionnaire indicate that the participants regularly read print books and electronic newspapers. They also used social media to keep in touch, watch videos, read texts and look at pictures, in descending order. Comparing print, electronic and social media, the participants were greatly involved in social media. The findings conform to the previous studies (Noor, 2011; Rajab & Ali-Sadi, 2015; Qenaway, 2019), which believe second language learners maintain a constant connection on social media and its domination has significant impact on their reading culture and interests. The popularity of social media among the participants could result from the fact that they are constantly available on the phone, user-friendly, require very less knowledge of the internet, ease sharing information and interactions and give people an opportunity to meet new people. Social media also offer a wide range of modes (e.g., writing, images, speech, and moving images) (Svärdemo Åberg & Åkerfeldt, 2017), which provide enormous potentials of engagement and meaning making in the world (Kress & van Leeuwen, 2001), and allow for more creative and expressive forms of communication through interactions of different modes (Svärdemo Åberg & Åkerfeldt, 2017).

Although the reading habit questionnaire reveals a significant level of engagement with social media among the participants, where they utilize multimodality in diverse and captivating contexts, the results from the online multimodal reading test suggest a preference for verbal over visual responses (e.g., Muñoz, 2017; Serrano & Pellicer-Sánchez, 2022; Pellicer-Sánchez, et al., 2020; Pellicer-Sánchez, et al., 2021). This could be due to the fact that what schools offer to students and prepare them for is quite different from the visual and pictorial world outside school (Kress, 2003). Moreover, English language proficiency for L2 learners has traditionally meant learning of linguistic knowledge and skills (e.g., vocabulary and grammar) for comprehension and communication (Hung et al, 2013). Thus, literacy has been conventionally considered only as a set of decoding and encoding skills, which L2 learners need to master. According to Jakobsen (2019), the process of acquiring language is distinct from learning other subjects or domains, as language continues to hold its position as a primary means of communication. Although all the modes on social media are considered as equal communicational forms (Svärdemo Åberg & Åkerfeldt, 2017), not all modes are appreciated and assessed in educational settings (Kress & Selander, 2012) and written texts have a more important status in teaching and assessment, compared to an image which is often recognized as a decoration to the written text (Jewitt, 2009; Oldakowski, 2014). Moreover, students are usually assessed based on the end product rather than the process of learning in educational settings (Burgess, 2016). The widespread monomodal methods of assessment and the lack of multimodal assessment tools (Unsworth, 2008) stop students from forming multimodal reading skills and perceptions. The presence of questions in the multimodal reading test may have inadvertently influenced participants' responses, possibly creating a washback effect akin to that observed in academic assessments conducted in schools and universities (Alderson & Wall, 1993). As a consequence, it may have hindered the participants from adopting a multimodal approach when reading multimodal texts. This assertion is supported by the fact that certain participants expressed their apprehensions regarding their potential grades to the researcher prior to the test. Besides, considering the respondents' extensive experience with social media, it is plausible that they are well aware of the potential for visual information to be misleading or deliberately manipulated. Consequently, disregarding such information could even serve as a crucial "multimodal reading strategy" in certain circumstances.

With regard to the follow-up questions, the majority of the participants held the belief that images enhanced their reading comprehension, leading to increased number of visual responses. Furthermore, the participants found the combination of visual and verbal elements more advantageous than each mode individually, which resulted in longer reading durations. Additionally, the participants who perceived "text" as valuable for answering questions primarily favored verbal responses, while those who regarded "image" as beneficial tended to opt for more visual responses.

Longer reading duration in the ‘both’ condition could result from the fact that different cognitive subsystems are used to process and encode visual and verbal information, which involves a higher cognitive demand (Florax & Ploetzner, 2010; Mayer, 1997). According to the research conducted by Baddeley (1992), individuals who engage with multimodal texts employ a selective approach by choosing words that hold relevance, constructing a prepositional representation, and subsequently transforming the verbal information into a mental model conveyed by the text. When it comes to the visual information, they make deliberate choices regarding pertinent images, construct a visual representation, and then transform it into a mental model represented by the pictures. Ultimately, a connection is established between the visual and verbal mental models, both of which coexist within the working memory concurrently. Therefore, the participants who found the interaction of visual and verbal discourse more helpful spent longer time mapping the visual and verbal mental models onto each other to create their final understanding, compared to the participants who selected only ‘image’ or ‘text’. The participant’s approval for intermodal benefits may also confirm Paivio’s dual coding theory suggesting that when two modes are processed through two different channels, the simultaneous activation of the verbal and non-verbal systems fosters learning (Clark & Paivio, 1991; Paivio, 1986). There are several studies that claim pictures help L2 reading comprehension (e.g. Elley & Mangubhai, 1983; Tang, 1992; Wright, 2010). Visuals strengthen the links between the verbal and visual systems, facilitate word decoding and reading comprehension by providing additional contextual information (Tang, 1992), activating the schema relevant to comprehension and influencing the type and depth of semantic processing (Seel, 1986; Brody & Legenza, 1981).

While the use of static images in the reading process has generally been viewed positively by learners (Tang, 1992), the extended reading duration observed in the "both" condition may also stem from confusion or uncertainty arising from the interaction between image and text. This aspect will be briefly explored here. In a multimodal text

where there is a relationship of tension between image and text, a reader who is equally engaged with both modes may require more time to align their visual and verbal mental models in order to reach a final conclusion. The discrepancy presents in the picture (e.g., ITT-Address, IEP-Hotel) can activate an inappropriate schema (Lehman et al., 2007) for readers to base their verbal mental model on, leading to a period of confusion and uncertainty. This, in turn, increases the interactions between the visual and verbal elements, resulting in a longer reading time.

Furthermore, the results indicate a preference for visual cues as the primary response when both modes are engaged in multimodal reading. According to Paivio (1986), pictures are more likely to be coded dually, processed and stored in the imaginal and verbal system as a 'copy', and have high imageability, therefore, they are more likely to be retrieved more successfully. Wolfe (2018) mentions that more visual engagement during visual-verbal interactions in reading helps the readers understand the multimodal text and improve their memory for the difficult text. Mayer et al. (2008) also state that interesting information uses up more of the reader's processing resources than low-interest information, which might explain why visual responses are preferable when readers are equally involved in both modes. Nyström and Ögren (2012) also acknowledge the persuasive effect of pictures and suggest that illustrations persuade students that an answer is true, irrespective of whether or not it actually is. The higher visual responses and shorter reading duration for the participants who acknowledged the contribution of visual information ('image' condition) to their comprehension might also emphasize that participants did not need much time to extract a lot of information from pictures. There could be alternative explanations for this phenomenon, such as the possibility that these students solely relied on the pictures and hastily responded without giving much attention to the accompanying text. This behavior could stem from personal preference or simply a motivation to finish the test quickly. Furthermore, the findings from the follow-up questions might indicate that the participants approached the task differently - while some participants found visual-verbal interactions helpful to their reading comprehension, others relied only on verbal or visual modes. These idiosyncrasies may explain that besides the bottom-up controls (e.g., image-text relations) (Bruce & Tsotsos, 2006; Kress, 2010), reading texts are also driven by the readers' top-down mechanisms (e.g., modal preferences) (Tinio, 2013; Wang & Pomplun, 2012).

5.2. Age and L2 learners' reading habits and modal preferences

With regard to the age of the participants, the results reveal that, in general, Group A read print and electronic books and comics, and social media more frequently than Group B. Group B spent time on print and electronic newspapers and magazines more regularly. While Group B used social media to keep in touch, watch videos, read texts and pictures, in descending order, Group A used social media more regularly to watch videos and keep in touch, read pictures and texts. The results demonstrated that different age groups showed rather different reading preferences (Chauhan & Lal., 2012; Fusco, 1986; Platt, 1986). Nevertheless, when Group A and Group B were compared, a significant difference was observed specifically in the activities of reading pictures and watching videos on social media, as indicated by the Mann-Whitney U test. The findings from the online multimodal reading test also show that it took longer for Group B (31-46) to read and respond to the multimodal texts (İyigün et al., 2018) and they relied on images more frequently to respond to the multimodal texts. Moreover, according to the results of the follow-up questions, Group B believed that images and their interactions with the verbal discourse improved their reading comprehension while Group A found the verbal information more helpful. The means plots show that there was a tendency for the reading and response speed to increase from the age of 13 to 29, start to decrease around 30 and reach its lowest speed at the age of 46. Moreover, as the participants got older, the number of verbal responses decreased while there was an increase in the number of visual responses, accordingly. The Gabriel post-hoc test showed significant difference in the means of response duration between the adolescent (13-17) and the middle-aged (39-46) and the young adult (24-29) and the middle-aged participants. There was also a significant difference in the means of visual and verbal responses between the adolescent and middle-aged participants.

Firstly, reading speed increases with age but deteriorates when adults reach 40 and above (Calabrèse et al, 2016; Chen et al., 2019). The results are in line with Solan et al's (1995) findings, who believe that elderly adults read more slowly, which could be due to the decline across different domains of cognitive function (Cullum et al. 2000), memory decline (Grady &Craik, 2000), poorer working memory (Park et al. 2002) and reduced information processing speed (Eckert et al. 2010). Thus, they may look for visual cues, which reinforce the information provided in the text and help improve their comprehension. Secondly, the online reading habit questionnaire showed Group B read

newspapers and magazines more frequently. This information might indicate that they read mostly for pleasure or as their hobby and not for academic purposes, thus they were less stressed-out and took their time to enjoy reading the texts.

Thirdly, the reading habit questionnaire results led to the conclusion that Group A tended to exhibit a higher level of engagement in reading books compared to Group B. As the choice of book was not specified in the questionnaire, one explanation could be that young readers are considered to have particular emotional and academic needs (Hopper, 2005). In Group A, the participants might be mostly school or university students which obliged them to have extensive readings on their subjects. Besides, the psychological constructs of adolescence require them to search for their identity through living the roles in the fictions (Erikson, 1974). They also read fictions to forget the difficulty growing up and experience their "virtual futures" (Sarland, 1991: 110). This greater exposure to written texts might have developed their reading skills in L1 and subsequently in L2. According to the European Commission (2015) developed L1 is a pre-requisite to all kinds of learning, including foreign language learning, and has positive correlations with L2 achievements (Ardasheva, 2016). Bernhardt (2011) also believes in the contribution of reading proficiency in one's L1 to their L2 reading because a good reader in L2 deploys similar cognitive and physiological resources and the same types of skills necessary for fluent reading in L1. And attention to print increases with reading proficiency (e.g., Roy-Charland et al, 2007). Due to Group B's potential distance from educational settings and lack of regular obligation to read books, it is plausible that their reading skills might have deteriorated, resulting in a decrease in their reading speed. Fourthly, in addition to language schools, what other educational settings (e.g. schools, colleges and universities) offer students and prepare them for in domain learning differs from the visual and pictorial world outside educational environments (Kress, 2003), so students are mostly habituated into learning the language content. Hardiess and Weissert (2021: 2) mention that "people have a habit of reading text because our prior experiences have taught us that text often conveys important information and incorporate text as high-level (feature) object". This habit formation during education – starting from early childhood and continuing in adulthood at universities – might gradually fade away as students graduate. This might explain why Group B selected visual information more frequently to answer the question as they may be no longer in formal educational settings, where language and its content are rigorously assessed and emphasized.

Last but not least, Group A are ‘digital natives’ (Marsh, 2005) who are surrounded by multimodal information, videogames and social media and interact with and process the knowledge, parallel process and multi-task faster than their prior generation (Prensky, 2001). Using the internet also involves reading a lot of creative and non-linear multimodal texts (e.g., hypertext and hyperlinks) with eye-catching visual effects in wide varieties of formats, lengths, and topics, which constantly compete for the attention of the reader (Afflerbach & Cho, 2010). Therefore, the students might power down and lose interest when they read static linear multimodal texts which are different from the digital multimodal world and may end up with lower visual engagement and skipping several words, which demonstrate shorter reading duration.

5.3. Eye-tracking experiment

The results obtained from the eye-tracking experiment indicate that the participants primarily directed their initial attention towards the text, and they devoted a significant portion of their time to reading the verbal stimuli. The results support the previous findings that L2 readers spend more time reading the text than the image (e.g., Rayner, 1998; Serrano & Pellicer-Sánchez, 2022; Tragant & Pellicer-Sánchez, 2019). The findings also invalidate the concerns about the potential distracting effect of the image on attention to the text (Tragant & Pellicer-Sánchez, 2019). Conversely, the participants processed verbal information in the presence of pictures and, actually, spent most of their time processing the text. Furthermore, a significant positive correlation was observed between the word counts and the frequency of regressions in the text. This relationship can be attributed, in part, to the fact that regressions account for approximately 15% to 25% of eye movements during normal reading (Rayner, 1989) and typically occur immediately to the left of the previously fixated word (Vitu & McConkie, 2000). On the other hand, regressions tend to be influenced by the complexity of the text and primarily occur due to comprehension difficulties. In other words, as the text becomes more challenging to process, the number of regressions increases, and the tendency to skip words decreases (Castelhano & Rayner, 2008; Frazier & Rayner, 1982).

In this study, regressions could potentially be attributed to the increased difficulty readers encountered in retaining larger amounts of information in their short-term memory. As a result, more cognitive effort was required, leading to a higher frequency of regressions during reading as a means to refresh their memory of previously read content. The results further demonstrate that as the frequency of regressions increased, there was a

correspondingly higher number of fixations on the images. This observation suggests that visual information, as supported by Tang (1992) and Wright (2010), played a beneficial role in aiding second language (L2) reading comprehension. The inclusion of pictorial elements assists in generating additional contextual information and establishing connections between the verbal and visual systems, thereby facilitating the process of reading comprehension (Tang, 1992). Notably, pictures contribute to the creation of a mental representation of the text (Boerma et al., 2016), serving as an initial framework upon which readers can construct their mental models (Eitel & Scheiter, 2015).

The results also show that while the number of fixations on the texts was much higher than the visual discourse, there was a tendency for the fixation durations on the images to be longer than the fixation duration on the written texts. The findings are consistent with the broader literature on differences in eye movements when viewing text and pictures (Carroll et al., 1992; Rayner et al., 2008; Rayner & Castelhano, 2008; Rayner, 2009; Rayner et al. 2001). In the written text, the information is more densely packed and eyes need to recognize letters and eye movements tend to be much smaller than when viewers look at the images, thus the number of fixations on texts increase (Rayner, 1998). Accordingly, the tendency for longer fixation durations on the pictures could be due to the fairly wide field of view in the images during which useful information could be collected (Rayner, 2009).

5.4. Integration of online and eye-tracking experiments

Adopting an exploratory approach, this section discusses the findings of 6 multimodal texts in the online and eye-tracking experiments. As stated by Castelhano and Rayner (2008), the quantity and duration of fixations directly correspond to the level of cognitive and perceptual effort required to process the information, serving as indicators of the difficulty and complexity involved in reading comprehension. According to the eye-tracking experiment results, more fixations and complete fixation time on the image and longer fixation durations on the text and the image in the TSI-Vegetables might indicate the participants' cognitive effort to integrate the information in both modes to fill in the missing information between the verbal and imagery systems and later map them onto each other for a coherent mental model. The discrepancy between the text and the image in the TSI-Vegetables was likely to be a stronger incentive for the readers to examine the picture more closely in order to provide additional contextual information (Tang, 1992) and fill in the semantic gap between the image and the text. However, the IST-Student,

in which the picture only repeats the concept of ‘hardworking student’, does not add any other contextual information and may consequently discourage further scrutiny of the image (Molitor et al., 1989). The complexity of the picture in the TSI-Vegetables (e.g., Oliver’s close shot and facial expression, warm colors) might also encourage more eye fixations. Therefore, the mere presence of the images did not necessarily encourage multimodal reading and that the participants were actively engaged with more complex and interactive multimodal texts.

According to the online experiment results, on the one hand, the presence of the information ‘vegetables’ in the image and subsequent close intermodal interactions in the TSI-Vegetables encouraged high percentage of visual responses. On the other hand, the visual and verbal mental models in the IST-Student were mapped onto each other to reinforce the concept of ‘hardworking student’, which resulted in a large percentage of visual responses. In other words, the high percentage of the visual responses in the IST-Student could be due to the redundancy (Lee & Mayer, 2018) and the fact that both modes repeat the hardworking personality of the student.

With regard to the image-text enhancement relation, the results suggest that the IEP-Hotel exhibits greater intermodal interactions compared to the TET-Reading. Furthermore, the participants exhibited longer fixations on the text and a higher number of fixations on the image in the IEP-Hotel. These findings might suggest that the inclusion of the word 'here' in the IEP-Hotel, which encompasses both London and the hotel, could potentially generate confusion or uncertainty by creating a semantic gap. As a result, the participants exhibited more intermodal regressions, experienced higher cognitive demand, increased fixations on the image, and longer fixation durations on the text in order to resolve the reference to the word 'here'. The visual density of the image, along with the presence of multiple characters, might also contribute to an elevated number of fixations on the image. In contrast, the limited intermodal interactions observed in the TET-Reading might be attributed to the text enhancing the image by providing specific information about Anne's sleep time (midnight), while the image itself does not introduce new elements but merely reinforces the concept of 'a girl reading in bed'. The reason behind the greater number of verbal responses in both multimodal texts in the online experiment may also be due to the fact that the accompanying images did not contain repeated or reinforced information about the location (London) and time (midnight) in the text.

Regarding the complementary relationship between image and text, the findings indicate that the ITC-Jungles elicited longer fixations on the text compared to the ITC-

Earth. However, both multimodal texts evoked a relatively similar level of eye-movement measurements. The outcomes of the online experiment additionally indicate that certain extralinguistic factors, such as top-down mechanisms, might influence the reading of multimodal texts. This implies that the analysis of multimodal input is influenced by the observer's prior experience, knowledge, and cognitive processes (e.g. Hwang et al, 2009; Pelowski & Akiba, 2011; Tinio, 2013; Wang & Pomplun, 2012).

The attentional allocations and scanpath graphs also show that there were some interconnected stages in multimodal reading during which the number of fixations and fixation durations changed. The reading process started off with briefly skimming the written text immediately followed by a few fixations on the image (Stage 1). This stage shows that texts capture attention early in the process of inspection (An et al., 2017). At this stage, there were fewer and shorter eye fixations because the participants quickly skimmed for specific facts in the text and the image to get a general impression of the message. In the later stage, Stage 2, there was an increase in the number of fixations and fixation durations on the verbal discourse, potentially reaching a peak. This pattern might suggest a shift towards a less automatic and more deliberate processing mode (Hy¨on¨a & Lorch. 2004; Hy¨on¨a & Nurminen, 2006). Thus, it is cognitively much more demanding than the first stage. This stage also shows that the readers did not alternate back and forth between the text and the picture, they gave early priority to the text (Rayner et al., 2001), and the comprehension process was largely text directed (Hegarty, 1992a, 1992b). The visual engagement mostly increased relatively after reading the text (Stage 3) (Carroll et al., 1992; Rayner et al. 2001). During this stage, the participants made efforts to combine and establish connections between the visual and verbal information. In the final stage, as the participants aimed to conclude their reading task, there was a noticeable decline in both the number and duration of fixations. This reduction could potentially be attributed to their repeated exposure to the presented information and their growing familiarity with the text (Pieters et al., 1996).

To summarize, extended fixations might suggest that information is either engaging due to visual complexity or confusing and demanding for readers to process. This could be also interpreted as an effort to integrate text and image and construct a visual-verbal mental model. Additionally, a greater number of fixations on image and text likely indicates readers' attempts to integrate information from both modes, accompanied by increased cognitive effort. Schwonke et al (2009) argue that the frequency of transitions

between image and text serves as an indicator of integrative effort, even if not necessarily successful or informed.

The research conducted by O'Halloran (2011) emphasizes the complexities introduced by multimodality in communication, as it brings forth diverse semiotic resources that individuals must navigate and comprehend. Additionally, studies by Kress and van Leeuwen (2001) highlight the unique affordances and constraints associated with different modes, further complicating the communication process. Therefore, while multimodality presents the potential for enhanced and engaging communication experiences, it also presents challenges in terms of information processing, interpretation, and the coordination of multiple modalities.

The findings of this study, which are derived from the analysis of the multimodal reading test, underscore the significance of recognizing the limitations inherent in the research instrument. Therefore, the subsequent chapter delves into a comprehensive evaluation of the test's reliability and validity, takes into account the potential drawbacks and constraints that may influence its accuracy, and offers valuable insights and recommendations to address these limitations.

CHAPTER 6

INSIGHTS INTO THE MULTIMODAL READING TEST

Within this section, careful scrutiny is given to assessing the multimodal reading test's reliability and validity, taking into account potential drawbacks and limitations that may affect its accuracy. Additionally, valuable insights and recommendations are put forth to address these shortcomings and optimize the test's performance. Section 6.1 delves into the examination of the reliability and validity of the multimodal reading test, while section 6.2 sheds light on the obstacles encountered when implementing the test, explores its limitations, and offers recommendations to address them.

6.1. Reliability and validity of the multimodal reading test

This section aims to provide a comprehensive evaluation of the reliability and validity of the test employed in this study. Reliability refers to the consistency and stability of the test's results, while validity pertains to the extent to which the test measures what it is intended to measure. Ensuring the validity and accuracy of the research findings, the assessment of test reliability and validity holds significant importance. To measure internal consistency, the Cronbach's alpha coefficient was utilized. Cronbach's alpha coefficient for our test was calculated to be 0.70. This indicates an acceptable level of reliability. Generally, a Cronbach's alpha value of 0.7 is considered satisfactory, suggesting that the items in the test are fairly reliable and contribute to the overall consistency of the measure. However, it is important to note that a higher value would indicate stronger internal consistency.

In order to assess the validity of the test, a Pearson Correlation analysis was conducted to determine the correlation between each question in the multimodal reading test and its overall score. If the significance value is equal to or less than 0.05, it indicates that the question is considered valid.

Table 23 shows that the correlation between each question in the multimodal reading test and its overall score is significant. The significant p-values indicate that there is strong statistical evidence supporting the validity of the multimodal reading test and the observed correlations are unlikely to have occurred due to chance alone. Therefore, these results provide support for the notion that the test accurately measures the construct of interest.

Table 23. The results of the Pearson Correlation between each question in the multimodal reading test and its overall score

Image-text relations	Pearson correlation	Sig. (2-tailed)
ITI-Rain	0.29	0.01
ITC-Earth	0.52	0.00
TGI-Volleyball	0.37	0.00
IGT-Lake	0.37	0.00
IST-Granny	0.37	0.00
TSI-Vegetables	0.40	0.00
IEP-Airport	0.25	0.03
IER-Football	0.25	0.03
TET-Reading	0.41	0.00
ITT-Shopping	0.50	0.00
ITI-Routine	0.40	0.00
ITC-Jungles	0.34	0.00
TGI-Violin	0.43	0.00
IGT-Fruits	0.42	0.00
IST-Student	0.28	0.01
TSI-Coffee	0.46	0.00
IEP-Hotel	0.34	0.00
IER-Birthday	0.51	0.00
TET-Picnic	0.34	0.00
ITT-Address	0.52	0.00

While the multimodal reading test has demonstrated satisfactory reliability and validity, they do not necessarily imply practical significance. Therefore, the subsequent section aims to address the challenges encountered with the test during the research and explores the avenues for further development.

6.2. Unveiling Challenges: Exploring test limitations and issues in research

In the realm of research, tests and questionnaires serve as valuable tools for gathering data and insights. However, like any methodological approach, they are not immune to limitations and issues that can arise during their implementation. This chapter delves into the exploration of these challenges, aiming to shed light on the potential limitations and

issues associated with the multimodal reading test in research. Unveiling these challenges will help improve future research outcomes.

6.2.1. Some questions are confusing

The ITI-Routine text initially states, "Jack gets up. He has breakfast. He takes a shower." Nevertheless, the corresponding question asks, "What does Jack do when he is awake?" This wording could potentially create confusion as the question aims to identify the participants' choice for Jack's first action upon waking up. To enhance clarity and minimize ambiguity, it would be more suitable to revise the question as follows: "What does Jack do when he gets up?" This modified wording explicitly indicates the specific moment when the action occurs, thereby eliminating any potential confusion arising from the term "awake."

Moreover, the text in the IGT-Fruits states, "Emma is strong. She likes bananas." However, the accompanying question asks, "Why is Emma healthy?" This question could potentially be misleading for the students, as the inclusion of the word "healthy" might prompt them to choose the option "fruits," which seems more logical in relation to health. To improve clarity and minimize confusion, it would be more appropriate to rephrase the question as follows: "What does Emma like?" This revised question focuses on the intended inquiry about Emma's preferences, avoiding any potential ambiguity or misinterpretation arising from the term "healthy.

The IST-Granny text also describes "David’s grandmother had grey hair. She was always happy. She helped the family a lot." The question is, "Why was David's grandmother a nice person?" The use of the word "nice" may inadvertently guide participants to choose the option "because she helped the family a lot" instead of "because she was always happy," as it may seem more logically connected to being nice. To address this potential bias and encourage a broader perspective, it would be preferable to rephrase the question as follows: "What does David appreciate about his grandmother?" or "Why does David like his grandmother?". This revised question focuses on exploring David's personal views and allows for a wider range of responses, capturing various aspects of his grandmother's character beyond a singular label like "nice."

Additionally, the text in the IST-Vegetables does not explicitly include the word 'eat,' but it is implicitly conveyed through the image. Consequently, it would be advisable to remove the word 'eat' from responses that indirectly refers to the information depicted in the image.

In the ITT-Shopping, the text states that "Mary is in a clothes shop. She likes buying a dress. She likes the yellow dress." However, the accompanying image shows the girl holding a red dress and displaying a cheerful smile. The question originally asks, "What color dress does she want to buy?" This question should be modified to "What color dress does she like?" since her preference for a particular color may differ from the one she ultimately purchases.

Finally, in the IST-Student, the question asks ‘What made Oliver a good boy?’. This question could be perplexing since both "helping friends" and "studying hard" are capable of making individuals good, and it doesn't necessarily prompt students to seek information from the image or text. Therefore, it would be preferable to substitute it with a question that doesn't rely on the reader's moral judgments.

6.2.2. Some texts are confusing

Based on the online and eye-tracking experiments, it was observed that the students spent the most time on the ITI-Rain and frequently shifted their focus between the image and text. These findings might suggest that the ITI-Rain likely provides a confusing explanation of the water flow from clouds to the sea. Consequently, there is a need to replace it with a more structured and comprehensible text that effectively conveys the process.

The TSI-Coffee was also found to be confusing. This confusion may arise from the presence of three objects on the desk (laptop, cup, car key) and the pronoun 'it' in the sentence 'He is putting it on the desk,' which could initially be associated with any of these objects. Therefore, it is advisable to include an image that depicts only one object that aligns with the pronoun mentioned in the text. This will help provide clarity and enhance understanding for the readers.

6.2.3. Texts have different length and level of difficulty

I encountered challenges when comparing the multimodal texts, particularly those with similar image-text relations like TGI-Volleyball and TGI-Violin, due to differences in their word counts and word frequencies. The eye-tracking experiment revealed that the difficulty of the text influenced the eye movements between the image and the text and the amount of time spent on them. Additionally, in the context of the online multimodal reading, the presence of varying word counts and frequencies made it challenging to determine whether the reading duration was influenced by these textual characteristics or by the specific image-text relations. Therefore, it is crucial to design future research tests

that incorporate texts with uniform word and character count and frequency to address this issue.

6.2.4. Image design affects multimodal reading

Comparing the multimodal texts posed challenges due to variations in image sizes and formats, particularly when texts share the same image-text relations. For example, the multimodal texts ITI-Rain and ITI-Routine have identical image-text relations, but the image sizes and formats differ. The image in the ITI-Rain is a square picture, while the image in the ITI-Routine is a horizontal rectangle, impacting the eye movements between the image and text.

Furthermore, the heat map from the eye-tracking experiment revealed that faces in the images captured attention. Since some images lack characters or faces, comparing multimodal texts with the same image-text relations became difficult. For instance, the texts in the IGT-Lake and the IGT-Fruits share the same image-text relations, but the presence of Emma in the IGT-Fruits, with her captivating pose, blinking smile, and muscular arm, potentially resulted in higher visual engagement among participants. In contrast, the IGT-Lake only featured a picture of a house and a lake, which might not be as visually appealing as Emma's pose. Therefore, it is important to include images of consistent sizes, shapes, and complexity in future research to ensure fair comparisons.

6.2.5. Lack of central focal points might affect the reading process

The participants were not given specific instructions regarding where to begin their reading process, which could potentially account for their shorter time until first fixation on the text. To mitigate this issue, it would be advantageous to include a fixation point at the center of the screen before introducing any slides. This would assist in directing their gaze and establishing a central focal point for their eyes to focus on. Additionally, it is important to present the questions and responses in the middle of the page rather than on the left side. This is necessary to avoid biasing the readers towards texts that were previously presented on the left side. By maintaining a central positioning, the presentation of questions and responses can be more balanced and impartial, fostering fair evaluation and interpretation.

6.2.6. The questions and responses include different word counts and frequencies

The variation in word counts and frequencies among the questions and responses posed challenges when comparing response durations in the multimodal texts. Addressing this

issue would significantly enhance the research's value. For future studies, it is crucial to design responses and questions with consistent word count and frequency. This approach would enable a meaningful comparison of response durations in multimodal texts with different image-text relations, and a precise analysis of response duration in relation to visual and verbal responses. Such an investigation would provide valuable insights into this area of study.

6.2.7. The verbal response is in the first sentence

The placement of visual responses in the second sentence and verbal responses in the first sentence might introduce bias in participants' responses, favoring the verbal response. To address this, it is recommended to alter the order of the visual and verbal responses. In other words, in one set of image-text relation the first multimodal text should feature the verbal response in the second sentence, while the second multimodal text can have the verbal response in the first sentence. This adjustment can help mitigate the potential bias towards the verbal response.

In summary, the statistical analysis showcased a satisfactory level of validity and reliability for the multimodal reading test. However, taking into consideration the aforementioned issues and addressing them would greatly improve the test's reliability and validity.

Next chapter seeks to synthesize the research findings, explore their implications, and propose actionable recommendations that can positively impact the field of study while addressing the broader research question.

CHAPTER 7

CONCLUSION

Reading a powerful and indispensable way of learning, guarantees academic and professional success (Hassan et al, 2021) and has a key role in second or foreign language education (Grabe, 1991). It is thus essential for L2 learners to form good reading habits which teach life-long learning experience and activate their thinking power. However, the traditional reading habits have been gradually replaced with the modern technological equipment and the stimulating environment of the internet, which have resulted in reduced contacts with books and other reading materials. Qenaway (2019) highlights that while social media offer advantages in terms of incorporating interactive modes that engage readers, their extensive impact has caused a decrease in the reading practices of instructors and students in second languages. Additionally, this dominance has brought about changes in their reading behaviors both within educational settings and outside of them.

Second language learning is not a linear monomodal, but it rather represents the incorporation of different modes. From the very early stages, L2 learners are exposed to, interact with and comprehend multimodal texts. Pictures in reading passages serve a variety of functions, such as depicting and supporting the message of text, making reading experience more engaging and creating a context for making predictions and meaning making (Wright, 1989). Second language learners are also actively engaged with multimodal texts, in which images are conceived to support reading comprehension and foster learning (Pellicer-Sánchez, 2022; Wright, 1989). Thus, it is vital for teachers and researchers to carry out reception studies, which enhance their understanding of how readers integrate different semiotic resources for meaning making (Pellicer-Sánchez et al, 2020). However, research into this area is highly challenging due to the difficulty of mapping learners' cognitive processes in real time. Therefore, despite the widespread use of imagery in L2 learning materials such as textbooks and the significance of understanding the cognitive processes involved, few studies have investigated the processing of text and image in multimodal L2 reading (Pellicer-Sánchez, 2022). To fill this gap, the present study adopts a mixed-method approach (online and eye-tracking experiments) to investigate the effect of age on L2 learners' reading habits and modal preferences and explore how L2 learners read multimodal texts.

The online experiment attempted to investigate L2 learners' reading habits and modal preferences. The results indicate that the participants were mostly involved in social media than print and electronic media. However, the findings from the multimodal reading test show that the participants preferred verbal to visual responses. Moreover, a positive correlation was found between the reading duration and the visual responses, that is the longer the reading duration was, the more the number of visual responses were. In contrast, the verbal responses were negatively correlated with the reading duration. In addition, most of the participants believed that images helped their reading comprehension, which subsequently increased the number of their visual responses. They also found the interaction of visual and verbal discourse more beneficial to answer the questions than each mode singly, which consequently increased the number of visual responses and the reading duration.

With regard to the age of the participants, Group A (13-30) read print and electronic books and comics and used social media to watch videos more frequently than Group B (31-46). Group B used social media to keep in touch and read print and electronic newspapers and magazines more regular. However, the results of the Mann-Whitney U test demonstrated a statistically significant difference between Group A and Group B solely in terms of their engagement with reading pictures and watching videos on social media. The findings of the online multimodal reading test show that as the participants got older, they demonstrated stronger visual preferences. Additionally, Group B recognized the contribution of images and their interactions with the texts to their reading comprehension while Group A considered verbal information more informative. The Gabriel post-hoc test results also show that the reading and response speed tended to increase from adolescence (13-17) to young adulthood (24-29), however, it started to decrease towards the middle age (39-46).

The eye-tracking experiment aimed to investigate L2 learners' multimodal reading strategies. The descriptive statistics of the eye-measurements reveal that the participants spent most of their time and fixated more on the texts. However, the fixation durations tended to be longer on the images than the text. Moreover, as the number of the words in the texts increased, the complete fixation time, the number of fixations and regressions, and time until the first fixation on the image escalated, which is likely to indicate that the participants preferred to have a quick skim over the whole text before looking at the image. Regressions in company with the fixations on the images might also support reading comprehension and deal with linguistic difficulty.

Furthermore, some interwoven stages in multimodal reading were detected when the number of fixations and fixation durations changed. Stage 1 started with skimming the text, when the number of fixations and fixation durations were low. Stage 2 involved longer and higher number of fixations on the text because participants started to process the new information in the text. Stage 3, which involved visual engagement, started relatively after the text was read by the participants. Visual-verbal relations were likely to affect the amount of the L2 learners' eye-movements between the visual and verbal stimuli at this stage. In Stage 4, the number and duration of fixations dropped as the participants started to finish the reading task. Moreover, Visual-verbal relations (e.g., text-subordinate-to-image and image-subordinate-to-text, image-text-complementary) might determine the degree of the participants' involvement with the image.

Together the findings of two experiments suggest that, on the one hand, although L2 learners regularly spent time in the multimodal environment of social media, they mainly considered text as the main source of information in texts with educational purposes. On the other hand, different age groups might grow different modal preferences in multimodal reading tasks. One reason might be that they grew different reading habits. Moreover, although the participants followed relatively similar stages to read multimodal texts, visual-verbal relations might change the eye measurements and movements between text and image.

7.1. Limitations of the study

It is important to acknowledge the limitations of the present study. Only nine participants attended the eye-tracking experiment based on convenience sampling and voluntary response sampling because the research was conducted during COVID-19 lockdown, when in-person and on-campus activities and the access to the laboratory equipment were strictly limited. When the lockdown was eased, six more participants were selected for the eye-tracking experiment, however, the eye-tracker was not available because either it was being used by other departments of university for their research or it wasn't working stably. When I finally managed to make a two-week appointment to use the eye-tracker, the Faculty of Engineering and Information Technology at the University of Pannonia announced that the eye-tracker was not accessible anymore due to software errors and bugs.

Moreover, a set of print standardized Pannon language exam tests was initially selected as the language proficiency test, which was handed in by the teachers of a secondary

school in Veszprém to 40 L2 students before COVID-19 started. Twenty eight participants received the passing score for the eye-tracking experiment. However, due to COVID-19 restrictions and school closures, onsite access to the students and collecting further data were not possible anymore, therefore, they were excluded from the final population.

Two different populations were used for data collection. When the first selected population (28 participants) was excluded from the research, I contacted several other secondary schools in Veszprém to recruit more participants. Nine secondary school students finally attended the online experiment, however, the COVID-19 pandemic created serious difficulties recruiting secondary school participants, I thus decided to reformulate the research questions, alternate the subject parameters and enlist the students at the language school where I worked and could easily access. Although there were two different EFL proficiency tests (Euroexam test and placement test) for two different populations (secondary school students and language school students), the data of nine secondary school students were not excluded from the final research population because I believed that they could provide an invaluable insight into school students' reading habits and modal preferences and generate constructive ideas for future research.

Due to the data collection being conducted exclusively online during the COVID-19 lockdown, the standardized Pannon language test was not utilized as the English language proficiency test. The Pannon Language Examination Center declined permission to administer the test online due to concerns regarding data protection, access, and security. Initially, I chose the Pannon Language Exam due to its inclusion of multiple reading texts that aligned with the objectives of my research. Furthermore, Hungarian universities and employers acknowledge this examination as a reliable indication of language proficiency. Additionally, the test adhered to standardized criteria, ensuring consistency and fairness.

The aforementioned external limitations imposed on my pre-existing research plan negatively impacted the research progress for almost a year and a half and consequently resulted in frequently reformulating the research questions, restructuring the research plan and procedure and altering the subject parameters.

7.2. Implications of the research

Textbook developers, web and app designers and teachers can benefit from the present research. Textbook developers need to integrate multimodal texts with different degrees of redundancy to provoke different interpretations from L2 learners, engage them in

multimodal thinking and cognitive flexibility and create multidirectional entry points into textual analyses and interpretations (Ajayi, 2012). Web and app designers can also adopt multimodal input with different intermodal relations. While, at the beginning of the learning process, redundancy in multimodal texts might provide a firm ground for L2 learners to learn, for instance, vocabulary, semantic gaps can later create opportunities for them to practice the knowledge and increase its retention. Moreover, teachers can effectively incorporate different modes in teaching materials to develop student's engagement, interpretation and attention. Duplications and semantic gaps in multimodal texts could scaffold language learning and develop learners' critical thinking and engagement, respectively. Although the use of technology may facilitate teaching and learning in language classrooms (Derakhshan & Hasanabbasi, 2015), the mere induction of social media does not bring about valuable learning experience. In fact, teachers need to use the available ESL reference resources to plan and design collaborative activities on social media, which engage the learners in language experience both effectively and cognitively and improve their communicative competence and purposes (Nur & Syarifuddin, 2018). Educational institutions should also consider the modification of ESL/EFL curricula and the incorporation of social media into education, hoping to offer more meaningful engagement and enriched learning experience to deal with students with different needs and preferences.

Different reading habits and modal preferences among L2 learners of different age ranges may encourage teachers to modify their approaches according to students' particular needs and interests and tailor teaching materials and assignments in various formats to support their students during learning processes and completion of assignments. For instance, teaching genres as a particular type of writing, the teacher may assign young adults to look into trendy comics on social media while middle-aged adults can explore electronic newspaper genres.

The findings from this research also demonstrate a clear gap between inside and outside school world. While the reading habit questionnaire shows the students' favor for social media, their responses to the multimodal texts and the eye-tracking experiment evidence that they relied on text more often. Therefore, in order to improve multimodal literacy at schools the reconceptualization of assessment and the incorporation of multimodal tasks, which encourage students to display metacognitive and multimodal literacy skills, are recommended. Teachers and curriculum designers can develop

meaningful multimodal assessment tools and create attractive and useful interactive materials, which maximize learning potentials.

The present research can help material and textbook developers, web and app designers and teachers to analyze available multimodal materials in educational settings to see if they are pacing up with the current needs of students and help develop student's engagement, interpretation and attention through an effective combination of semiotic modes in teaching materials. In sum, the present piece of research emphasizes that the age of technology has changed the reading habits. While it is important to recognize the importance of traditional reading habits on second language proficiency, it is essential for teachers and curriculum designers to listen, learn and adjust their approach to pace up with the changing needs of students, tailor the curricula, teaching, learning and assessment materials, and assignments accordingly and design lessons which lend themselves to different types of learning.

7.3. Future directions

Only nine participants attended the eye-tracking experiment because the research was conducted during COVID-19 lockdown. Future research could include a larger sample of participants to investigate, refine and generalize our findings related to L2 learners' multimodal reading processes.

Moreover, the present research is about the effect of social media on reading multimodal texts. And since using the internet involves reading a lot of creative and eye-catching visual effects in wide varieties of formats, lengths and topics, the multimodal reading test includes texts with varied image-text relations. Further research is required to focus on only two image-text relations (e.g., image-more-general-than-text and text-more-general-than image) and include more trial per relation.

Besides, this paper only examined the modal preferences of the participants and their mental representations in the online experiment. Further research could explore which image-text relation (e.g., image-subordinate-to-text or text-subordinate-to-image) develops a better comprehension of the multimodal texts.

Furthermore, the present research only presents the preliminary results of the multimodal reading test from the pilot study. The findings from the online and eye-tracking experiments could be used to modify and validate the present multimodal reading test for future replications.

In addition, eye-tracking experiments, read-aloud protocols and interviews may provide further detailed data about L2 learners' cognitive processes while reading and responding to multimodal texts. This valuable information may help researchers investigate how readers deal with, recall and retrieve multimodal information.

Further research could also explore the broader impact of autonomous learning on individuals' engagement with the external environment, as well as its influence on their motivation, reading habits, and preferred modes of learning. Investigating these aspects can shed light on the potential implications for future educational studies.

To enhance the validity and potentially the reliability of the research, an option could be to replicate the study using open-ended questions within the online multimodal reading test. Instructing participants to respond with brief and straightforward sentences would contribute to the overall robustness of the findings.

Furthermore, this study could be replicated with L2 Learners from different countries to investigate the effect of culture on their multimodal reading. More research could be also done to examine which modal preferences (visual or verbal) lead to a better reading comprehension

Finally yet importantly, more research is required to investigate the drawbacks and benefits of different research instruments and explore which combinations of methods demonstrate better and more comprehensive results. We strongly believe that collecting further empirical evidence contributes to developing a more comprehensive and efficient model of L2 multimodal language learning environment.

References

Abidin, M. J. Z., Ahmad, N., & Kabilan, M. K. (2010). Facebook: An online environment for learning of English in institutions of higher education. Internet and Higher Education, 13(4), 179–187.

Abraham, P., & Farías, M. (2017). Reading with eyes wide open: Reflections on the impact of multimodal texts on second language reading. Íkala, Revista de Lenguaje y Cultura, 22(1), 57–69.

Afflerbach, P. A., & Cho, B. Y. (2010). Determining and describing reading strategies: Internet and traditional forms of reading. In H. S. Waters & W. Schneider (Eds.). Metacognition, strategy use, and instruction (pp. 201-255). New York: Guilford Press.

Ajayi, L. (2012). How teachers deploy multimodal textbooks to enhance English language learning. TESOL Journal, 6, 16–35.

Alderson, J., & Wall, D. (1993). Does washback exist? Applied Linguistics, 14, 115–129. https://doi.org/10.1093/applin/14.2.115

An, L., Wang, Y., & Sun, Y. (2017). Reading Words or Pictures: Eye Movement Patterns in Adults and Children Differ by Age Group and Receptive Language Ability. Front. Psychol, 8. doi: 10.3389/fpsyg.2017.00791

Anankulladetch, Ph. (2017). The impact of social media on ESL students' learning performance. IJSRT, 1(6), pp. 3535-5754.

Anstey, M., & Bull, G. (2006). The literacy labyrinth (2nd ed.). Sydney, Australia: Pearson Education.

Ardasheva, Y. (2016). A structural equation modeling investigation of relationships among school-aged EL's individual difference characteristics and academic and second language outcomes. Learning and Individual Differences, 47, 194–206. https://doi.org/10.1016/j.lindif.2016.02.010

Arumugam, W., Shanthi, W. S., & De Mello. (2019). Impact of social media on reading comprehension among undergraduates. International Journal of Advanced Science and Technology, 28(8), 457-470.

Baddeley, A. (1992). Working memory. Science, 255(5044), 556–559. https://doi.org/10.1126/science.1736359

Baird, D., & M. Fisher. (2006). "Neomillennial user experience design strategies: Utilizing social networking media to support 'Always On' learning styles". Journal of Education Technology Systems, 34(1), 5-32. https://doi.org/10.2190/6WMW-47L0-M81Q-12G1

Bal, M. (2017). Narratology: Introduction to the theory of narrative (4 ed.). Toronto: University of Toronto Press.

Barthes, R. (1997a). The Photographic Message. In R. Barthes (Ed.). Image–Music–Text (pp. 15–31). London: Fontana.

Barthes, R. (1997b). Rhetoric of the Image. In R. Barthes (Ed.). Image–Music–Text (pp. 32–51). London: Fontana.

Bateman, J. (2014). Text and image: A critical introduction to the visual/verbal divide. Routledge, Milton Park 5.

Beardsley, M.C. (1981). Aesthetics: Problems in the philosophy of criticism. Indianapolis, IN: Hackett Publishing.

Beck, I. L., & McKeown, M. B. (2006). Improving compre¬hension when questioning the author: A fresh and expanded view of a powerful approach. New York, NY, London, Sydney: Scholastic.

Berger, J. (1972). Ways of seeing. London: Penguin.

Bernhardt, E. B. (2011). Understanding advanced second-language reading. New York: Routledge. https://doi.org/10.4324/9780203852408

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: a social semiotic account of designs for learning. Written Communication, 25, 166–195. http:// dx.doi.org/10.1177/0741088307313177

Bezemer, J., & Kress, G. (2010). Changing text: a social semiotic analysis of textbooks. Design for Learning, 3, 10–29.

Bisson, M-J., W. Van Heuven., K. Conklin., & R. Tunney. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. Quarterly Journal of Experimental Psychology, 68, 1306–1326. DOI:10.1080/17470218.2014.979211

Blattner, G., & Fiori, M. (2009). Facebook in the language classroom: Promises and possibilities. International Journal of Instructional Technology and Distance Learning, 6(1), 17–28.

Boerma, I. E., Mol, S. E., & Jolles, J. (2016). Reading pictures for story comprehension requires mental imagery skills. Frontiers in Psychology, 7 (1630), 1–10. https://doi.org/10.3389/fpsyg.2016.01630

Brody, P.J., & Legenza, A. (1981). Can pictorial attributes serve mathemagenic functions? Educational Communication and Technology Journal, 29, 93–100. https://doi.org/10.1007/BF02791377

Bruce, N. D. B., & Tsotsos, J. K. (2006). Saliency based on information maximization. Advances in Neural Information Processing Systems, 18, 155–162.

Burgess, M. Ø. (2016). Fra novelle til film: elevproduserte multimodale tekster. Doctoral thesis, University of Oslo, Oslo.

Calabrèse, A., Cheong, A. M. Y., Cheung, S. H., He, Y., Kwon, M. Y., Mansfield, J. S., Subramanian, A., Yu, D., & Legge, G. E. (2016). Baseline MNREAD measures for normally sighted subjects from childhood to old age. Investigative Ophthalmology and Visual Science, 57(8), 3836–3843. https://doi.org/10.1167/iovs.16–19580

Carney, R. N., & Levin, J, R. (2002). Pictorial illustrations still improve students' learning from text. Educational Psychology Review, 14(1), 5–26.

Carr, N. G. (2011). The shallows: what the Internet is doing to our brains. Norton pbk. [ed.] New York, W.W. Norton.

Carroll, P. J., Young, R. J., & M. S. Guertin. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.). Eye movements and visual cognition: Scene perception and reading (pp. 444- 461). New York: Springer.

Castelhano, M. S., & Rayner, K. (2008). Eye movements during reading, visual search, scene perception: an overview. In K. Rayner, D. Shem, X. Bai & G. Yan (Eds.). Cognitive and cultural influences on eye movements (pp. 175–195). Tianjin: Tianjin People's Publishing House.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and Instruction, 8(4), 293-332. https://doi.org/10.1207/s1532690xci0804_2

Chang, Y., & Choi, S. (2014). Effects of seductive details evidenced by gaze duration. Neurobiology of Learning and Memory, 109, 131–138. doi:10.1016/j.nlm.2014.01.005

Chartrand, R. (2012). Social networking for language learners. Creating meaningful output with Web 2.0 tools. The International Journal of Knowledge Management and E-learning, 4(1), 97–101.

Chauhan, P., & Lal, P. (2012). Impact of information technology on reading habits of college students. International Journal of Research Review in Engineering Science and Technology, 1(1), 101–106.

Chen, X., Fang, H., Lin, T-Y, Vedantam, R., Gupta, S., Doll´ar, P, & Zitnick, C. L. (2015). Microsoft coco captions: Data collection and evaluation server. arXiv preprint arXiv:1504.00325.

Chen, A. H., Khalid, N. M., & Buari, N. H. (2019). Age factor affects reading acuity and reading speed in attaining text information. International Journal of Ophthalmology, 12(7), 1170–1176. https://doi.org/10.18240/ijo.2019.07.19

Celik, B. (2019). The role of extensive reading in fostering the development of grammar and vocabulary knowledge. International Journal of Social Sciences & Educational Studies, 6(1), 215-223. https://doi.org/10.23918/ijsses.v6i1p215

Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in year 2. Journal of Research in Reading 22, 241–256. doi.org/10.1111/1467-9817.00088

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. Educ. Psychol. Rev, 3, 149–210.

Comenius, J. (1985). A didática magna [Didactica Magna]. Lisboa: Fundação Calouste Gulbenkian.

Cullum, S., Huppert, F.A., McGee, M., Dening, T., Ahmed, A., Paykel, E.S., & Brayne, C. (2000). Decline across different domains of cognitive function in normal ageing: Results of a longitudinal population-based study using CAMCOG. International Journal of Geriatric Psychiatry 15, 853–862. https://doi.org/10.1002/1099-1166(200009)15:9<853::AID-GPS211>3.0.CO;2-T

Cunningham, S., Chris, R., & Moor, P. (2005). New cutting edge elementary. Student's book. London, Pearson Education.

Danielsson, K. (2011). Att närma sig en naturvetenska¬plig diskurs. Text och textanvändning i svenska och finlandssvenska klassrum (Getting closer to the discourse of science. Text and text use in Swedish and Finland-Swedish classrooms). In I. Eriksson (Ed.). Kemiundervisning, text och textbruk i finlandss¬venska och svenska skolor – en komparativ tvärveten¬skaplig studie (pp.161-237). Stockholm: SU Förlag.

Danielsson, K., & Selander, S. (2016). Reading multimodal texts for learning—A model for cultivating multimodal literacy. Designs for Learning, 8(1), 25–36.

Depew, K. E. (2011). Social media at academia's periphery: Studying multilingual developmental writers' Facebook composing strategies. Reading Matrix: An International Online Journal, 11(1), 54–75.

Derakhshan, A., & Hasanabbasi, S. (2015). Social networks for language learning. Theory and Practice in Language Studies, 5(5), 1090–1095.

Eckert, M. A., Keren, N. I., Roberts, D. R., Calhoun, V. D., & Harris, K. C. (2010). Age-related changes in processing speed: Unique contributions of cerebellar and prefrontal cortex. Frontiers in Human Neuroscience, 4, 10. https://doi.org/10.3389/neuro.09.010.2010

Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. Educational Psychology Review, 27, 153–180. https://doi.org/10.1007/s10648-014-9264-4

Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. Reading Research Quarterly, 19(1), 53–67. https://doi.org/10.2307/747337

Engebretsen, A. (2012). Balancing cohesion and tension in multimodal rhetoric. An interdisciplinary approach to the study of semiotic complexity. Learning, Media and Technology, 37(2), 145-162. DOI: 10.1080/17439884.2012.655745

Erikson, E. (1974). Identity: Youth and crisis. London: Faber & Faber.

European Commission. (2015). Improving policy and provision for adult learning in Europe. Luxembourg: Publications Office of the European Union.
Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: evidence from eye movement monitoring. Psychol. Sci, 16, 913–920. DOI: 10.1111/j.1467-9280.2005.01636.x
Faghri, F., Fleet, D. J., Kiros, J. R., & Fidler, S. (2018). Improving visual semantic embeddings with hard negatives. Presented in BMVC conference, Northumbria, England.
Faust, M. (1994). Alone but not alone: Situating readers in school contexts. In B. Corcoran, M. Hayhoe, & G.M. Pradl (Eds.). Knowledge in the making: Challenging the text in the classroom (pp. 25–33). Portsmouth, NH: Boynton-Cook.
Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. Learning and Instruction, 20, 216–224. https://doi.org/10.1016/j.learninstruc.2009.02.021
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178–210. https://doi.org/10.1016/0010-0285(82)90008-1
Fusco, E. (1986). Reading interests and cognitive development. SIGNAL, 11(2), 1–3.
George, M. S. (2018). Developing listening and reading skills through social media using apps. An Interdisciplinary Journal of Literary Studies, 93-101.
Gibbons, P. (2002). Scaffolding language, scaffolding learn¬ing: Teaching second language learners in the main¬stream classroom. Portsmouth, NH: Heinemann.
Godfroid, A. (2020). Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide. New York, NY: Routledge.
Gombrich, E. H. (1961). Art and illusion: A study in the psychology of pictorial representation (2nd ed.). Princeton, NJ: Princeton University Press.
Grabe, W. (1991). Current developments in second language reading research. TESOL Quarterly, 25(3), 375–406. https://doi.org/10.2307/3586977
Grady, C. L., & Craik, F. I. (2000). Changes in memory processing with age. Current Opinion in Neurobiology, 10(2), 224–231. https://doi.org/10.1016/S0959-4388(00)00073-8
Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. Annu. Rev. Psychol. 48, 163–189.
Haglund, J. (2013). Collaborative and self-generated analogies in science education. Studies in Science Education. http://dx.doi.org/10.1080/030572 67.2013.801119
Häikiö T., Bertram R., Hyönä J., & Neimi P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. Journal of Experimental Child Psychology, 102, 167–181. doi:10.1016/j.jecp.2008.04.002.
Halliday, M. A. K. (1994). An introduction to functional grammar (2nd ed.). London: Arnold.
Halliday, M. A. K., & Martin, J. R. (1993). Writing science: Literacy and discursive power. Pittsburgh, PA: University of Pittsburgh Press.
Halliday, M., & Matthiessen, C. (2004). An introduction to functional grammar (3rd ed.). London: Arnold.
Hamilton, B. (2009). Young adult literature. Library media Connection, 28(3), 14–15.
Hardiess, G., & Weissert, C. (2021). Interaction between image and text during the process of biblical art reception. Journal of Eye Movement Research, 13(2) https://doi.org/10.16910/jemr.13.2.14
Hartman, D. K. (1992). Intertextuality and reading: The text, the reader, that author, and the context. Linguistics and Education, 4, 295–311.

Hassan, I., Azmi, M. N. L., Muhamad, S. N., & Abdullah, A. T. H. (2021). Reading Habits and their Correlation with Reading Achievement among ESL Learners in Selected Malaysian Secondary Schools. Arab World English Journal, 12 (3) 385-399. https://doi.org/10.24093/awej/vol12no3.27

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. Trends in Cognitive Sciences, 9(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Hegarty, M. (I992a). The mechanics of comprehension and comprehension of mechanics. In K. Rayner (Ed.), Eye movements and visual cognition: Scene perception and reading (pp. 428-443). New York: Springer-Verlag.

Hegarty, M. (1992b). Mental animation: Inferring motion from static diagrams of mechanical systems. Journal of Experimental Psychology: Learning, Memory & Cognition, 18, 1084-1102.

Hertel, T. (2003). Using an e-mail exchange to promote cultural learning. Foreign Language Annals, 36(3), 386–396.

Holsanova, J. (2014). Reception of multimodality: Applying eye tracking methodology in multimodal research. In C. Jewitt (Ed.). Routledge handbook of multimodal analysis (2nd edition, pp. 285–296). London: Routledge.

Hopper, R. (2005). What are teenagers reading? Adolescent fiction reading habits and reading choices. Literacy (Formerly Reading), 39(3), 113–120. https://doi.org/10.1111/j.1467-9345.2005.00409.x

Hung, H., Chiu, Y. J., & Yeh, H. (2013). Multimodal assessment of and for learning: A theory-driven design rubric. British Journal of Education Technology, 44(3), 400–409. https://doi.org/10.1111/j.1467-8535.2012.01337.x

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. Journal of Vision, 9(5), 1–18. https://doi.org/10.1167/9.5.25

Hy¨on¨a, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. Learning and Instruction, 14, 131–152. https://doi.org/10.1016/j.learninstruc.2004.01.001

Hy¨on¨a, J., & Nurminen, A. M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. British Journal of Educational Psychology, 97, 31–50. https://doi.org/10.1348/000712605X53678

İyigün, E., Bekircan, M., & Maviş, İ. (2018). Yetişkin Bireylerde Yaş, Eğitim ve Cinsiyet Değişkenlerinin Okuma Hızına Etkisi . Dil Konuşma ve Yutma Araştırmaları Dergisi, 1(2), 162–178.

Jakobsen, I. K. (2019). Inspired by image: A multimodal analysis of 10th grade English school-leaving written examinations set in Norway (2014-2018). Acta Didactica Norge, 13(1), 1–27. https://doi.org/10.5617/adno.6248

Jewitt, C. (2005). Technology, literacy, and learning: A multimodal approach. London: Routledge Falmer.

Jewitt, C. (2009). The Routledge handbook of multimodal analysis. London: Routledge.

Johnson-Smaragdi, U., & Jönsson, A. (2006). Book reading in leisure time: Long-term changes in young people's book reading habits. Scandinavian Journal of Educational Research, 50(5), 519-540.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. Journal of Experimental Psychology: Applied, 18, 178–191. doi:10.1037/a0026923

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split attention and redundancy in multimedia instruction. Applied Cognitive Psychology, 13, 351-371.

Keln, O. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. Foreign Lang Ann, 25(2), 441- 454.
Kitsis, S. M. (2008). The Facebook generation: Homework as social networking. English Journal, 98(2), 30–36.
Kornalijnslijper, D.S. (2007). A study of three models for image-text relations. Retrieved from https://pdfs.semanticscholar.org/fba4/243011cf80cffe36d333368bffd653e9ca9b.pdf?_ga=2.206275856.246584535.1587158450-2045160697.1585391590
Kress, G. (2000). Multimodality. In B. Cope & M. Kalantzis (Eds.). Multiliteracies (pp. 182–202). London: Routledge.
Kress, G. (2003). Literacy in the new media age. London: Routledge.
Kress, G. (2009). What is mode? In C. Jewitt (Ed.). The Routledge handbook of multimodal analysis (pp.54-67). London: Routledge.
Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. London: Routledge.
Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). Multimodal teaching and learning. The rhetorics of the science classroom. London: Continuum.
Kress, G., & Selander, S. (2012). Multimodal design, learning and cultures of recognition. The Internet and Higher Education, 15, 265–268. https://doi.org/10.1016/j.iheduc.2011.12.003
Kress, G., & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. London: Edward Arnold.
Kress, G., & van Leeuwen, T. (2006). Reading images. The grammar of visual design (2nd Ed.). London: Routledge.
Kroonenberg, N. (1995). Developing communicative and thinking skills via e mail. TESOL Journal, 4(2), 24 27.
Kruk, J., Lubin, J., Sikka, K., Lin, X., Jurafsky, D., Divakaran, A. (2019). Integrating text and image: determining multimodal document intent in Instagram posts. Published in the Proceedings of the conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), Hong Kong, China. DOI: https://doi.org/10.18653/v1/D19-1469
Lee, H., & Mayer, R. (2018). Fostering learning from instructional video in a second language. Applied Cognitive Psychology, 32, 648-654. doi:10.1002/ACP.3436
Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. The Modern Language Journal, 102, 557–577. https://doi.org/10.1111
Lehman, S., Schraw, G., McCrudden, M., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. Contemporary Educational Psychology, 32, 569–587. https://doi.org/10.1016/j.cedpsych.2006.07.002
Lemke, J. (1990). Talking science: Language, learning and values. Norwood, NJ: Ablex.
Levie, H. W., & Lentz, R. (1982). Effects of text illustrations: A review of research. Educ. Commun. Technol. J. 30, 195–232.
Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.). The Psychology of Illustration (Vol. 1, pp.51-86). Springer, New York.
Li. S. (2011). Multimodal discourse analysis of New Horizon College English Coursebook. MA Thesis, Southwest University, China.
Liu, X., & Qu, D. (2014). Exploring the multimodality of EFL textbooks for Chinese college students: A comparative study. RELC Journal, 45(2), 135–150. DOI: 10.1177/0033688214533865

Løvland, A. (2010). På jakt etter svar og forståing. Samansette fagtekstar i skulen [Chasing for answers and understanding. Multimodal texts in school]. Bergen: Fagbokforlaget.
Lowry, M., Koneman, P., Osman-Jouchoux, R., & Wilson, B. (1994). Electronic discussion groups. TECH trends, 39(2), 22-24.
Lutkewitte, C. (2013). Multimodal composition: A critical sourcebook. Boston: Bedford/ St. Martin's.
Machin, D. (2007). Introduction to multimodal analysis. London: Hodder Arnold.
Macken-Horarik, M. (2004). Interacting with the multimodal text: Reflections on image and verbiage in Art Express. Visual Communication, 3(1), 5-26.
Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. International journal of educational research, 58, 61-68. https://doi.org/10.1016/j.ijer.2012.12.002
Marefat, F., & Marzban, S. (2014). Multimodal analysis of gender representation in ELT textbooks: Reader's perceptions. Procedia—Social and Behavioral Sciences, 98, 1093–1099.
Marsh, J. (2005). Navigational correlates of comprehension in hypertext. Paper presented at hypertext 2000, San Antonio, TX.
Marsh, E. E., & White, M. D. A (2003). Taxonomy of Relationships between Images and Text. Journal of Documentation, 59 (6), 647-672.
Martin, J. R., & Rose, D. (2012). Learning to write. Reading to learn. Genre, knowledge and pedagogy in the Sydney school. Sheffield: Equinox.
Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. Visual Communication, 4(3), 337-3.
Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? Educ. Psychol, 32, 1–19.
Mayer, R. E. (2001). Multimedia learning. Cambridge: Cambridge University Press.
Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge: Cambridge University Press.
Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. Journal of Experimental Psychology: Applied, 14(4), 329–339. https://doi.org/10.1037/a0013835
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 38, 43–52.
Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. Computers & Education, 55(2), 444–453.
Mirza, Q., Pathan, H., Khatoon, S., Hassan, A., (2021). Digital Age and Reading habits: Empirical Evidence from Pakistani Engineering University. TESOL International Journal,16 (1), 210-136.
Molitor, S., Ballstaedt, S., & Mandle, H. (Eds.). (1989). Problems in knowledge acquisition from text and pictures. Elsevier Science Publishers.
Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. System, 41, 720–739. doi:10.1016/j.system.2013.07.013
Morin, C., Mercier, A., & Atlani-Duault, L. (2019). Text–Image Relationships in Tweets: Shaping the Meanings of an Epidemic. Societies, 9 (1), 12. https://doi.org/10.3390/soc9010012

Multimodality (2023, January, 6). In Wikipedia. https://en.wikipedia.org/wiki/Multimodality#:~:text=Multimodality%20is%20the%20application%20of,%2C%20rain%2C%20etc.).

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. System, 67, 77–86. https://doi.org/10.1016/j.system.2017.04.015

Noor, N. M. (2011). Reading Habits and Preferences of EFL Post Graduates: A Case Study. Indonesian Journal of Applied Linguistics, 1(1). https://doi.org/10.17509/ijal.v1i1.95

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. Visual Communication, 10(2), 129–147.

Norris, S., & Phillips, L. (2003). How literacy in its fundamental sense is central to scientific literacy. Science Education, 87, 224–240.

Nur, R., & Syarifuddin, R. (2018). Designing an Instructional Model of Youtube-Based Materials of Listening Comprehension at Umpar Indonesia. Retrieved from https://www.asian-efl-journal.com/wp-content/uploads/AEFLJ-Volume-20-Issue-7-July-2018.pdf

Nyström, M., & Ögren, M. (2012). How illustrations influence performance and eye movement behaviour when solving problems in vector calculus [Paper presentation]. LTHs 7: e Pedagogiska Inspirationskonferens.

O'Halloran, K. 2011. Multimodal discourse analysis. In K. Hyland & B. Paltridge (Eds.). Continuum Companion to Discourse Analysis (pp. 120-137). US: Continuum International Publishing Group.

Oldakowski, T. (2014). A Multimodal Assignment that Enriches Literacy Learning: The Problem. InSight: A Journal of Scholarly Teaching, 9, 70–77. https://doi.org/10.46504/09201406ol

Omaggio, C. (1979). Pictures and second language comprehension: Do they help? Foreign Language Annals, 12, 107-116.

O'Regan, K., & Levy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.). Attention and Performance XII: The psychology of reading (pp. 329-348). Hillsdale, NJ: Erlbaum.

Owusu-Acheaw, M. (2016). Social media usage and its impact on reading habits: a study of Koforidua Polytechnic students. Int. J. Soc. Media Interact. Learn. Environ., 4, 211-222.

Paivio, A. (1986). Mental representations: A dual coding approach. Oxford, England: Oxford University Press.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension: Fostering and comprehension-monitor activities. Cognition and Instruction, 1(2), 117–175.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D. & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. Psychology and Aging, 17(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. International Journal of Applied Linguistics, 173(1), 2–17. https://doi.org/10.1075/itl.21039.pel

Pellicer-Sánchez, A., Conklin, K., Rodgers, M., & Parente, F. (2021). The Effect of Auditory Input on Multimodal Reading Comprehension: An Examination of Adult Readers' Eye Movements. The Modern Language Journal, 105(4), 936–956. https://doi.org/10.1111/modl.12743

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, A. (2020). Young learners' processing of multimodal input and its impact on reading

comprehension: An eye-tracking study. Studies in Second Language Acquisition, 42(3), 577–598. https://doi.org/10.1017/S0272263120000091
Pelowski, M., & Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. New Ideas in Psychology, 29(2), 80–97. https://doi.org/10.1016/j.nwidasych.2010.04
Pennington, M. (1989). Teaching languages with computers: The state of the art. La Jolla, CA: Athelstan.
Pieters, R., Rosbergen, E., & Hartog, M. (1996). Visual Attention to Advertising: The impact of motivation and repetition. Advances in Consumer Research, 23, 242–248.
Pillai, A. M. (2019). Image vs. text in narratives: The case study of National Geographic's Instagram [Dissertation submitted in part fulfilment of the requirements for the degree of MBA - General] At Dublin Business School.
Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. Journal of Educational Psychology, 90, 25–36.
Platt, B. (1986). Reading preferences of adolescents. Illinois Reading Council Journal, 14(1), 37–39.
Prensky, M. (2001). Digital natives, digital immigrants, part 1. On the Horizon, 9(5), 1-6. DOI: 10.1108/10748120110424816
Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. Second Language Research. doi:10.1177/02676583211049741
Qenaway, Kh. (2019). The impacts of using social media on second language reading habit among Arab students. Academia. Retrieved from https://www.academia.edu/40216569/the_impacts_of_using_social_media_on_second_language_reading_habit_among_arab_students.
Rafiq, M., Khan, T. M., Asim, A., & Arif, M. (2019). The effects of social media on reading habits. Pakistan Journal of Information Management & Libraries, 21, 46–65.
Rajab, H., & Al-Sadi, A. (2015). An empirical study of reading habits and interest of Saudi University EFL learners. International Journal of Linguistics 7(2), 1–17. https://doi.org/10.5296/ijl.v7i2.7034
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372– 422.
Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology, 62, 1457–1506. https://doi.org/10.1080/17470210902816461
Rayner, K., & Castelhano, M. S. (2008). Eye movements during reading, scene perception, visual search, and while looking at print advertisements. In M. Wedel & R. Pieters (Eds.). Visual marketing: From attention to action (pp. 9-42). Taylor & Francis Group/Lawrence Erlbaum Associates.
Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. Scientific Studies of Reading, 10(3), 241–255. https://doi.org/10.1207/s1532799xssr1003_3
Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye movements when looking at print advertisements: The goal of the viewer matters. Applied Cognitive Psychology, 22, 697–707. https://doi.org/10.1002/acp.1389
Rayner, K., & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., Rotello, C. M., Stewart, A., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. Journal of Experimental Psychology: Applied, 7, 219–226

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. Psychonomic bulletin & review, 17(6), 834–839. https://doi.org/10.3758/PBR.17.6.834

Richards, J.C. (1999). Interchange. English for international communication. Student's book. Cambridge: Cambridge University Press.

Richards, J.C., Hull, J., Proctor, S. (2001). New interchange. English for international communication. Student's book (19th ed). Cambridge: Cambridge University Press.

Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. Calico Journal, 34, 20–38. doi:10.1558/cj.29522

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. ITL-International Journal of Applied Linguistics, 171(2), 191-220. doi:10.1075/itl.18034.rod

Royce, T. (2002). Multimodality in the TESOL Classroom: Exploring Visual-Verbal Synergy. TESOL Quarterly, 36(2), 191–205.

Roy-Charland, A., Saint-Aubin, J., & Evans, M. A. (2007). Eye movements in shared book reading with children from kindergarten to Grade 4. Journal of Reading and Writing, 20, 909–931. https://doi.org/10.1007/s11145-007-9059-9

Salbego, N., Heberle, V. M., & Soares da Silva Balen, M. G. (2015). A visual analysis of English textbooks: Multimodal scaffolding learning. Calidoscopio, 13(1), 5–13.

Sarland, C. (1991). Young People Reading: Culture and response. Milton Keynes: Open University Press.

Scales, A. M. & Rhee, O. (2001). Adult reading habits and patterns. Reading Psychology, 22(3), 175-203. https://doi.org/10.1080/027027101753170610

Schaffhauser, D. (2009). Boundless opportunity. T.H.E. Journal, 36(9), 13–18.

Schleppegrell, M. (2004). The language of schooling: A functional linguistics perspective. Mahwah, NJ: Lawrence Erlbaum.

Schnotz, W. (1994). Aufbau vonWissenstrukturen. Untersuchungen zur Koh¨arenzbildung beim Wissenserwerb mit Texten. Pschologie Verlags Union: Weinheim.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. Educational Psychology Review, 14(1), 101–120.

Schnotz, W., & Baadte, C. (2008). Domain learning versus language learning with multimedia. In M. Farías & K. Obilinovic (Eds.). Aprendizaje multimodal/Multimodal learning (pp. 21-49). Santiago de Chile: Publifahu USACH.

Schnotz, W., & Bannert, M. (1999). Support and interference effects in learning from multiple representations. In S. Bagnara (Ed.). European Conference on Cognitive Science (pp. 447– 452), Istituto di Psicologia Consiglio, Nazionale delle Ricerche, Rome, Italy.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. Applied Cognitive Psychology, 23, 1227–1243. doi: 10.1002/acp.1526

Seburn, T (2017). Learner-sourced visuals for deeper text engagement and conceptual comprehension. In K. Donaghy & D. Xerri (Eds.). The image in English language teaching (pp. 79-88). Malta: Gutenberg Press.

Seel, N. M. (1986). Wissenserwerb durch Medien und "mentale Modelle".Unterrichtswissenschaft, 14(4), 384–401.

Selander, S., & Kress, G. (2010). Design för lärande – ett multimodalt perspektiv [Desings for learning – a multimodal perspective]. Stockholm: Norstedts.
Selfe, C. (1990). Technology in the English classroom: Computers through the lens of feminist theory. In C. Handa (Ed.). Computers and community: Teaching composition in the twenty-first century (pp. 118-139). Portsmouth, NH: Heinemann.
Serafini, F. (2012). Expanding the four resources model: Reading visual and multi-modal texts. Pedagogies: An International Journal, 7(2), 150-164.
Serrano, R., & Pellicer-Sánchez, A. (2022). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye movements. Applied Linguistics Review. https://doi.org/10.1515/applirev-2018-0102
Shen, L. (2006). Computer Technology and College Students' Reading Habits. Chia-Nan Annual Bulletin, 32, 559-572.
Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior, 6, 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7
Siegel, M. (2012). New times for multimodality? Confronting the accountability culture. Journal of Adolescent & Adult Literacy, 55(8), 671-680.
Smagorinsky, P. (2001). If meaning is constructed, what is it made from? Toward a cultural theory of reading. Review of Educational Research, 71(1), 133–169.
Solan, H. A., Feldman, J., & Tujak, L. (1995). Developing visual and reading efficiency in older adults. Optometry and Vision Science, 72, 139–145. https://doi.org/10.1097/00006324-199502000-00012
Spivey, N.N. (1997). The constructivist metaphor: Reading, writing, and the making of meaning. San Diego, CA: Academic Press.
Ståhl, M., & Kaihovirta, H. (2019). Exploring visual communication and competencies through interaction with images in social media. Learn Cult Soc Interact, 21, 250-266.
Sturgeon, C. M., & Walker, C. (2009). Faculty on Facebook: Confirm or Deny? Retrieved from https://eric.ed.gov/?id=ED504605
Svärdemo Åberg, E., & Åkerfeldt, A. (2017). Design and recognition of multimodal texts: Selection of digital tools and modes on the basis of social and material premises?. Journal of Computers in Education. 4, 283–306. https://doi.org/10.1007/s40692-017-0088-3
Tan, Y. (2012). A multimodal discourse analysis of the senior high school English textbooks Advance with English. MA Thesis, Hunan University of Science and Technology, China.
Tang, G. (1992). The effect of graphic representation of knowledge structures on ESL reading comprehension. Studies in Second Language Acquisition, 14, 177. https://doi.org/10.1017/S0272263100010810
Tang, K-S., Delgado, C., & Moje, E. B. (2014). An integrative framework for the analysis of multiple and multimodal representations for meaning-making in science education. Science Education, 98(2), 305–326.
Tinio, P. P. (2013). From artistic creation to aesthetic reception: The mirror model of art. Psychology of Aesthetics, Creativity, and the Arts, 7(3), 265–275. https://doi.org/10.1037/a0030872
Tragant, E., & Pellicer-Sánchez. A. (2019). Young learners' engagement with multimodal exposure: An eye-tracking study. System, 80, 212–223. https://doi.org/10.1016/j.system.2018.12.002

Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. English Teaching: Practice and Critique, 5(1), 55-76.
Unsworth, L. (2008). Multiliteracies and metalanguage: Describing image/text relations as a resource for negotiating multimodal texts. In J. Coiro. et al. (Eds.). Handbook of Research on New Literacies (pp. 377- 405). New York: Lawrence Earlbaum.
Unsworth, L., & Chan, E. (2009). Bridging multimodal literacies and national assessment programs in literacy. Australian Journal of Language and Literacy, 32(3), 245–257.
Van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.
Varaporn, S., & Sitthitikul, P. (2019). Effects of multimodal tasks on students' critical reading ability and perceptions. Reading in a foreign language, 31, 81–108.
Verdi M. P., & Kulhavy R. W. (2002). Learning with maps and texts: An overview. educational psychology review, 14(1), 27–46.
Vergara, A., Siles, I., Castro, A, C., & Chaves, A. (2021). The Mechanisms of "Incidental News Consumption": an Eye Tracking Study of News Interaction on Facebook. Digital Journalism, 9(2), 215–234, DOI: 10.1080/21670811.2020.1813047
Vitu, F. M., & McConkie, G.W. (2000). Regressive saccades and word perception in adult perception. In A. Kennedy, R. Radach, & J. Pynte (Eds.). Reading as a perceptual process (pp. 89–118). Amsterdam, the Netherlands: Elsevier. https://doi.org/10.1016/B978-008043642-5/50015-2
Walker, A. (2010). Using social networks and ICTs to enhance literature circles: A practical approach. Paper Presented at the School Library Association of Queensland and the International Association of School Librarianship Conference. Brisbane, Australia.
Wang, H. C., & Pomplun, M. (2012). The attraction of visual attention to texts in real-world scenes. Journal of Vision, 12(6), 1–17. https://doi.org/10.1167/12.6.26
Wang, S., & Vasquez, C. (2014). The effect of target language use in social media on intermediate-level Chinese language learners' writing performance. CALICO Journal, 31(1), 78-102.
Wellington, J., & Osborne, J. (2001). Language and literacy in science education. Philadelphia: Open University Press.
Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. Language Learning & Technology, 14, 65–86. doi:10125/44214.
Wolfe, P. (2018). Using eye movement to study adolescents' comprehension of visual texts. In A. August (Ed.). Visual imagery, metadata, and multimodal literacies across the curriculum (pp. 235–248). Hershey PA, USA: IGI Global. https://doi.org/10.4018/978-1-5225-2808-1.ch013
Wright, A. (1989). Pictures for language learning (1st ed.). Cambridge University Press. Cambridge.
Wright, A. (2010). Pictures for language learning (Illustrated edition). Cambridge: Cambridge University Press.
Wyer, R. S, Jr., Hung, I. W., & Jiang, Y. (2008). Visual and verbal processing strategies in comprehension and judgment. Journal of Consumer Psychology, 18(4), 244–257.

APPENDICES

Appendix A. Reading habit questionnaire

Kedves Résztvevő!

A kérdőív a közösségi média használatával kapcsolatos szokásaival kapcsolatos kérdéseket tesz fel. Kérem, hogy jelölje a legmegfelelőbb választ.

Kérem, hogy ne felejtse el használni a 'submit' gombot, mielőtt befejezné a feladatsort!

	SOHA	NÉHA	GYAK-RAN	REND-SZE-RESEN
1: Nyomtatott újságot olvasok.				
2: Nyomtatott magazint olvasok.				
3: Nyomtatott könyvet olvasok.				
4: Nyomtatott képregényt olvasok.				
5: Elektronikus újságot olvasok.				
6: Elektronikus magazint olvasok.				
7: Elektronikus könyvet olvasok.				
8: Elektronikus képregényt olvasok.				
9: Használom a közösségi médiát.				
10: Kapcsolatokat tartok a közösségi médián.				
11: Szövegeket olvasok a közösségi médián.				
12: Képeket nézegetek a közösségi médián.				
13: Videokat nézek a közösségi médián.				

Appendix B. The first draft of the multimodal reading test

1- Image-text independent

Noisy neighbours

The top problems!

- The baby cries.
- The dog barks.
- They talk loudly.
- They argue a lot.
- They have noisy parties.
- They have the TV on very loud.
- They play musical instruments.

1- **What does the neighbor play loud?** Musical instrument ☐ The guitar ☐

2- Image-text complementary

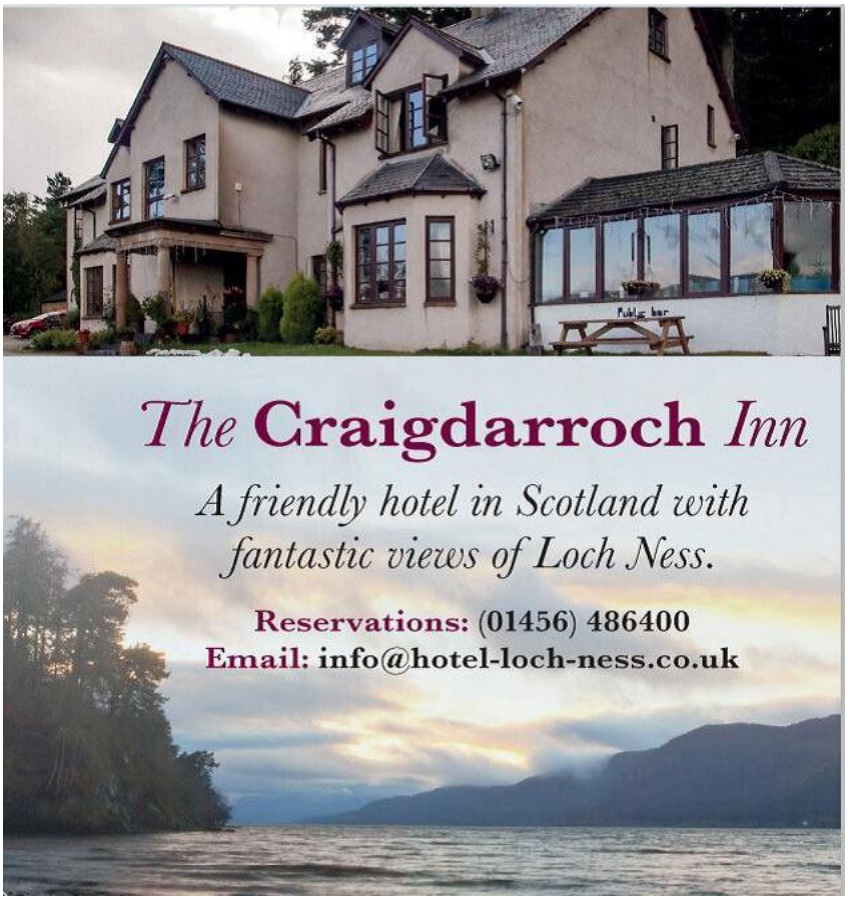
1- How do the people feel when they gossip?

Good ☐ Happy ☐

3- Text more general than image

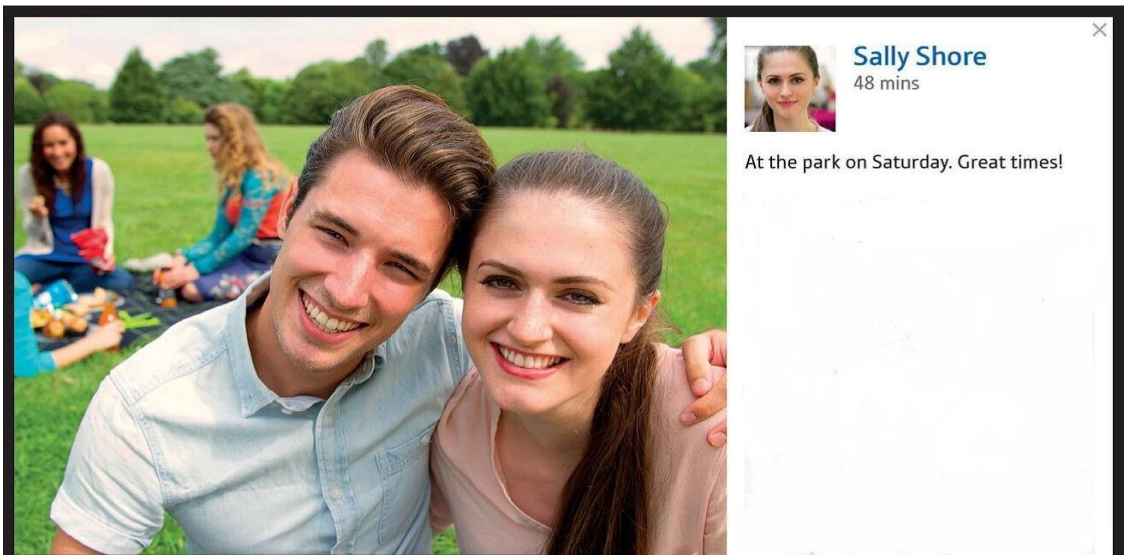
1- **What have you been afraid of?** A snake ☐ it ☐

4- Image more general than text:



1- **What is the view of the hotel?** the lake ☐ Loch Ness ☐

5- Enhancement, temporal

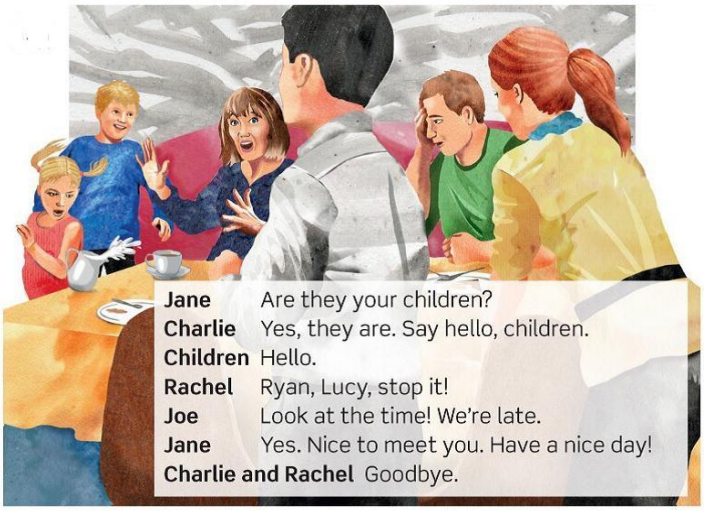


1- **When are they at the park?** On Saturday. ☐ When the weather is good. ☐

6- Enhancement, spatial

1- Where are they? Texas ☐ Restaurant ☐

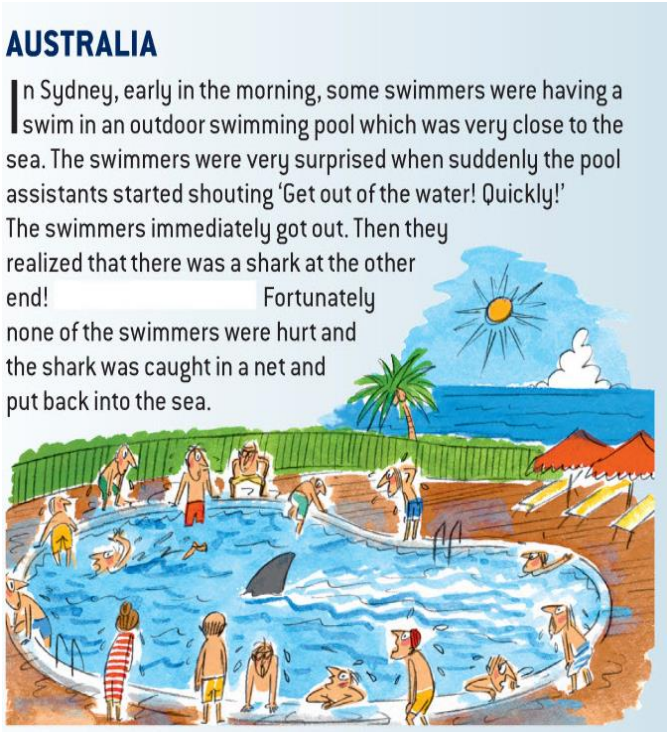
7- Enhancement, reason



1- **Why is Rachel angry?** Because the kids do not stop it. ☐

Because the kids are playing around the table. ☐

8- Image subordinate to the text



1- Where was the shark? In the pool ☐ At the other end. ☐

9- Text subordinate to the image

Helena Bonham Carter, the actress, was born in London in 1966. Her mother is half Spanish and her father, who died in 2004, was the grandson of Herbert Asquith, the British Prime Minister from 1908 to 1916. Her first big role was as Lucy Honeychurch in *A Room with a View*, and her other roles include The Red Queen in *Alice in Wonderland*, Bellatrix Lestrange in the Harry Potter films, and Queen Elizabeth in *The King's Speech*. She has two children with her partner Tim Burton, a film director.

1- **What is she?** She is a writer. ☐ She is an actress. ☐

Appendix C. Multimodal reading test

Kedves Résztvevő!

A feladatsor 20 részfeladatot tartalmaz, amelyek részben képi, részben szöveges információt közölnek. Mindegyik részt követően kérdéseket talál, amelyekhez két megadott válasz közül választhat (mindkettő helyes megoldást tartalmaz). Nincs jó vagy rossz válasz! Kérem, hogy nézze meg a szövegeket és a képeket, és azt a választ jelölje meg, amelyet elsőként választana.

Image-text independent 1 (the ITI-Rain)

Water falls from the clouds as rain.
The rain runs into rivers.
The rain runs from the rivers into the sea.

1. Where does the rain go?

a. It goes to the rivers.
b. It goes to the sea.

Image-text independent 2 (the ITI-Routine)

Jack gets up.
He has breakfast**.**
He takes a shower.

2. What does Jack do when he is awake?

a. He has breakfast.

b. He takes a shower.

Image-text-complementary 1 (the ITC-Earth)

The earth is melting.
We can save our planet.

3. What problem does the earth have?

a. The earth is melting.

b. The earth is getting hot.

Image-text-complementary 2 (the ITC-Jungles)

Jungles are lungs of the earth.
Jungles are very important.

4. Why should we protect jungles?

a. Because they are lungs of the earth.

b. Because they make oxygen for the earth.

Text-more-general-than-image 1 (the TGI-Volleyball)

They like to do sports.
They do sports in the park.

5. How do they usually spend their free time?

a. They usually play sports.

b. They usually play volleyball.

Text-more-general-than-image 2 (the TGI-Violin)

Rosie is 8 years old.
She plays musical instruments.

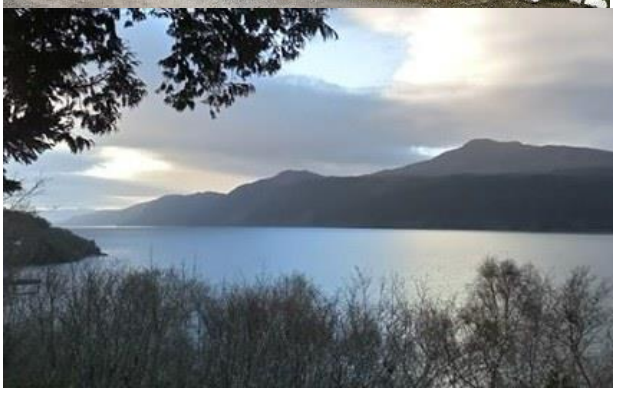
6. What does she do in her free time?

a. She plays musical instruments.

b. She plays the violin.

Image-more-general-than-text 1 (the IGT-Lake)

The George Inn is a friendly hotel.
It has a fantastic view of Loch Ness.



7. What can you see from the window of the hotel?

a. Loch Ness

b. The lake

Image-more-general-than-text 2 (the IGT-Fruits)

Emma is strong. She likes bananas.

8. Why is Emma healthy?

a. She eats bananas.

b. She eats fruits.

Image-subordinate-to-text 1 (the IST-Granny)

David's grandmother had grey hair.
She was always happy.
She helped the family a lot.

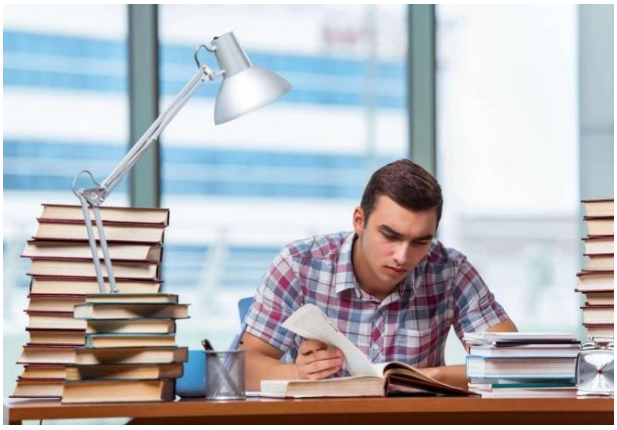
9. Why was David' grandmother a nice person?

a. Because she helped the family a lot.

b. Because she was always happy.

Image-subordinate-to-text 2 (the IST-Student)

William was a hardworking student.
He was also very kind.
He helped friends a lot.



10. What made William a good boy?

a. **He helped his friends a lot.**
b. **He was a hard-working student.**

Text-subordinate-to-image 1 (the IST-Vegetables)

Look at Oliver.
He is sad.
He does not like them.

11. Why is Oliver unhappy?

a. He does not want to eat them.

b. He does not want to eat vegetables.

Text-subordinate-to-image 2 (the TSI-Coffee)

James is working on his laptop.
There is a cup of coffee on the desk.
He is putting it on the desk.

12. What is James putting on the desk?

a. He is putting a coffee on the desk.

b. He is putting a car key on the desk.

Image-enhance-text-place 1 (the IEP-Airport)

John was too late.
He missed his flight to Paris.

13. What did he have to do on time?

a. He had to get the flight on time.

b. He had to get to the airport on time.

Image-enhance-text-place (the IEP-Hotel)

The family is in London for a trip.
They are going to stay here for a week.

14. Where are they going to stay for a week?

a. They are going to stay in London

b. They are going to stay in a hotel.

Image-enhance-text-reason 1 (the IER-Football)

John played football at home.
Mary was angry.
John was sad.

15. Why was Mary angry with John?

a. Because he played football at home.

b. Because he broke the bottle.

Image-enhance-text-reason 2 (the IER-Birthday)

It is Liam's birthday party.
He is 6 years old now.
He is very excited.

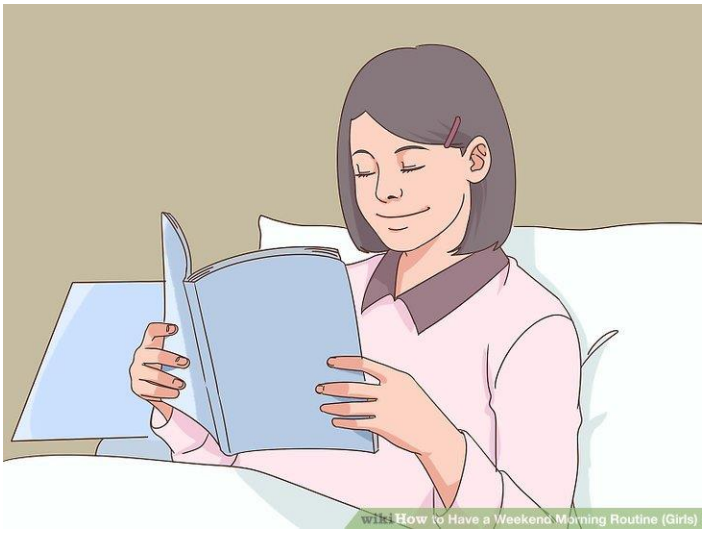
16. Why is he very happy?

a. Because it is his birthday.

b. Because he gets a beautiful gift.

Text-enhance-image-time (the TET-Reading)

Anne always reads a book in bed at night. She sleeps around midnight.



17. When does she sleep?

a. She sleeps around midnight.

b. She sleeps after reading a book.

Text-enhance-image-time 2 (the TET-Picnic)

They are having a picnic in the park on Saturday.
It is sunny.
They are having a great time.

18. When do they usually have a picnic?

a. They usually have a picnic on Saturdays.

b. They usually have a picnic when it is sunny.

Image-text-tension 1 (the ITT-Shopping)

Mary is in a clothes shop.
She is buying a dress.
She likes the yellow dress.

19. What color dress does she want to buy?

a. She wants to buy the yellow dress.

b. She wants to buy the red dress.

Image-text-tension 2 (the ITT-Address)

Go straight on North Street. Then turn left at Hillside Road. The chemist is next to the cinema.

20. What direction should you go after North Street?

a. You should take a left turn.

b. You should take a right turn.

Appendix D. Follow-up questions

21. Segítették-e a képek a szövegek elolvasását és jobb megértését?

a. Igen
b. Nem

22. A képekre vagy szövegekre támaszkodott leginkább a kérdések megválaszolásakor?

a. Képekre

b. Szövegre

c. Mindkettőre

Appendix E. Consent form for the online experiment

Consent form for the participation in the research

You are requested to participate in this research conducted by Marzban Sheida, the PhD student of Multilingualism at the University of Pannonia. This study aims to investigate how English language learners in Hungary read multimodal texts . This experiment is part of the completion of the PhD research at the University of Pannonia. In this study, you are asked to fill in some tests.

The results of this research will be published in international conferences or journals. However, your identity will be confidential and protected. Participation in this study is voluntary. You have the right to choose not to participate or withdraw during the process of research.

If you have questions about your participation in the study or you would like to receive a report of results when the study has been completed, please contact:

Name: Marzban Sheida

Affiliation: Doctoral School of Multilingualism, University of Pannonia, Hungary

Telephone number: +36205049105

Supervisor: Dr. Gyöngyi Fábián

Thank you for your support and help.

◯ Please check the box and click on 'Submit' button if you agree to the terms.

Appendix F. Consent form for the eye-tracking experiment

Consent form for the participation in the research

You are requested to participate in this research conducted by Marzban Sheida, the PhD student of Multilingualism at the University of Pannonia. This study aims to investigate how English language learners in Hungary read multimodal texts of different image-text relations. This experiment is part of the completion of the PhD research at the University of Pannonia. In this study, you are asked to fill in some tests and do the eye-tracking experiment. Your eye-movements will be recorded by an eye-tracker. The eye-tracking experiment will take around 20 minutes.

The results of this research will be published in international conferences or journals. However, your identity will be confidential and protected. Participation in this study is voluntary. You have the right to choose not to participate or withdraw during the process of research.

If you have questions about your participation in the study or you would like to receive a report of results when the study has been completed, please contact:

Name: Marzban Sheida
Affiliation: Doctoral School of Multilingualism, University of Pannonia, Hungary
Telephone number: +36205049105
Supervisor: Dr. Gyöngyi Fábián

Signature: