Education and Information Technologies (2024) 29:4699–4743
https://doi.org/10.1007/s10639-023-12017-8

How do L2 learners of Turkish read texts with graphics? An eye-tracking study

Mustafa Armut[1] · **Mehmet Kara**[2]

Received: 2 January 2023 / Accepted: 28 June 2023 / Published online: 11 July 2023

Abstract

This study aimed to determine the variables affecting the eye-movements of second-language (L2) learners of Turkish in Türkiye while solving a test consisting of questions that included graphics. A total of 115 L2 learners participated in the research (21 in the pilot study and 94 in the main study). We presented a test consisting of five multiple-choice questions with graphics (GT) as a stimulus on the screen and investigated the following variables to see whether participants' eye-movements were affected by age, gender, native country, mother tongue, reading direction in the mother tongue, department at the university, education level, opting to have questions with graphics in the Turkish Learning, Research and Application Center (TÖMER) and the Language Teaching Research and Application Center (DİLMER) exams and success on the GT. We investigated whether these independent variables had an effect on the eye-movement measurements and examined the relation between the GT success and success on the TÖMER/DİLMER comprehension exam. Results indicate that gender, education level, opting to have graphic questions in the TÖMER/DİLMER exams, and GT success had significant effects on various eye-movements.

Keywords Eye-tracking · Graphics · L2 learners · Learning Turkish · Test

This study was based on the data of the doctoral thesis titled "Investigation of Learners' Reading Skills in Graphic Questions with Eye Tracking Technique in Teaching Turkish to Foreigners" conducted by the first researcher under the supervision of the second researcher.

The data of this study were obtained at Gazi University Turkish Learning, Research and Application Center in Ankara-Turkey.

Mustafa Armut is responsible for all stages of refereeing and publication, as well as post-publication correspondence.

Extended author information available on the last page of the article

1 Introduction

Turkish is a language that is increasingly in demand, especially by individuals from the Middle East, Africa, and Central Asian countries. Many of these people come to Türkiye for undergraduate and graduate studies. Before starting their education, they learn Turkish for about a year with preparatory courses. Language teaching centers such as TÖMER and DİLMER established within universities do most of the teaching of Turkish to L2 learners.

Preparing assessment-evaluation materials suitable for various language teaching levels is an important need for Turkish-teaching centers, and centers should know L2 learners' characteristics to prepare accurate materials based on learners' levels. However, no evidence exists about the effect of variables such as gender, mother tongue, and reading direction in the mother tongue on the test achievement of L2 learners of Turkish. In addition, there are no studies on the cognitive processes of L2 learners of Turkish during test situations. Eye-movements can be very informative about these cognitive processes. As a matter of fact, there is a relation between eye-movements and cognitive processes (Just & Carpenter, 1980) and the eyes can be described as the windows to the mind (Duffy, 1992). In this respect, the eye-tracking technique, which provides important clues about cognitive processes, can provide important information about the cognitive processes of L2 learners of Turkish while taking the GT.

The Common European Framework of Reference for Languages (CEFR; Council of Europe [CE], 2020), which is accepted as the main reference source in second/foreign language teaching in Europe, places an important emphasis on graphics. There are important explanations and descriptors in the CEFR for the use of graphics at different levels in language teaching. However, reading and comprehending graphics, multiple texts accompanied by graphics, and questions with graphics do not have an important place in language teaching centers in Türkiye. Every language has its own visual literacy (Bruski, 2011). In this respect, graphics accompanied by Turkish texts also form their own grammar and vocabulary as a visual literacy category. Therefore, it would be meaningful to investigate the cognitive processes of L2 learners of Turkish while taking a test consisting of questions with both text and graphics. Eye-tracking is a very suitable technique for investigating such processes.

1.1 Eye-tracking

Eye-tracking is a technique based on the online measurement of eye-movements with the help of infrared light. In this system, infrared light reflected on the cornea follows eye-movements on a screen or with the help of a small camera in natural environments. Movements are recorded with the help of a special software that has an eye-tracking algorithm (Dolgunsöz, 2015). In this process, several near-infrared light sources create reflection patterns on the cornea that are invisible to the eyes of the viewer (Al-Edaily et al., 2013).

The eye-tracking technique can help researchers understand visual attention. With eye-tracking, it is possible to learn where users have looked at one point in time and one can get an idea of how long something has been looked at. The path followed by users' eyes can be investigated (Schall & Romano Bergstrom, 2014).

The eye-tracking technique is a relatively new technology. It is increasingly in demand in applied linguistics, second language acquisition, and in investigating tasks that are traditionally measured using offline tecniques (e.g., paper–pencil tests, assessment tasks, think-aloud protocol, and interviews). Eye-tracking is also commonly used to monitor eye-movement behaviors while participants are looking at still images, viewing scenes, or watching videos (Conklin & Pellicer-Sánchez, 2016). Advantages and applications of the eye-tracking technique include access to pre-attentive behavior, understanding sequential strategies, and error analysis (Goldberg & Helfman, 2011).

According to Duffy (1992), eye-movement recording is currently the best way to get information about moment-to-moment operations during reading comprehension. In this context, eye-tracking is widely used to obtain answers to some questions during reading. Many researchers interested in the language process use eye-tracking technology to view the eyes while reading (Conklin & Pellicer-Sánchez, 2016). Eye-movement, which is considered to be closely related to cognitive processes, represents the attention of the user. Therefore, the aforementioned processes can be observed via the eye-tracking technique (Biedert et al., 2010).

The literature on the eye-tracking indicates that a large number of eye-movement measures are available. However, Duchowski (2007) reported that the primary aim in eye-movement analysis is to identify fixations, saccades, and straight pursuits. Rayner et al. (2006) stated that there are three basic dimensions of eye-movement during reading: saccades, fixations, and regressions. Winke et al. (2013) also explained that eye-movement behavior is divided into two distinct categories: fixations and saccades.

Fixations refer to instances when the eye-movements stop scanning and focus on a particular part of a visual field (Schall & Romano Bergstrom, 2014). Fixations are eye-movements that focus the retina on an immobile object of interest. These movements are actually described as miniature eye-movements such as *tremors, drifts,* and *micro-saccades* (Duchowski, 2007). During reading, a fixation takes about 200–250 ms (ms; Rayner et al., 2006).

Rayner et al. (2006) stated that while eyes are generally thought to move straight across the page, in reality the eyes make a number of rapid movements separated by pauses called fixations that typically last about 200–250 ms. They added that the new information is decoded only during fixations, citing that vision is suppressed during saccades. For skilled readers, the eyes typically advance seven to ten letter gaps per saccade.

Saccades are the rapid movements of the eye from one fixation to the next to help unite the whole scene one is looking at (Schall & Romano Bergstrom, 2014). These saccades are rapid eye-movements used to reposition the fovea to a new span in the visual environment (Duchowski, 2007). Saccades occur when a person moves their eyes from one area to another, providing new visual information to the cognitive system (Winke et al., 2013).

Rayner (1998) stated that while reading, looking at a scene, or scanning an object, continuous eye-movements called saccades occur. He added that between these jumps, the eyes remain relatively still for about 200–300 ms during fixation. Saccadic movements are both systematic and reflexive. These movements can be performed voluntarily or run as a confirmatory optokinetic or vestibular measure (Duchowski, 2007).

Rayner (1998) reported that the average saccade size when reading English is seven to nine letter gaps. Reichle et al. (1998), on the other hand, stated that the duration of most fixations in reading is between 200 and 350 ms, and most saccades (from left to right when reading English) are between five and nine characters. They indicated that the saccades take about 15 to 40 ms.

There are functional relations between the fixation duration and the properties of the fixed words. Readers spend more time fixating on more difficult and important words than focusing on easier words, and they are more likely to focus on long words than short words. Words that are likely to be skipped are function words that are shorter (Dussias, 2010). Dehaene (2017) stated that in reading, fixation must occur at least once on almost all content words such as nouns or verbs, although small grammatical words such as *this* and *that* can sometimes be omitted. There is also a common misconception that readers are not fixated on every word, but on only a small part of the text, perhaps every two or three words (Just & Carpenter, 1980).

Regressions are eye-movements in which the reader refixes to the previously fixed information (Duffy, 1992). Reichle et al. (1998) emphasized that not all saccades go from left to right. They reported that there are return movements that take the reader from the end of a line to the beginning of the text. They also stated that the reader moves backwards in the text about 10% of the reading time. Skilled readers regress to the previously read part 10 to 15% of the time they read (Rayner, 1998; Rayner et al., 2006). Rayner (2009) stated that regressions are not particularly well understood because it is difficult to control them experimentally. Because it is difficult to deduce empirically from regressions, examining them is more difficult than examining forward saccades (Rayner, 1998).

If the return is low during reading, the comprehension level is high. This hypothesis is based on speed reading, also known as native speaker reading (Hayashi et al., 2002). Longer regressions in a text largely reflect failure to understand (Conklin & Pellicer-Sánchez, 2016). Also, regressions must be distinguished from return sweeps. Return sweep refers to saccades from right to left from the end of the line to the beginning of the next line (Rayner, 2009).

Another term in eye-tracking studies is *areas of interest*. Areas of interest are areas that can be used to detect fixations on the stimulus image and to identify scan sequences and transitions (Goldberg & Helfman, 2010; Holmqvist et al., 2011). Simply put, areas of interest define the regions in the stimulus from which the researcher aims to collect data (Holmqvist et al., 2011). Likewise, based on the objectives of the research, it may be important to obtain information about where the participants first looked on the page, how long they looked there, and whether they focused on the content or quickly moved to another area of the screen (Olmsted-Hawala et al., 2014). Area of interest is a region of stimulus relevant to one of the study's research

questions. This area can be a navigation bar, software widget, a paragraph of text, a product on a shelf, an information board, or a wayfinding sign at an airport. Customizing the areas of interest is necessary for quantitative analysis. Areas of interest are needed to calculate most eye-tracking measurements (Bojko, 2013).

Another related subject is the devices used for tracking the eyes. Eye-movement-tracking devices generally collect information about the location and duration of eye fixation in a particular area. When objects such as words and pictures are shown, a person's eye-movement can be observed while they are looking at those words or pictures (Tai et al., 2006). This technique usually works by illuminating the eye with infrared light from one or more light sources, and reflections from the eye are recorded with the camera. One of the advantages of infrared lights is that they are invisible to the user. Thus, users are not disturbed while working with the eye-tracking device (Biedert et al., 2010).

1.2 Graphics

Graphics began to be used extensively by William Playfair about 200 years ago to show data (Cleveland & McGill, 1984). Playfair developed almost all of his basic graphic designs, wishing to replace traditional number tables with systematic visual representations of his linear arithmetic (Tufte, 2006).

Acartürk (2015) stated that graphics are communication tools that are frequently used in daily life. Graphics are used in conjunction with texts in scientific journals as well as written and electronic media in daily life (Habel & Acartürk, 2011). Documents containing figures (e.g., graphics, drawings, and photographs), tables, and equations along with text are common in both print and electronic media (Habel & Acartürk, 2007). Graphics can be important to students to turn complex data into an easy-to-understand visual (Stoerger, 2018). According to Schiff (1998), one of the best ways to make information more useful is to use graphical representations. One of the striking facts about human consciousness is that it prefers to process quantitative information graphically (Pinker, 1990). Therefore, graphic representations can be one of the most appropriate tools that can help present quantitative information in a more concise and clear way.

Graphics are diagrams that display data (Reed, 2010). They are technical elements within the scope of visual literacy together with advertisements, posters, brochures, digital signs, and tables (Maden & Altunbay, 2016). According to Tufte (2006), data graphics visually display measured quantities through the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color.

Tufte (2006) stated that the simplest and most powerful of the methods for analyzing and communicating statistical information is well-designed data graphics. Spence (2005) cited that when Playfair presented statistical data for nineteenth century European countries, he did so in graphic form, because he believed that catching the eye is the best way to convey a particular idea when it comes to proportion and size.

The scope of graphics and what constitutes graphics are issues to consider when working with graphics. Plass and Jones (2005) considered graphics, together with

diagrams, photographs, and maps, in the category of picture that has the property of non-dynamic material. Acartürk (2015) wrote that statistical graphics, also called information graphics, are descriptive representations that reflect data. Simkin and Hastie (1987) described bar, pie, and line graphs as graphics.

It is also possible to consider what qualifications a competent individual should have in reading graphics. Ainley et al. (2000) argued that expertise in graphics requires three distinct competencies: (a) an understanding of how to interpret and use graphics; (b) knowledge of the layout and technique of graphics, such as using scale; and (c) the practical skills required to produce graphics manually. Strobel et al. (2016), on the other hand, stated that identifying trends and comparing score differences are regular graphic reading tasks that are mostly necessary to understand and interpret data sets. Pinker (1990) stated that graphic readers must do two things: First, they must accurately represent the objects on the graph. Second, they must remember or understand which dimensions of the visual components of the graphic take the place of which of the mathematical scales the graph is trying to link.

One issue with graphics is their design. Tufte (2006) revealed the following principle regarding graphic design: First of all, data should be displayed. He added that this principle is also the basis of a theory for data graphs. In this context, Tufte explained the *Data-Ink* theory that he put forward. Accordingly, most of the ink in a graph should present data. As the data change, the ink changes. Data ink is the indelible core of a graphic. After discussing this theory through various examples within the scope of Data-Ink theory, Tufte listed the following principles on the subject: (a) first of all, show the data; (b) increase the data ink rate; (c) delete ink that is not data related; and (d) review and edit.

1.3 Multiple-choice questions

Today, the multiple-choice question format is still widely used in educational circles, especially on standardized exams. According to Boylu (2019), the multiple-choice question is the most preferred question type in teaching Turkish as L2, especially at the B2 and C1 levels. This question type is also frequently preferred in centers that teach Turkish as L2. Despite the widespread use of multiple-choice questions, we need more work to develop, implement, and standardize multiple-choice-type questions, especially in the context of teaching Turkish as L2. The study conducted by Işıkoğlu (2015) on teaching Turkish as L2 criticized the questions on proficiency exams prepared by two institutions. Gedik (2017) also focused on multiple-choice questions in regards to measurement and evaluation in teaching Turkish as L2. The thesis study by Boylu (2019) on setting standards with measurement evaluation in teaching Turkish as L2 allocated an important part to multiple-choice-type questions.

According to Özçelik (2010), a multiple-choice question is a type of question in which a question is asked and the answer to the question is given with three or four statements that are not the correct answer. In such questions, what is expected from the student is to read the question posed, to think about the answer, and to mark the correct answer found among the given ones. Multiple-choice questions require

students to choose the correct response from a list of options. There are variations such as choosing a single-most probable answer or multiple probability options. The most well-known examples are based on choosing a single correct answer from two, three, or four options (McKenna, 2019).

It is possible to classify multiple-choice questions within their type. One of these question classes is the context-based item. While Haladyna (2004) discussed the context-based item, the stimulus for such a question is a picture, photograph, chart, graphic, figure, table, written text, poem, story, problem, or narrative. In this respect, it is natural for graphics to be included in context-based questions.

Haladyna et al. (2002) emphasized that the traditional multiple-choice question, which is one of the multiple-choice item formats, is well known and is widely used in achievement tests at all levels and with many content types. The authors stated that all books recommend this format, and that all test publishers use this traditional form for their standardized tests. There are many reasons why multiple-choice questions are in high demand. For example, since they can be marked automatically and objectively in principle, multiple-choice questions save time, self-control, and feedback in terms of marking (McKenna, 2019).

Haladyna et al. (2002) stated that the science of writing multiple-choice items is advancing, but writing items is still a creative activity. For this reason, we can expect an eye-tracking study to make a significant contribution to the field of preparing multiple-choice questions.

1.4 Significance of the study

When it comes to language and reading research, many people focus only on written texts. However, visuals accompany many written texts in real-life situations. Moreover, according to Reed (2010), most learning is based on combining pictures and words.

Tepgeç and Seferoğlu (2019) stated that multimedia learning is one of the areas in which eye-tracking measurements are widely used in the context of education. However, these studies are insufficient when compared with reading studies on this subject (Acarturk et al., 2008b; Habel & Acartürk, 2009, 2011).

The Assessment Areas table of the Programme for International Student Assessment (PISA) 2012 National Final Report positioned graphics within the scope of reading, not within the scope of mathematics and science. Here, together with lists, figures and diagrams, they were shown as reading materials (Anıl et al., 2015).

Reading and interpreting graphics are skills that may not require any training. However, in a study conducted on the topic of misunderstanding hierarchical graphs, Körner (2005) emphasized the significance of training on graph reading, and stated that hierarchical graphs can be understood very little without training. Acartürk (2015) also emphasized that little is known about how people interpret graphics, although they are used frequently.

The interaction between language and graphics is not a well-researched area when compared to research on multiple understanding of pictorial and diagrammatic representations (Habel & Acartürk, 2009). Research on the relations based on

multiple indicators consisting of graphics and vocabulary is rather limited (Parodi et al., 2018).

According to Escalante-Talavera (2018), little is known about how language learners read and solve problems while being tested. Bayrak-Karslı et al. (2020) emphasized that few studies using eye-tracking technology are applied to text-based problem solving. While it is important to work on perceptions of different graphic types, it is also important to investigate the problem situations containing the graphic visuals in L2 learning situations and how the minds of L2 learners work in this process. In this way, we can obtain significant findings in terms of both the language learning processes of the learners and their future in the education system, for which they are getting ready.

According to Özer et al. (2020), studies in the field of eye-tracking are limited in Türkiye. Tepgeç and Seferoğlu (2019) examined the theses that explored the eye-tracking technique in Türkiye and they could not identify any studies conducted in the field of teaching Turkish as L2 during the period of 2006 to 2018.

Haladyna et al. (2002) stated that researchers rarely evaluate the effect of item-writing techniques according to individual characteristics such as age, gender, and ability level. In this respect, it is important to investigate in the current study the effects of many individual variables on eye-movements, as addressed in the context of multiple-choice graphic questions.

1.5 Previous studies

There are studies reported in the literature that mostly associate the eye-tracking technique with multiple-choice questions and problem solving in the context of native language. In one of these studies, Tsai et al. (2011) investigated the use of the eye-tracking technique to study how students solve science questions. Results showed that, when solving an image-based multiple-choice science problem, students generally placed more importance on the selected options than the unpreferred options.

In addition, successful problem solvers focused more on relevant factors. Those who solved problems unsuccessfully faced difficulties in solving the problems, recognizing the relevant factors and focusing. Tsai et al. (2011) stated that students spend more time in general screening on their preferred option and successful problem solvers examine options in a different pattern than unsuccessful solvers.

In their study, Sajka and Rosiek (2015) sought to confirm these hypotheses on multiple-choice science questions based on the eye-tracking technique. Results showed that the participants generally spent more time analyzing the selected areas. Sajka (2017) conducted an eye-tracking study on a graphic multiple-choice math question about motion and determined that experienced and inexperienced people have different problem-solving approaches.

In Tai et al.'s (2006) pilot study on multiple-choice science questions, they tried to make inferences via the eye-tracking technique about the components of students' performance. They worked with only six preservice teachers. The results revealed differences in eye-gazing behavior between different disciplines and similarities

among participants with similar scientific backgrounds. Lindner et al. (2014) compared the multiple-choice-item solving processes of students with high and low prior knowledge. Results of theirs eye-tracking study revealed that students with both high prior knowledge and low prior knowledge showed a bias toward subjectively preferred answer options. Students with high prior knowledge spent more time on objectively correct answers. The researchers concluded that students with both high and low prior knowledge showed similar patterns of attention distribution over time.

Bayrak-Karslı et al. (2020) conducted a study to investigate the effect of different problem-solving strategies used in multiple-choice questions on success, time to solve the question, the number of fixations, and the amount of regression to the previously read field was investigated. Time spent on the question part, the number of fixations, and the number of returns to the read field were found to be significantly higher for those who used the strategy of reading the question part first compared to those who used the paragraph strategy.

In Karaoğlan-Yılmaz and Yılmaz's (2019) study, they evaluated one question asked on the exam within the scope of the Instructional Technologies and Material Design in the Educational Sciences section of the 2018 Public Personnel Selection Examination (KPSS). In the study conducted with 13 university students, the researchers determined that the student who took the longest time to answer the question exhibited a lot of saccadic motion between the question part and answer options. They reported that this student found the correct answer by comparing each option with the question part. They also determined that the participant who answered the question in the shortest amount of time exhibited less transition between the question part and the answer options.

In some studies, researchers investigated via the eye-tracking technique issues such as how graphics are perceived or processed instead of focusing on multiple-choice questions or problem situations. In one of these studies, Kim and Lombardino (2015) examined the effectiveness of graphic processing according to text. Results showed that students did not always process the graphics more effectively than the text conveying the same information. Students processed the graphics much faster than the text only when more complex questions were presented. The advantage of graphics over text was less pronounced when more complex graphic patterns were presented. Students also spent most of their time looking at specific information about axes and labels, with the increasing complexity of charts and tasks.

In his research on hierarchical graphics, Körner (2004) determined via the eye-tracking technique that all visual features in terms of overall response times and fixation numbers have an effect on graphic comprehension, to varying degrees. He revealed that planarity was the dominant visual feature, while other features had a smaller effect. He determined that there was no significant effect for slope and level in terms of total fixation numbers and that the visual features are independent of each other.

Acarturk et al. (2008a) studied how the graphic form (e.g., straight or bent shape) and the annotation mark affect the interpretations of the temporal features of the annotated event. They also investigated how the salient points change in the case of the legend text and the legend mark. In the experimental session, the subjects examined the graph shown to them on each screen and indicated their comments by

clicking on the response scale given under the x-axis on the same screen. According to the study results, most of the subjects reported that the annotated event occurred at a certain point in time under flat graph conditions. On the other hand, almost half of the subjects reported that the annotated event occurred in a time interval under curved graph conditions. Results showed that texts in the graphs did not reveal significant changes in the distribution of the mean number of fixations and gaze time values in the previously mentioned areas of interest compared to the distributions in the unexplained graphs. Values increased with the addition of annotated text. The addition of a legend icon (and the addition of legend text) caused major changes in the distribution of average fixation numbers and the distribution of average gaze duration values in curved graphs.

In another study, Habel and Acartürk (2011) investigated people's causal inference structures while interpreting verbally labeled graphics. In this study, they explored the standard use of annotations in simple line graphs. Their results revealed that the position of the annotation and the shape of the chart line influenced reader judgments regarding the causal attributions between the event presented with the annotation and the process lines and status lines in the graph.

In another study, Acartürk (2014) examined participants' one-sentence verbal definitions of graphics and, besides the presence or absence of graphical cues, how the type of graph (i.e., agonic line, angular line, and column graph) affects the conceptualization of graph formation. After the graphic display was presented as a stimulus on the screen, the participants wrote a single sentence describing the graphic. Acartürk stated that the presence of graphical cues had a strong effect by attracting the participants' attention. While the participants spent similar amounts of time on angular and agonic line graphs, the time for column graphs was shorter.

In another study, Lohse (1993) introduced a computer program that simulates graph and table perception. This program estimates the time to answer a question embedded in a graphical representation from assumptions about the order of eye fixations, short-term memory capacity and time limits, and the degree of difficulty in obtaining information at each glance. Lohse found that the participants answered the table questions faster than the column and line graph questions. Carpenter and Shah (1998) proposed a model for understanding line graphs. The model was calculated in terms of the pattern and duration of eye fixations of college students interpreting three-dimensional line graphs.

Strobel et al. (2016) aimed to find out whether graphic readers prefer the numerically advantageous graphic type as a function of the task within the scope of the optional graphic processing hypothesis in their research. Results revealed that the participants changed their graph preferences depending on the task type and improved their preferences throughout the graph task. In addition, the rate of graphic literacy was found to be relatively high. The mean total fixation time was longer for the bar graph on the difference task and longer for the line graph on the trend task.

In some studies on graphics, researchers did not use the eye-tracking technique to investigate participants' skills such as interpreting graphics and drawing. Kıranda and Akpınar (2020) determined what kind of difficulties seventh-grade students had with graphics. Results revealed that most of the participants thought they did not have any difficulties while reading and drawing graphics. Students who thought they

had difficulties in drawing graphs indicated that they found the graphs confusing and failed to draw graphs according to the data. Those who thought they had difficulty in reading graphics found the graphics difficult, complex, and boring. The study revealed that the participants found themselves most competent with bar graphs, and stated that they had the most difficulty with line graphs.

Oruç and Akgün (2010) aimed to reveal the level at which seventh-grade students had acquired graphic reading skills in a Social Studies course. They found that the students' general graphic reading skills were below 50% and determined that the students were at a very low level in reaching an average result or finding differences based on presented data. The students' ability to read one-dimensional graphics was high, but their graphics reading skills were insufficient in two-dimensional situations such as the display of bar and line graphs together.

Sezgin-Memnun (2013) examined seventh-grade students' ability to draw and read line graphs and whether these skills differed according to mathematics course scores. Results showed that the majority of the participants had the ability to read line graphs. However, a significant number of students' line graph-drawing skills were found to be inadequate. Sezgin-Memnun concluded that the students' mathematics course achievement had a positive effect on their ability to read and use line graphs.

The eye-tracking technique has been used in research in which other visual aids besides graphics were the topics. In one of these studies, Hallowell and Heuer (2005) compared the distribution of eye-movements in controlled and uncontrolled image designs. Based on the irregular fixation distributions obtained at the end of the research, the researchers stated that thumbnails posed a great risk of distraction compared to the carefully prepared images. Jian (2017) used the eye-tracking technique to investigate the cognitive processes and reader characteristics of sixth-grade students in illustrated scientific texts. Eye-movement data showed that well-performing readers spent significantly more reading time on the whole text, text portion, and diagram section than poor performers. Well-performing readers showed significantly longer average fixation durations than poor-performing readers. Also, well-performing readers made more transitions between text and diagrams.

Hannus and Hyöna's (1999) study examined the effect of authentic textbook materials on learning among 10-year-old primary school children with high and low abilities. Children's eye-movements were measured while learning picture textbook passages to examine how children divided their attention between text and pictures. Results showed that learning was largely guided by the text, and children viewed the pictures at the lowest level.

Jian and Wu (2015) investigated the strategies readers use while reading a diagrammatic science article. Undergraduate students read a scientific article while their eye-movements were tracked and then completed a reading comprehension test. The researchers found that the text-diagram referencing strategy was widely used. However, they found that some readers also used other strategies, such as reading the diagram or the text first.

Mason et al. (2015) used the eye-tracking tecnique to investigate the processing of illustrated text. In their study, seventh-grade students read an illustrated science text while their eye-movements were recorded. A cluster analysis revealed two

processing patterns that differed for the time spent on key concepts in text and image during the first-pass reading. During the reread, two models of stronger and weaker integrative processing emerged. The combination of verbal and graphic information emerged with the frequency of the second jumps of transition from text to picture and picture to text, and with the time of reexamination of the picture when rereading the text information and rereading it. Hierarchical regression analysis revealed that integrative processing models uniquely predicted verbal and graphical recall and information transfer during the second-pass reading only.

Mason et al. (2013b) used the eye-tracking technique to examine primary school students' reading of an illustrated science text. They identified three patterns based on clustering analysis findings: high integrators, intermediate integrators, and low integrators. The researchers explained the characteristics of these patterns in the context of eye-movements.

Schnotz et al. (2014) investigated whether image processing differs from text processing. They analyzed pupils' eye-movements in terms of the number of fixations in different areas of interest and eye-movement transitions between these areas. Results revealed that text and images served fundamentally different functions associated with different processing strategies in goal-directed knowledge acquisition and that texts were more likely to be used for context-oriented general processing.

Canham and Hegarty (2010) investigated psychology students' understanding of weather maps and whether the information reduction hypothesis could be generalized to the understanding of weather maps. They also investigated whether extraneous information in a graphical representation reduces students' performance and how the extra information interacts with subject area knowledge. They tracked the students' eyes before and after they taught them the meteorological principles.

The eye-tracking technique is a methodology used in both native language and L2 studies. In one of the studies within this scope, Baştuğ et al. (2019) compared university students' eye-movements in oral and silent reading. They found that the number and duration of the participants' blinks were higher in oral reading than in silent reading, but this difference was not statistically significant. The researchers determined that the number of pauses and the average pause time were significantly higher in oral reading than in silent reading and that the distance of the navigation paths was greater in oral reading.

Harada (1988) used the eye-tracking technique to examine the effect of three different orthographic presentations of a text in Japanese on the reading behavior of native and nonnative Japanese readers. Results indicated that the intermediate group whose native language was not Japanese was significantly more fixed than those whose native language was Japanese. This group also showed more regressive fixations than native Japanese speakers. To accommodate unfamiliar text presentations, the nonnative Japanese intermediate group showed significantly greater fixation than native Japanese speakers. They also showed significantly more regression than the native language group. Beginner nonnative speakers of Japanese also showed a significantly longer fixation duration than the native language group.

Lim (2017) also used the eye-tracking technique to examine the reading processes of Chinese L2 learners of English in the iBT TOEFL reading section. Only eye-tracking data of high scorers were examined here (i.e., eye-movement pattern and

fixation duration). Results revealed that the participants predominantly exhibited careful reading below the paragraph level. Lim determined that the questions at the word level were also unsuccessful in revealing the learners' inference ability.

Escalante-Talavera (2018) used the eye-tracking technique to study native Spanish speakers of English and determine how quickly learners at different levels read texts at different difficulty levels. Escalante-Talavera determined the variability in speed reading by the interaction of the learner and text level. Regardless of the learner level, the reader groups adapted to the text difficulty and read more carefully as the difficulty level increased.

In his eye-tracking study, Ayhan (2019) aimed to determine what kind of differences and similarities existed in the Turkish reading performance of L2 learners of Turkish. Another purpose of his study was to level the reading texts prepared by different institutions. Ayhan included in his report the relations between gender, native language reading direction, and age independent variables and eye-tracking measurements in the context of the books titled *Gazi, Hitit*, and *Yunus Emre*.

Yurtman-Kaçar (2018) aimed to validate multiple text reading tasks on foreign language proficiency exams through oral-protocol and eye-tracking techniques. In the study, she presented questions about reading multiple texts on language proficiency exams such as the The Examination for the Certificate of Competency in English (ECCE), The Michigan English Test (Met), Integrated Skills in English (ISE) II to 10 students. Results revealed that the Integrated ISE II exam did not aim to adequately and accurately measure multiple text reading skills. On the other hand, the Met was not special enough in terms of making skill definitions functional. Yurtman-Kaçar concluded that ECCE aimed to adequately and comprehensively measure multiple text-reading skills.

Based on this review of the existing literature, it is clear that there are no reports of studies based on the eye-tracking technique regarding the role of graphics which have an important emphasis in the CEFR and in L2 teaching.

1.6 Research questions

The following research questions served as the framework of this study:

1. Is there a relation between success on the GT and success on the TÖMER/DİLMER comprehension exam?
2. Does the duration of the first fixation eye-movement parameter differ according to gender, direction of reading in the mother tongue, university department, education level, opting for graphic questions on the TÖMER/DİLMER, GT success, age, country, and mother tongue?
3. Does the duration of the total fixation eye-movement parameter differ according to gender, direction of reading in the mother tongue, university department, education level, opting for graphic questions on the TÖMER/DİLMER, GT success, age, country, and mother tongue?
4. Does the number of the fixation eye-movement parameter differ according to gender, direction of reading in mother tongue, university department, education

level, opting for graphic questions on the TÖMER/DİLMER, GT success, age, country, and mother tongue?
5. Does the total visit time eye-movement parameter differ according to gender, direction of reading in mother tongue, university department, education level, opting for graphic questions on the TÖMER/DİLMER, GT success, age, country, and mother tongue?
6. Does the number of visits eye-movement parameter differ according to gender, direction of reading in mother tongue, university department, education level, opting for graphic questions on the TÖMER/DİLMER, GT success, age, country, and mother tongue?

1.7 Limitations

This research is limited to: (a) 115 learners at the B1 level learning Turkish and 53 learners at the test development stage, (b) eye-tracking data obtained during the solution of the test consisting of five multiple-choice graphic questions, (c) technical specifications of the eye-tracking device and software used in the application, (d) the possibilities of the application laboratory, and (e) demographic information from the participants.

2 Method

2.1 Research design

The study is basically relational survey research as it focuses on the relations between the independent variables determined for the participants and their eye-movement measurements. In relational research, the relation between two or more variables is examined without manipulating these variables in any way (Büyüköztürk et al., 2015).

In addition, in the observation study based on the eye-tracking technique, the issue of what eye-tracking conditions should be was also taken into consideration. In this context, it is necessary to know the distinction between experiments and non-experimental observational studies. This distinction is based on the manipulation of an independent variable. Observational studies are often done without manipulation of an independent variable (e.g., consider gender as an independent variable; it cannot be manipulated; Duchowski, 2007). On the other hand, if multiple stimuli are to be tested, a decision must be made whether to choose a between-subjects approach or a within-subjects approach. In a between-subjects design, each participant is exposed to one of the tested interfaces or products. Therefore, if there are two interfaces to test, half of the participants will interact with Interface A and the other half with Interface B. On the other hand, in a within-subjects design, each participant is exposed to the tested products one after the other (Bojko, 2013). We conducted this research as an observation based on an eye-tracking study using a within-subjects design.

2.2 Participants

The study group of this research consisted of 115 foreign learners at the B1 level learning Turkish in the Turkish teaching centers of two state universities. Some eye-tracking systems may not successfully calibrate certain eyes or eyes with glasses. Therefore, experts recommend including 10–20% more participants in such studies (Goldberg & Helfman, 2010). Based on this recommendation, this study aimed to reach a relatively large number and used another group of 53 learners from the B1 level during the development of the multiple-choice GT. In their study examining the theses related to eye-tracking in Türkiye, Tepgeç and Seferoğlu (2019) showed that the study groups mostly did not exceed 50 participants. The literature recommends working with large samples in eye-tracking studies as well (Mason et al., 2013a, 2015; Sajka & Rosiek, 2015; Strobel et al., 2016). In this respect, using a relatively large number of participants was a positive aspect of this study. All participants voluntarily participated in the study and signed a voluntary participation form.

2.3 Procedures

2.3.1 Development of questions

Before creating the multiple-choice graphic questions, we examined the programs related to both teaching Turkish as L2 and teaching it as a mother tongue in Türkiye and abroad besides the CEFR. In this way, we examined a total of 18 programs and identified the descriptors and acquisitions that refer to the graphics in these programs. Next, we wrote the acquisitions down with the help of a program development specialist with a doctorate degree in the field. Before writing the questions, we examined the standardized exams for both L2 learners of Turkish and for Turkish students learning Turkish as a part of their general education. In this way, we investigated a total of 466 exams and identified a total of 152 multiple-choice graphic questions, and found that 195 line charts and 71 bar charts were used in the exams. Hence, we decided to prepare the questions predominantly on line and bar charts. In addition, we avoided questions that required specific content knowledge and advanced math skills. While designing the questions, we took into consideration the findings and recommendations regarding graphic and question design in the literature (Carpenter & Shah, 1998; Haladyna et al., 2002; Kim & Lombardino, 2015; Körner, 2005; Lohse, 1993; Özçelik, 2010; Schiff, 1998; Tufte, 2006).

Next, we created a draft test with questions about the acquisitions prepared for the B1 level and presented the draft test to two experts in teaching Turkish as L2 and an assessment-evaluation expert who all had at least 3 years of experience in the field. These experts examined the questions in terms of their compatibility with the teaching of Turkish as L2 at the B1 level and the measurement-evaluation principles. We finalized the questions based on expert opinions and reduced the number of options in the questions from five to four. In addition, experts in teaching Turkish as L2 anticipated a 3-min response time per question. Consequently,

we gave this draft test to a group of 53 learners at the B1 level as a pilot study. After the pilot application, we determined the item difficulty and discrimination values of the questions and identified five multiple-choice graphic questions with medium difficulty (0.40 – 0.60) and high discrimination (> 0.30). In addition, we compared the scores obtained from the GT and the scores obtained from the TÖMER/DİLMER B1 level comprehension exam to understand whether the developed test could be used in the TÖMER/DİLMER as a valid test.

2.3.2 Pilot eye-tracking study

We conducted a pilot eye-tracking study with 21 B1-level participants for the five multiple-choice graphic questions. We used the pilot study to identify the problems that might be encountered during the actual implementation. After the pilot study, we optimized the questions for reading from the screen.

2.3.3 Main implementation

A total of 94 learners at the B1 level participated in the main application which lasted 4 weeks. Each participant took a 15-min multiple-choice test. Before starting the test, we utilized two extra questions to ensure that each participant was familiar with the process. In this preliminary application, we informed the participants about how to position themselves in front of the screen, how to answer the questions, how to move on to the next question after answering a question, and how to finish the test. Before starting the test, we explained the procedures to follow:

- A five-point calibration
- Recalibration in case of calibration error
- Appearance of the first question on the screen
- Reading and responding to the question
- Moving on to the next question by pressing the space bar
- Ending the test by pressing the space bar at the end of the test. If the time is over, the researcher ends the test.

We provided the questions to all participants in the same order (e.g., Klein et al., 2021; Tai et al., 2006). We informed the participants that they would not have a chance to return to the same question after solving it (e.g., Lindner et al., 2014; Schnotz et al., 2014; Tai et al., 2006).

One of the researchers recorded the answers provided by the learners on the form in front of him. To avoid mistakes that could be made by the researcher, we discussed each question with each student at the end of the exam and had them confirm their responses. In their study, Tsai et al. (2011) also asked their participants to verbally confirm their choices on multiple-choice questions.

2.3.4 Device

We recorded participants' eye-movements while they read the graphic multiple-choice questions with a Tobii brand TX300 model eye-tracking device with a sampling rate of 300 Hertz (Hz) per sec.

The Tobii brand was a preferred tool in various studies in which visuals were investigated (e.g. Acartürk & Habel, 2012; Acarturk et al., 2008a; Goldberg & Helfman, 2010; Mason et al., 2013b). We processed the recorded data with Tobii Pro 3.5.8 software.

2.3.5 Test conditions

We recorded participants' eye-movements in an environment with no windows and which was illuminated with fluorescent lamps. Participants sat in a height-adjustable chair during testing and we did not use a head or chin stabilizer to achieve a more natural process.

2.4 Data analysis

2.4.1 Identifying the areas of interest

For each question, we identified basically three different areas of interest: an explanation paragraph, a graphic, and a question statement with options. In the literature, a similar path is followed in eye-tracking studies on multiple-choice questions with images such as graphics (Klein et al., 2021; Schnotz et al., 2014; Tai et al., 2006). We identified the areas of interest represented by the explanation paragraph, graphic, and option-question statements in pixels horizontally and vertically. In the horizontal dimension, the areas of interest such as the explanation paragraph and the question statement with options had an average size of about 100 pixels, and the graphic areas of interest had a size of 96 pixels. In the vertical dimension, the explanation paragraph had an average size of approximately 75 pixels, the graphic areas of interest had 179 pixels, and the question statement with options had a size of 109 pixels. Thus, we can argue that the text and visual fields in the test, consisting of graphic multiple-choice questions, had horizontal and vertical dimensions close to each other.

2.4.2 Data cleaning

In this research, we followed two data elimination methods. In the first method, we determined a fixation lower limit as was done in several studies (Abegg et al., 2015; Acarturk et al., 2008b; Cook et al., 2008; Mason et al., 2013a; Tuninetti et al., 2015). The second method was related to the inability to record the data correctly (Carpenter & Shah, 1998; Jian, 2017; Lindner et al., 2014). In this study, we determined the fixation lower limit to be 60 ms. Additionally, we examined each participant's

eye-movement patterns. Similarly, Strobel et al. (2016) visually investigated the patterns to determine whether the eye-movement data were recorded accurately. We excluded participants whose eye-movement patterns could not be detected in general on the questions from the data set. We also excluded participants whose eye-movement patterns could not be calibrated by repeated measurement during the application. We did not analyze the data for eight participants due to poor eye-tracking data and poor measurement in the original application. Thus, we excluded from the analysis the data of 8.51% of the 94 people who participated in the actual application.

2.4.3 Statistical analyses

We used the SPSS 21.0 package program to analyze the quantitative data in the research. Before analyzing the data, we used normal distribution graphs, skewness/kurtosis coefficients, and the Kolmogrov-Smirnov test to check whether the continuous variables were normally distributed. We accepted the span of + 2/-2 as the appropriate span in the skewness/kurtosis coefficient and decided that the data were not normally distributed by taking 0.05 as the significance threshold in the Kolmogrov-Smirnov test results and since the obtained value was below this criterion. We used Mann–Whitney U and Kruskal–Wallis H tests, which are nonparametric tests, in the cases where the data were not normally distributed. We used the Friedman test and the Wilcoxon signed rank test to compare group means. In the case of normal distribution of the data, we utilized independent groups *t*-tests and one-way analysis of variance while using dependent groups *t*-test and analysis of variance for related groups for comparison of group means. In classifying the participants as a lower and upper group members according to the scores obtained from the test consisting of graphic multiple-choice items, we sorted the scores and accepted those with over 60 points as the upper group, and accepted those with scores lower than 60 as the lower group similar to Jian's (2017) implementation.

2.5 Baseline eye-tracking metrics

We took the following eye-tracking metrics as the dependent variables in the study:

Duration of first fixation indicates the duration of initial fixation during the first transition to an area of interest or a group of areas of interest (Tobii Studio Version 3.4.8 User's Manual [TSVUM], 2017).
Duration of total fixation represents the sum of the durations of all fixations regarding an area of interest or a group of areas of interest (TSVUM, 2017).
Fixation number refers to the number of times an area of interest or a group of areas of interest has been fixated on. The fixations displayed when leaving the fixated media element and returning back to the same element are also taken into account (TSVUM, 2017).
Total visit time measures the duration of all visits to an active area of interest or a group of areas of interest (TSVUM, 2017). Total visit time represents the sum of saccade and fixation times (Wong et al., 2014).

Number of visits indicates the number of visits within an active area of interest or a group of areas of interest (TSVUM, 2017).

3 Findings

In this section, we presented the findings of the study in tables. Values regarding the relations between the *gender*, *direction of reading in mother tongue*, *department*, *education level*, *opting for graphic questions on the TÖMER/DİLMER*, *GT success*, *age*, *country.* and *mother tongue* variables and eye-movement parameters. Each separate table shows the relations between these variables and a single eye-movement parameter.

First of all, we examined the relation between participants' success on the GT and the comprehension exams in the TÖMER/DİLMER to see if the GT was an exam that could be used in the TÖMER/DİLMER. Thus, we tried to understand whether GT was a valid exam that could be used in these language teaching centers. According to Table 1, we found a moderate, positive, and significant relation between the GT and the TÖMER/DİLMER comprehension exams ($p = 0.000 < 0.05$). From this point of view, we can argue that the GT can be used in the TÖMER/DİLMER as comprehension tests.

Table 2 shows the relations between the parameter of duration of the first fixation and various variables. According to this table, only the level of education variable significantly affected the duration of the first fixation ($p = 0.029 < 0.05$). Accordingly, graduate-level participants exhibited a longer duration of first fixation. We can explain this situation as follows: The duration of the first fixation parameter refers to preattentive processes before the purposeful reading behavior in the areas of interest. Therefore, graduate-level participants showed longer preattention scanning behavior than other participants.

As for other variables, male and female participants exhibited almost equal durations of first fixation. Therefore, there was no difference in preattention scanning behavior in terms of gender. Participants who read and write from left to right in their native language, as in Turkish, exhibited slightly longer duration of

Table 1 Relations between success on the GT and success on the TÖMER/DİLMER B1 comprehension exam

			GT success	TÖMER/DİLMER b1 exam success
Spearman's rho	Success in GT	Correlation Coefficient	1.000	0.485**
		Sig	-	0.000
		N	86	86
	Success in TÖMER/DİLMER B1 Exam	Correlation Coefficient	0.485**	1.000
		Sig	0.000	-
		N	86	86

Table 2 Relations between the parameter of duration of first fixation and various variables

Independent variables		*N*	Mean rank	U	*p*	X^2	df
Gender	Female	35	43.94	877.000	0.892		
	Male	51	43.20				
Direction of reading	From left to right	45	46.64	781.000	0.221		
	From right to left	41	40.05				
Department	Other	42	40.81	811.000	0.329		
	Sciences	44	46.07				
Education level	High school-undergraduate	46	38.01	667.500	0.029		
	Graduate	40	49.81				
Opting for graphic questions	Yes	49	42.58	861.500	0.695		
	No	37	44.72				
GT success	Upper	37	39.92	770.500	0.235		
	Lower	49	46.28				
Age	Lower than 20	33	36.50		0.121	4.221	2
	Between 20 and 24	25	48.26				
	25 or higher	28	47.50				
Country	Africa	26	47.37		0.539	1.236	2
	Middle East	37	40.34				
	Asia and other	23	44.22				
Mother tongue	Afro-Asiatic	33	38.38		0.195	4.702	3
	Altaic	20	41.85				
	Indo-European	25	46.88				
	Other	8	58.19				

the first fixation than those who read from right to left. Therefore, it can be said that those who read from left to right tend to have a slower preattention process. However, the difference is statistically insignificant.

Participants from science departments also showed a slightly longer duration of first fixation than participants from other departments. Therefore, it can be said that participants from science departments tend to have longer preattention periods. However, this difference is also statistically insignificant. Those who opted for questions with graphics in their Turkish learning courses, that is, those who had a more positive attitude toward this subject, exhibited a slightly less duration of first fixation than those who did not opt for such questions. However, this attitude difference was not statistically significant.

Participants in the upper group in terms of success on the GT showed shorter durations of first fixation, indicating a faster preattention process. However, this difference is also not statistically significant. Although the group under the age of 20 exhibited shorter duration of first fixation, the age factor did not cause a statistically significant variability in this parameter. Although the participants from the Middle East countries showed slightly shorter duration of first fixation, the country group variable did not cause a statistically significant change in

this parameter. Participants speaking one of the Afro-Asiatic languages as their mother tongue exhibited the least duration of first fixation. Participants belonging to the Altaic language family, including Turkish, tend to exhibit the least duration after this group. However, the mother tongue variable did not cause a statistically significant change in this parameter. As a result, the fact that other variables such as age, gender, native country, and mother tongue did not have a significant effect on the duration of first fixation may be interpreted with a similar preattention process employed by the participants.

According to Table 3, gender and success on the GT significantly affected the duration of total fixation employed by the participants ($p = 0.048$ and $0.033 < 0.05$). Accordingly, male participants fixated longer than females while answering the GT questions. Here, we observed that the participants in the lower group regarding achievement fixated for significantly longer periods ($p = 0.033 < 0.05$). Therefore, we can argue that a shorter duration of total fixation is sufficient to be successful on a test consisting of graphic questions. We can add that this is also related to a faster scanning ability. Duration of the total fixation parameter is related to the process after the first preattention scanning behavior. This parameter indicates how long the reader stays in the areas of interest to understand the stimulus. Therefore, we can

Table 3 Relations between the parameter of duration of total fixation and various variables

Independent variables		*N*	Mean	*p*	*F*
Gender	Female	35	455.4306	0.048	
	Male	51	518.5539		
Direction of reading	From left to right	45	483.3760	0.530	
	From right to left	41	503.2780		
Department	Other	42	496.9688	0.800	
	Sciences	44	488.946		
Education level	High school-undergraduate	46	480.6185	0.421	
	Graduate	40	506.9468		
Opting for graphic questions	Yes	49	511.6976	0.169	
	No	37	467.9227		
GT success	Upper	37	454.4254	0.033	
	Lower	49	521.8894		
Age	Lower than 20	33	472.2527	0.583	.544
	Between 20 and 24	25	502.1024		
	25 and higher	28	508.9079		
Country	Africa	26	534.7108	0.111	2.257
	Middle East	37	491.7414		
	Asia and other	23	447.3657		
Mother tongue	Afro-Asiatic	33	496.1488	0.906	.186
	Altaic	20	509.8495		
	Indo-European	25	479.5596		
	Other	8	478.4288		

state that female participants and participants in the upper group in terms of GT success read and understand the questions faster.

Although those reading from left to right in their mother tongue exhibited a shorter duration of total fixation, this difference is not statistically significant. Participants from the science departments also showed a shorter duration of total fixation, which was not statistically significant. In terms of education level, the tendency of graduate-level participants to exhibit longer duration of total fixation is also not statistically significant. The duration of total fixation of the participants who opted for questions with graphics in the Turkish course did not change statistically. The duration of total fixation tends to increase as the age of the participants increases. However, the age factor did not cause a statistically significant difference on this parameter. Although the participants from African countries showed longer duration of total fixation, the country group did not make a statistically significant difference on the parameter. The mother tongue of the participants also did not cause a statistically significant difference on the duration of total fixation parameter. As a result, participants showed similar duration of total fixation on the GT in terms of variables other than gender and achievement.

According to Table 4, only the variable of opting for questions with graphics on TÖMER/DİLMER exams significantly affected the number of fixations exhibited by the participants while answering the GT questions ($p = 0.028 < 0.05$). Accordingly, the participants who opted to have questions with graphics on these exams, that is, those with positive attitudes toward questions with graphics, fixated more on the areas of interest on the questions. The number of fixations parameter is related to the intensity of fixations, which are eye-movements for understanding. Therefore, we can say that the participants opting for questions with graphics in Turkish courses concentrated their eye-movements on the questions more while reading the questions.

Other variables did not significantly affect the number of fixations. Although female participants exhibited fewer fixations as in the duration of total fixation parameter, the gender variable did not cause a statistically significant difference on this parameter. Participants who read from left to right in their mother tongue also showed fewer fixations. However, this variable did not make a statistically significant difference on the number of fixations. Although students from science departments exhibited slightly more fixations, the department variable did not cause a statistically significant difference on this parameter. Although graduate-level students showed slightly fewer fixations, education level did not cause a statistically significant difference on this parameter.

Participants in the upper group in terms of success on the GT showed slightly fewer fixations. However, being in the upper or lower group in terms of GT success did not make a statistically significant difference on the number of fixations. According to the age variable, the number of fixations did not change statistically significantly. Although the participants from African countries showed a slightly higher number of fixations, the country variable did not cause a statistically significant difference on this parameter. The number of fixation values of the participants belonging to the Altaic language family, including Turkish, were determined as the highest. However, the mother tongue variable was not reflected in the number of

Table 4 Relations between the parameter of number of fixations and various variables

Independent variables		*N*	Mean	*p*	*F*
Gender	Female	35	1643.4286	0.082	
	Male	51	1831.5490		
Direction of reading	From left to right	45	1710.2667	0.381	
	From right to left	41	1804.0732		
Department	Other	42	1741.1667	0.801	
	Sciences	44	1768.1818		
Education level	High school-undergraduate	45	1792.8043	0.449	
	Graduate	40	1711.5000		
Opting for graphic questions	Yes	49	1856.2041	0.028	
	No	37	1620.9459		
GT success	Upper	37	1704.4595	0.412	
	Lower	49	1793.1729		
Age	Lower than 20	33	1780.8970	0.792	0.234
	Between 20 and 24	25	1697.8000		
	25 and higher	28	1775.7500		
Country	Africa	26	1819.6923	0.583	0.543
	Middle East	37	1760.8108		
	Asia and other	23	1672.4783		
Mother tongue	Afro-Asiatic	33	1725.1818	0.147	1.838
	Altaic	20	1966.3500		
	Indo-European	25	1680.8400		
	Other	8	1581.2500		

fixations parameter as a significant difference. As a result, participants exhibited similar numbers of fixations in terms of variables other than opting for graphic questions on the TÖMER/DİLMER variable.

According to Table 5, only success on the GT significantly affected the total visit time, that is, the total time spent by the participants in the areas of interest on the questions ($p = 0.025 < 0.05$). Accordingly, those in the upper group regarding GT success visited the areas of interest on the questions for a shorter period of time. Total visit time is the sum of saccade and fixation durations (Wong et al., 2014). This may serve as information about the participants' fixation movements while reading the questions, as well as the time spent between these movements. Table 5 shows that the same group also spent less time between fixations, which are movements for understanding.

Although not statistically significant, female participants exhibited less total visit time than male participants. Again, although not statistically significant, those who read from left to right in their mother tongue showed slightly less total visit time. The department variable did not have a statistically significant effect on the total visit time. Although graduate-level participants tended to exhibit more total visit time, education level did not cause statistically significant variability on this

Table 5 Relations between the parameter of total visit time and various variables

Independent variables		*N*	Mean	*p*	*F*
Gender	Female	35	523.3497	0.091	
	Male	51	584.7198		
Direction of reading	From left to right	45	545.4449	0.404	
	From right to left	41	575.4373		
Department	Other	42	561.8076	0.911	
	Sciences	44	557.7734		
Education level	High school-undergraduate	46	546.6480	0.444	
	Graduate	40	574.8035		
Opting for graphic questions	Yes	49	588.1645	0.066	
	No	37	522.1051		
GT success	Upper	37	514.0800	0.025	
	Lower	49	594.2243		
Age	Lower than 20	33	541.3064	0.684	0.382
	Between 20 and 24	25	563.3180		
	25 and higher	28	578.2818		
Country	Africa	26	608.3212	0.110	2.269
	Middle East	37	557.0400		
	Asia and other	23	509.1791		
Mother tongue	Afro-Asiatic	33	564.0894	0.853	0.261
	Altaic	20	582.5805		
	Indo-European	25	541.9204		
	Other	8	540.4225		

parameter. Participants opting for graphic questions in the courses in which they were learning Turkish displayed more total visit time than participants who did not. However, this difference was not reflected in the results as statistically significant. As the age of the participants increased, the total visit time value also increased. However, the age variable did not cause a statistically significant difference on the total visit time parameter.

Participants from African countries also tended to get high values in the total visit time parameter. However, the country variable did not make a statistically significant difference on the results. Although the total visit time values of the Altaic speaking participants tended to be the highest, the mother tongue variable did not cause a statistically significant variability on this parameter. As a result, the participants exhibited similar total visit time in terms of variables other than the variable of GT success.

According to Table 6, the variables of level of education and opting for questions with graphics on TÖMER/DİLMER exams significantly affected the number of visits exhibited by the participants ($p = 0.044$ and $0.009 < 0.05$). Accordingly, the participants at the graduate level visited the areas of interest on the questions less frequently. On the other hand, the participants who opted for questions with

Table 6 Relations between the parameter of number of visits and various variables

Independent variables		*N*	Mean rank	U	*p*	X^2	df
Gender	Female	35	39.07	737.500	0.173		
	Male	51	46.54				
Direction of reading	From left to right	45	39.80	756.000	0.150		
	From right to left	41	47.56				
Department	Other	42	40.69	806.000	0.308		
	Sciences	44	46.18				
Education level	High school-undergraduate	46	48.55	687.500	0.044		
	Graduate	40	37.69				
Opting for graphic questions	Yes	49	49.59	608.000	0.009		
	No	37	35.43				
GT success	Upper	37	44.01	887.500	0.868		
	Lower	49	43.11				
Age	Lower than 20	33	47.02		0.587	1.065	2
	Between 20–24	25	41.10				
	Higher than 25	28	41.50				
Country	Africa	26	43.50		0.658	0.836	2
	Middle East	37	45.82				
	Asia and other	23	39.76				
Mother tongue	Afro-Asiatic	33	40.14		0.119	5.853	3
	Altaic	20	54.05				
	Indo-European	25	43.16				
	Other	8	32.06				

graphics on TÖMER/DİLMER exams visited the areas of interest on the questions more often.

In terms of the gender variable, female participants exhibited fewer visits, but the difference between them and male participants was not statistically significant. Participants who read from left to right in their mother tongue also showed fewer visits, but again, the direction of reading in their mother tongue did not cause a statistically significant difference in the number of visits parameter. Participants from science departments visited the areas of interest on the questions a little more often. However, the values related to the department variable were not reflected in the results in a statistically significant way.

In terms of success on the GT, the participants in the upper and lower groups exhibited numbers of visits that were very close to each other. Although the group under the age of 20 exhibited a slightly higher number of visits, the age variable did not cause a statistically significant variability on this parameter. The country groups of the participants are also a variable that does not create statistically significant variability on the number of visits. Although participants speaking Altaic languages exhibited slightly more visits, the mother tongue variable did not cause statistically significant variability on this parameter. According to all of these,

the participants exhibited similar numbers of visits in terms of variables other than education level and opting for questions with graphics on TÖMER/DİLMER exams.

4 Discussion

The present study revealed that a number of variables affected the eye-movements of adult individuals learning Turkish as L2 in a Turkish test situation consisting of multiple-choice questions with graphics. Therefore, we can infer that these variables may have an impact on the mental processes of adult individuals learning Turkish as L2 in a test situation consisting of multiple-choice questions with graphics.

The gender of the participants in this study had a significant effect only on the duration of total fixation among the eye-movement measurements. Accordingly, male participants showed a higher duration of total fixation compared to female participants. Based on this finding, we can argue that male participants fixated longer to understand the stimulus parts. If we consider this as a reading comprehension case, we can claim that males have more difficulty in understanding stimuli. Moreover, we can argue that males tend to obtain higher values than female participants since the values were observed to approach the significance limit (0.05) in many measurements although there was no significant difference in other eye-movement measurements. Accordingly, males tended to fixate more on stimuli than females, to visit areas of interest longer and more frequently, and to reciprocate more in order to synthesize information between different parts of the questions. In measuring the duration of first fixation, indicating preattention movements before eye-movements for the purpose of deep understanding, the two groups achieved almost equal values. Therefore, they showed a partial difference in eye-movements that lead to understanding behavior after the first impression.

Based on all of these, we can report that male participants tended to exhibit longer and more eye-movements than females when the eyes started to move on the stimulus for the purpose of understanding. On the other hand, in his study on written texts, Ayhan (2019) found that the number of fixations were similar for males and females. In their study on weak and proficient Chinese readers, Zhan et al. (2020) found that proficient male readers were significantly more fixated on the text than proficient female readers. On the other hand, the situation was reversed for weak readers, where female participants were more fixated than males.

Considering that the current study represents research within the scope of visual literacy, we can investigate studies from this perspective. In this context, Arif and Hashim (2010) investigated whether there was a gender difference or similarity in visual literacy practices and found a limited difference between male and female participants in terms of visual literacy practices, as in the current study. On the other hand, Gültekin (2009) examined ninth graders' ability to read and interpret graphics about solutions and their properties and reached a conclusion in favor of female students.

Similar to reading written texts, female participants tend to be more successful in reading multiple formats both in terms of scores and in displaying fewer

eye-movements in shorter durations, which is thought to indicate successful reading behavior. Hence, in a multiple-choice test situation, when texts are accompanied by visuals, the expectation that women will be more successful was realized in this study.

While addressing perceptual span, Rayner (1998) stated that the span is not more than three or four letters from the beginning of the fixated word to the left of the fixation for the readers of alphabetic texts such as English, French, and German, and that there are approximately 14 or 15 letter spaces to the right of the fixation, hence the span is asymmetrical to the right for these readers. On the other hand, for texts written from right to left, such as Hebrew, this span is asymmetrical to the left of the fixation. Rayner (1998) stated that writing system features not only affect span asymmetry, but also affect all dimensions of perceptual span.

When Schiff (1998) discussed traditions in graphic design, he added that Americans and Europeans represent time progressively from left to right, which is evidence that it is tied to written language traditions. Therefore, we can say that it is possible for learners with different writing systems in their mother tongue to exhibit different eye-movements while solving the multiple-choice questions with graphics, but the mother tongue reading directions of the participants in the current study did not have a significant effect on eye-movement measurements. However, when it comes to measurements other than the duration of first-fixation measurement, we can say that learners who read from right to left in their mother tongues tended to exhibit more eye-movements for a longer time. Accordingly, those who read from left to right in the preattention period before they start examining the stimulus for deep understanding tended to spend longer time. When effort was expended to understand the stimulus deeply, those who read from right to left tended to expend more effort regarding time and number of visits.

Supporting the results of this study, Ayhan (2019) determined that the reading direction in the mother tongue did not have a significant effect on the number of fixations for written texts. Dehaene (2017) stated that there is a special area for reading even in the brains of Chinese and Japanese readers, and the position of this area is approximately the same as for those who read from left to right. It is worth noting that the reading direction (left to right or right to left) does not seem to affect the location of this region which remains unchanged at the border of the left occipitotemporal half, even for readers reading from right to left. Therefore, we can say that the direction of reading does not seem to cause different processing in the brain besides the fact that it does not affect eye-movements. Of course, we can recommend further investigating this subject with different groups and stimuli.

The learners' departments at their universities did not have a significant effect on their eye-movement measurements. Means indicated different trends for different measures. Accordingly, the learners studying science topics at their universities had higher values regarding the duration of the first fixation, indicating preattention processes; the number of fixations, indicating the number of eye-movements for understanding the areas of interest on the stimulus; and the number of visits, indicating the number of repetitive movements for the purpose of understanding the areas of interest.

We can argue that eye-movements tended to be associated with successful comprehension in the context of the department variable. Accordingly, we can state that the students who tended to spend more time in the preattention period during the first encounter with the stimulus, who tended to fixate more on the areas of interest, who tended to visit these areas more, and who tended to exhibit more mutual eye-movements between these parts to combine the different parts of the stimulus, but who tended to display short-term eye-movements after the preattention period may be more successful based on the variable of department. However, we should not forget that these trends were not statistically significant.

Based on the comparison of fixation and saccade density in their study of participants from three different fields (i.e., biology, chemistry, and physics), Tai et al. (2006) also stated that the distribution of these values was related to the expertise in the field. Moreover, based on the debate about whether experts or novices read faster, Thai et al. pointed out that some people spend more time on the subject in which they are experts. Considering that solving multiple-choice questions with graphics is more relevant to the expertise of learners who studied science, we can regard their tendency to use more eye-movements in the preattention period in this context.

In a study with learners from different departments, Christopherson (1997) rated the significance of visual literacy for different departments. Accordingly, among five departments, visual literacy was found to be more significant for the Fine Arts and Communication departments followed by Engineering, Business Management, Nursing, and Family and Social Sciences, respectively. The order of competence in visual literacy in these five departments was as follows: Fine Arts and Communication followed by Family and Social Sciences, Nursing, Engineering, and Business Management, respectively. The fact that science departments such as Nursing and Engineering did not take the first place in this study suggests that it is important to consider the context of visual literacy.

Learners' levels of education significantly affected two of the eye-movement measurements. Accordingly, duration of first fixation eye-movement measurements of the learners at the postgraduate level were significantly higher than those of the learners at lower education levels. Postgraduate learners may have spent more time on stimuli during the preattention period. The number of visits, which indicates the number of times and repetitive visits to areas of interest, was significantly lower in learners at the graduate level than in learners at the lower education level. Accordingly, the learners in the lower education level visited the areas of interest on the stimuli more often.

When we look at other eye measurements, we can say that although no significant relation was detected, learners at the postgraduate education level tended to exhibit eye-movements for a longer period of time. Accordingly, we can say that graduate level learners tended to have higher values in terms of fixation duration and visit time. On the other hand, learners at the lower education level tended to obtain higher values in the measurement of number fixations. The *postgraduate group* tended to obtain higher measurement values regarding the duration of eye-movements, while *the undergraduate-high school group* had higher measurement values regarding the number of eye-movements. Moreover, those who read the questions successfully

based on education level tended to display shorter preattention periods and to exhibit more eye-movements in a shorter time.

In this framework, Sajka (2017) found that academicians spent more time on problem statements than learners, as in the current study, but they visited this region less. Sajka showed that academicians did not visit the same region as often as learners. In addition, academicians fixated less on graphics, expressions, and white areas than learners, and they revisited these areas less often as well. Jiang (2012), on the other hand, determined that the level of education did not make a difference for those who learn English as a foreign language in terms of graphic organizer education or reading tests.

Gültekin (2014) could not detect a significant difference between university students and high school students in terms of reading and interpreting graphs regarding change of state, solutions, and solubility in the context of their mother tongue. In this study, the education level variable had a significant effect on some of the eye-movements during the test consisting of multiple-choice questions with graphics. In that case, we can regard level of education as an important variable to be considered while preparing such exams in teaching Turkish as L2. As a matter of fact, while classes are created for learners in many language teaching centers, different learners are in the same group in terms of education level. Therefore, these learners take the same exams. On the other hand, we can interpret the homogeneous test consisting of multiple-choice questions with graphics as more challenging for postgraduate students in the context of teaching Turkish as L2. Therefore, we should think and work more on the application of such tests in classrooms that are not homogeneous regarding the level of education.

Interest can affect comprehension in a foreign language (Ay & Bartan, 2012). In this context, Akbarov et al. (2018) investigated learners' attitudes toward the application of infographics for English learning and teaching purposes and determined that attitudes on this subject were positive. In this study, we can regard opting to have questions with graphics in the exams as interest. In addition, within the scope of the current research, we found that this interest has a certain effect on eye-movements related to the comprehension process. Accordingly, the fact that the learners opted for questions with graphics on TÖMER/DİLMER exams, that is, they had a positive attitude toward this subject, had a significant effect on the number of fixations and the number of visits. Accordingly, those who opted for questions with graphics on TÖMER/DİLMER exams were significantly more fixated on the stimuli and visited the areas of interest on the stimuli significantly more often.

Contrary to the results of this study, Chung and Kakizawa (2002) found in their study on children with congenital deafness that they displayed a lower frequency of fixation to the text that was interesting to them in the context of narrative texts. Although the variable of opting for questions with graphics on TÖMER/DİLMER exams did not significantly affect other eye-movement measurements, both means and values close to the level of significance indicate some general tendencies. Accordingly, we can argue that those who opted for questions with graphics on these exams tended to exhibit shorter preattention periods, which occurred with shorter initial fixation periods. On the other hand, in terms of eye-movements other than preattention processes, we can say that those who opted for questions

with graphics on exams tended to use more eye-movements in a longer period. As a result, we can state that those who opted for multiple-choice questions with graphics in the context of TÖMER/DİLMER exams in general tended to use a higher number of eye-movements in longer durations, except for eye-movements indicating preattention processes.

In this context, Reichle et al. (1998) reported a lot of corroborating evidence pointing to a close connection between eye-movements and cognitive processes. Sajka and Rosiek (2015) stated that too much time spent can be an indicator of trying to solve the problem by using information and at the same time having sufficient motivation. As for visual literacy in teaching Turkish as L2, we can say that there is almost no evidence about how learners' attitudes affect their reading. Therefore, we think that this study offers some foresight to researchers and practitioners about the importance of attitudes within the scope of visual literacy.

Participants' total achievement scores on the multiple-choice exam with graphics had a significant effect on duration of total fixation and total visit time. Accordingly, the learners who were more successful on the exam were fixated on the stimuli for less time and visited the areas of interest for a shorter time. Contrary to these results, Jian (2017) stated that the deep processing requires more mental resources based on the total reading-time measurement, and based on this understanding, reported that learners who understand well spend more time not only in the text section but also in the diagram section. On the other hand, relatively supporting the findings of the current study, Lindner et al. (2014) emphasized that eye-movements are highly related to option choices in decision-making situations and found that students with high content knowledge generally have shorter decision-making processes than those with low content knowledge.

On the other hand, Schnotz et al. (2014) reported that high-level students fixate more on pictures. According to Mason et al. (2013b), a longer fixation period after the first encounter with a text may be related to comprehension difficulties. In their study where they examined the eye-movements of students while solving a gas law problem, Tang and Pienta (2012) found that unsuccessful students fixated more on the questions while thinking about the solutions. Although it is not statistically significant, as in the two aforementioned eye-movement measurements (i.e., duration of fixation and visit), we can say that learners' achievement on the exam was associated with their other eye-movement measurements, showing a tendency to shorten the durations and lessen the number of visits. As an example, we can cite that successful learners tend to have shorter initial fixation times which indicate engagement in the preattention process.

This study's results show that the age variable did not significantly affect any eye-movement measurements. Therefore, the age groups in the study did not affect eye-movements (and the mind processes thought to be related to these movements) in a test situation consisting of multiple-choice questions with graphics. Working with age groups with more distinctions instead of close age groups may produce different results. As a matter of fact, in the context of using multiple document formats, the time spent by young students and adults on texts and diagrams may differ (Jian, 2017).

In addition, older readers read more slowly than younger readers, and their perceptual spans are smaller and less asymmetrical (Rayner, 2009). Some studies in the literature show that the age variable does not have an effect on eye-movements in line with the results of the current study (Ayhan, 2019) while other studies show that this variable affects eye-movements (Lee & Choi, 2019). However, some studies show that this variable partially affects eye- movements for some age groups revealed (Choo & Koh, 2019).

If cultural differences affect how participants view and encode images, there may also be differences in the pattern of eye-movement saccades and fixations of members from two different cultures (Chua et al., 2005). Rayner (2009) also examined whether culture affected where one looked in an image and reported that Chinese participants looked at the scenes differently than English speakers. Chua et al. (2005) revealed that eye-movements can change as a function of culture. Easterners and Westerners allocate their attention resources differently when viewing images. Rayner (2009), on the other hand, found no different perspectives on image according to culture.

Considering that learners' eye-movements are not independent of their cognitive processing styles, we can expect some differences between the eye-movements of learners from different countries. However, the current study found that the countries of origin did not significantly affect any eye-movement measurement. However, the means indicated some general trends. Accordingly, the *Asia and other* group, including learners from Turkic Republics, tended to obtain the lowest values in all measurements except the duration of first fixation measurement. In other words, we can say that learners in the *Asia and other* group tended to examine stimuli faster and with fewer eye-movements. Learners from African countries tended to obtain the highest values in the measurements of duration of first fixation, duration of total fixation, number of fixations, and total visit time. Therefore, we can argue that African learners tend to examine stimuli for longer durations and with a higher number of eye-movements to demonstrate a longer preattention process and understanding. Learners from the Middle Eastern countries tended to be ahead in the measurement of the number of visits. Based on these results, we can interpret that learners from the Middle Eastern countries tended to visit parts of stimuli more and made more efforts to combine different parts in their minds.

Learners' mother tongues did not significantly affect eye-movement measurements. However, the means indicated some general trends. Accordingly, we can say that the learners whose mother tongues were in the Altaic group, including Turkish, tended to achieve the highest value in all measurements other than duration of first fixation. Accordingly, when the learners who spoke one of the Altaic languages first encountered the stimulus, they tended to obtain high values in all movements except the eye-movements related to the first scanning behavior. We can infer that the tendency to use more eye-movements for longer periods of time in the mother tongue may be associated with successful reading. However, we need to think about the fact that the mother tongue did not have a significant effect on eye-movements.

The studies reported in the literature examined the effect of mother tongue on second language learning. In their study with native German, French, and Dutch speakers who were learning English as L2, Lemhöfer et al. (2008) found that the

participants' processing of second-language words was similar to their mother tongue. In their study based on English online newspaper data, Berzak et al. (2017) examined the eye-movements of native Chinese, Japanese, Portuguese, and Spanish speakers. In this study, the researchers revealed that the similarities in the mother tongues were reflected in reading in English. Based on these studies, we can say that there is a need for more studies on the effects of different mother tongues on the eye-movements that occur in the reading process in Turkish.

It is important to remember that many of the aforementioned studies were conducted in the context of the mother tongue. Moreover, Jian (2017) reported that although eye-tracking is used to observe cognitive processes in reading illustrated scientific texts, the findings obtained in these studies are inconsistent and contradictory, by adding that many eye-tracking studies report that better learning performance is associated with a greater combination of textual and graphical information (e.g., readers make more saccades between the text and the diagram part in articles). As a result, we may claim that successful learners tend to allocate less time and employ fewer movements in all eye-movement measurements except the number of visits. Accordingly, we can argue that learners who may achieve high success on the exam, which consists of multiple-choice questions with graphics, may tend to exhibit a lower number of eye-movements for shorter durations, including the first fixation period related to preattention processes. Based on this, we can say that it may be necessary to move the eyes for a shorter time and less frequently on stimuli in order to successfully read and understand the questions on a test based on multiple-choice questions with graphics. These results are actually compatible with eye-movements in the context of text-based reading. As a matter of fact, in successful text-based reading, eyes go backwards less often, and they show less fixation for a shorter period of time.

5 Conclusion

This study aimed to present how a homogeneous test consisting of multiple-choice questions with graphics with different structures and meanings can be used in teaching Turkish as L2 at the B1 level, and what differences can occur based on learners' independent characteristics and eye-movements. With the help of the eye-tracking technique, this study determined which variables could affect the eye-movements which impact both achievement and the cognitive processes. Thus, we tried to provide guidance on what should be taken into account when preparing such exams. We recommend a homogeneous exam with multiple-choice questions with graphics with different features as a test type in the field of teaching Turkish as L2 at different levels. Such tests can also be used to check for skills acquisitions after training based on the interpretation of graphics and texts with multiple formats.

As a suggestion, we emphasize again that TÖMER/DİLMER courses should provide training on reading and understanding graphics and documents with multiple

formats accompanied by graphics in Turkish. Training and materials can be prepared for the strategic reading behavior required for the interpretation of such documents. Eye-movement recordings of people who are experts in interpreting multiple documents can be watched as videos during these training sessions. Mason et al. (2013b) also maintained that eye-tracking records of good readers can be shown as an example to other learners.

Learners' metacognitive awareness can be increased both in text-based reading and in the context of visual literacy by showing them their own eye-movement records. Özer (2019) also stated that determining the eye-movement patterns of learners will have positive reflections on many areas of development, including their academic skills.

Studies can be conducted to acquire the vocabulary shaped around multiple documents accompanied by visuals and graphics.

The eye-movement records obtained in this study while the learners took the test consisting of multiple-choice questions with graphics and learners' achievement on this exam presented some significant results about individual variables.

Although some general trends emerged, learners' native languages and their university departments did not cause significant changes in the eye-movements.

We found that learners' gender and whether they opted for graphic questions on the TÖMER/DİLMER significantly affected their eye-movements when they took the test consisting of multiple-choice questions with graphics.

Level of education was the variable that had a significant effect on some of the eye-movements. Accordingly, the eye-movements of the postgraduate students differed significantly.

We can make some interpretations based on the three situations previously cited. First of all, the level of education variable is associated with eye-movements, which may be related to both achievement and cognitive processes during reading. We know that learners from different education levels learn together in language-teaching classes. on the TÖMER/DİLMER, multiple-choice questions with graphics should be used especially on exams used to determine student levels and on exams used to move students to higher classes. On the other hand, different education levels should be addressed more in the exams with multiple-choice questions with graphics.

Learners whose mother tongue is one of the Altaic languages can be expected to be more successful in an exam style that they encounter for the first time in Turkish. Also, the learners enrolled in departments related to science are expected to be more successful on this exam. However, the eye-movements can mean that the strategic reading behavior required to be successful on an exam consisting of multiple-choice questions with graphics. Therefore, these learners should not only be guided to achieve success on the exam, but studies should be conducted to make them gain strategic reading behaviors.

The gender of learners can be expected to affect eye-movements, but a significant effect occurred only in a single eye-movement parameter. Therefore, we

can say that an understanding that can cause differences in other eye-movements should be developed for more successful reading comprehension in the context of gender.

Opting for graphic questions significantly affected some of the eye-movements. Therefore, we can argue that developing positive attitudes toward multiple-choice questions with graphics may affect both achievement and eye-movements, associated with successful reading in the case of positive attitudes.

The fact that the learners who were more successful on the test consisting of multiple-choice questions with graphics in general, apart from the previously mentioned variables, tended to exhibit fewer eye-movements at shorter durations, some of them at significant levels, can mean that a reading strategy based on using faster and fewer eye-movements should be followed to be successful on such tests. Therefore, directing unsuccessful learners to use a faster scanning behavior during multiple-choice tests containing graphics can increase their achievement on these exams. As a matter of fact, tasks based on information search and associations should be basically faster.

In this study, we regarded the homogeneous test consisting of multiple-choice graphic questions as a psychological whole. We focused on the eye-movements during the solution of the whole test under time pressure, and did not examine the eye-movements for individual questions. Future studies can investigate the eye-movements for individual questions both under time pressure and without time limits.

All the questions on the test proposed in this study had the following format: explanation paragraph, graphic and question sentence with options. Future studies can compare different eye-movements (e.g., horizontal and vertical positioning of the problem and positioning the graphic area on the top, right, or left). The most appropriate number of options in multiple-choice questions for teaching Turkish to foreigners can be determined by the eye-tracking technique. Question situations with three, four, and five options can be compared. The principles set forth regarding the preparation of multiple-choice items can be validated with the eye-tracking technique in the context of teaching Turkish as L2. For example, the situations in which the question sentence is positive and negative can be compared.

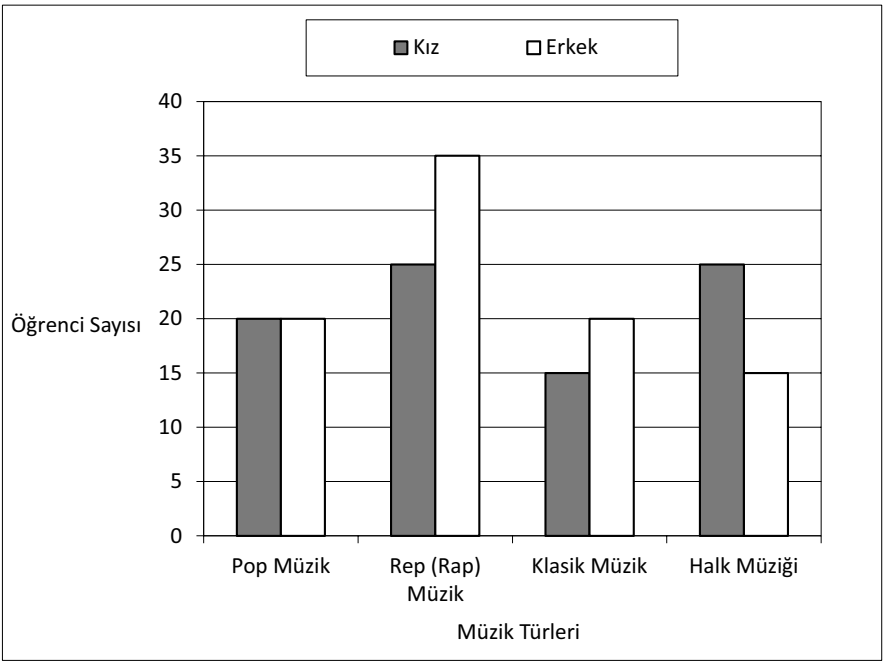
If possible, eye-tracking studies can be conducted on other language fields in teaching Turkish as L2. For example, the writing processes of learners can be examined in depth. Again, if possible, topics related to interaction such as the use of smartboards in the classroom can be studied with the eye-tracking technique.

Finally, we remind you that the results obtained from this study are limited to screen reading. Of course, it is possible that different behaviors were exhibited compared to reading from paper. As a simple example, the participants in this study did not use a pencil. Moreover, learners' attitudes toward screen reading may have also affected the findings. Therefore, we can say that it would be appropriate to conduct further studies on the subject with devices that can be attached to the eyes, while the learners solve the questions presented on paper by using a pen.

Appendix

Figures 1, 2, 3, 4, and 5

Aşağıdaki grafik, TÖMER'deki 175 öğrencinin müzik tercihlerini gösteriyor. Örneğin 20 kız ve 20 erkek öğrenci, yani toplamda 40 öğrenci pop müzik dinliyor.



Grafiğe göre bu öğrenciler, hangi müzik türlerini eşit (aynı) sayıda dinliyor?

A) Klasik Müzik-Halk Müziği

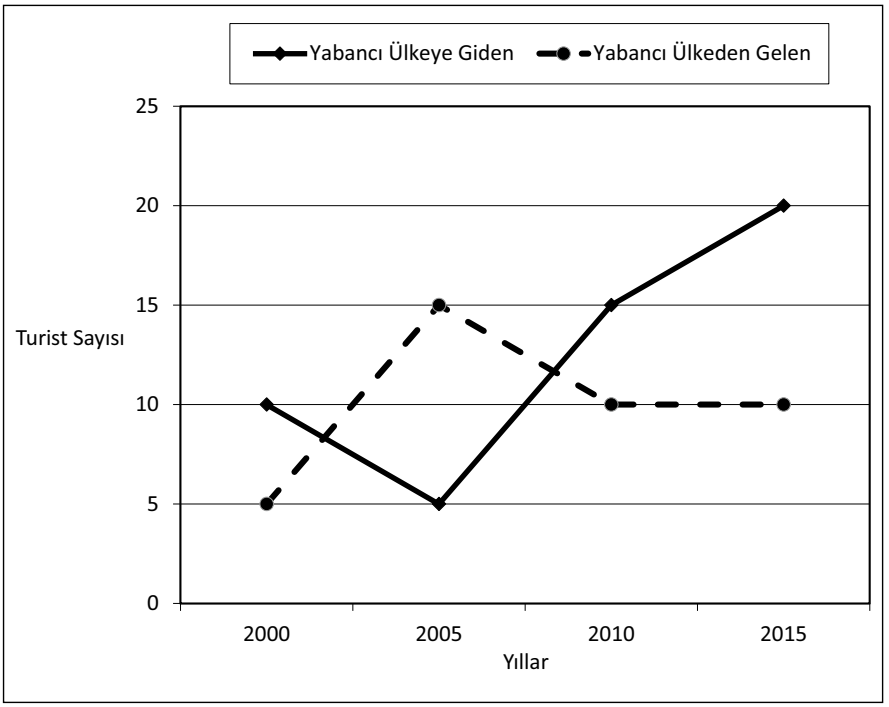
B) Rep Müzik-Klasik Müzik

C) Pop Müzik-Halk Müziği

D) Pop Müzik-Rep Müzik

Fig. 1 The first multiple-choice question that students solved

Aşağıdaki grafik, 2000-2015 arasındaki bazı yıllarda bir ülkeden yabancı ülkeye giden ve yabancı ülkeden bu ülkeye gelen turist sayılarını gösteriyor.



Yukarıdaki yazıya göre bu grafikte aşağıdaki bilgilerden hangisi yoktur?

A) Verilen yıllarda bu ülkeye hangi ülkeden en az turist geldiği

B) Verilen yıllarda yabancı ülkeye hangi yıl en az turist gittiği

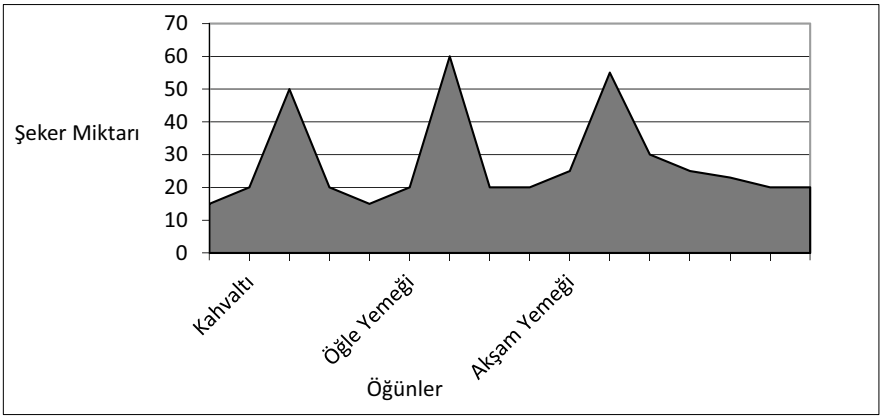
C) Verilen yıllarda bu ülkeye hangi yıl en çok turist geldiği

D) Verilen yıllarda yabancı ülkeye giden toplam turist sayısı

Fig. 2 The second multiple-choice question that students solved

Her insanın kanında bir miktar şeker vardır. Sağlıklı bir insanda, yemek yedikten sonra kandaki bu şeker miktarı yükselir. Diğer zamanlarda ise insanın kanındaki şeker miktarı genelde aynı seviyededir.

Aşağıdaki grafikte, bir insanın kanındaki şeker miktarının gün içindeki değişimi gösterilmiştir.



Buna göre aşağıdakilerden hangisi kesinlikle yanlıştır?

A) Şeker miktarı gün içinde en az üç kere yükselmiştir.

B) Şeker en çok akşam yemeğinden sonra yükselmiştir.

C) Bu kişi gün içinde üç kere yemek yemiştir.

D) Her yemekten sonra şeker yükselmiştir.

Fig. 3 The third multiple-choice question that students solved

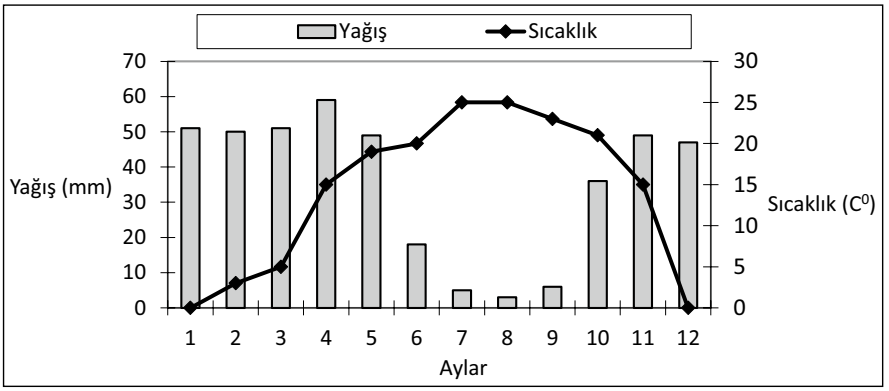
Buğday: Yetişme bölgesinde en çok ilkbahar mevsiminde yağmur yağar. Yetişme bölgesinde sıcaklık -5 ^{0}C ile 30 ^{0}C arasındadır.

Çay: Yetişme bölgesinde en çok sonbahar mevsiminde yağmur yağar. Yetişme bölgesinde sıcaklık 10 ^{0}C ile 25 ^{0}C arasındadır.

Muz: Yetişme bölgesinde en çok kış mevsiminde yağmur yağar. Yetişme bölgesinde sıcaklık 10 ^{0}C ile 30 ^{0}C arasındadır.

Kayısı: Yetişme bölgesinde en çok ilkbahar mevsiminde yağmur yağar. Yetişme bölgesinde sıcaklık 0 ^{0}C ile 25 ^{0}C arasındadır.

Aşağıdaki grafik, Türkiye'de bir yerin sıcaklık ve yağış durumunu gösteriyor.

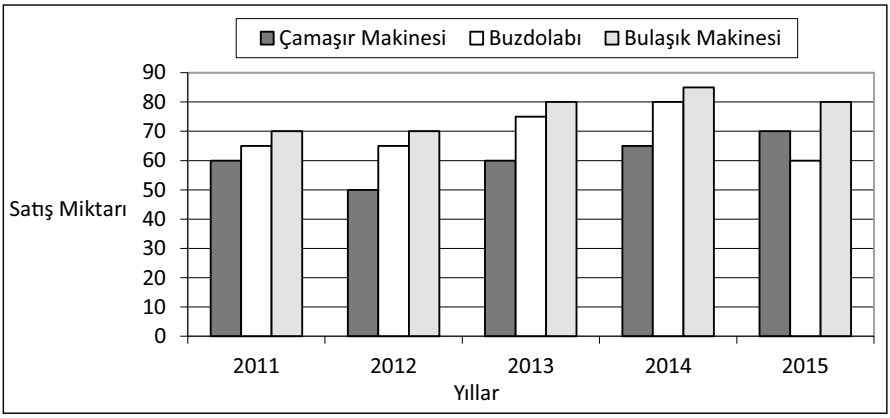


Buna göre bu yerde hangi ürünün yetişmesi <u>en uygundur</u>?

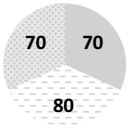
A) Buğday B) Çay C) Muz D) Kayısı

Fig. 4 The fourth multiple-choice question that students solved

Aşağıdaki grafik, bir mağazadaki eşyaların bazı yıllardaki satış miktarlarını gösteriyor.



Yukarıdaki grafikten faydalanarak aşağıdaki daire grafik çiziliyor.



Buna göre yukarıdaki daire grafik aşağıdakilerden hangisini gösteriyor?

A) 2011'deki çamaşır makinesi, buzdolabı ve bulaşık makinesi satışını

B) 2014'teki çamaşır makinesi, buzdolabı ve bulaşık makinesi satışını

C) 2012, 2013 ve 2014 yıllarındaki bulaşık makinesi satışını

D) 2011, 2012 ve 2015 yıllarındaki bulaşık makinesi satışını

Fig. 5 The fifth multiple-choice question that students solved

Acknowledgements We would like to express our gratitude to the students learning Turkish as a foreign language at Gazi University and Ankara Hacı Bayram Veli University, where we obtained the data of this study, to the managers and teachers of the language teaching centers.

Author contributions Conceptualization: [Mustafa ARMUT, Mehmet KARA], Methodology: [Mustafa ARMUT, Mehmet KARA], Formal analysis and investigation: [Mustafa ARMUT], Writing—original draft preparation: [Mustafa ARMUT]; Writing—review and editing: [Mustafa ARMUT, Mehmet KARA], Resources: [Mustafa ARMUT, Mehmet KARA], Supervision: [Mehmet KARA].

Funding The first researcher was supported by the Scientific and Technological Research Council of Turkey with a scholarship called BIDEB 2011-A General Domestic Doctorate Scholarship Program 2017/2 during his doctorate education, which also included the data collection process of this study.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abegg, M., Pianezzi, D., & Barton, J. J. (2015). A vertical asymmetry in saccades. *Journal of Eye Movement Research, 8*(5), 1–10. https://doi.org/10.16910/jemr.8.5.3

Acarturk, C., Habel, C., & Cagiltay, K. (2008). Multimodal comprehension of graphics with textual annotations: The role of graphical means relating annotations and graph lines. In J. H. J. Lee & G. Stapleton (Eds.), *Diagrammatic representation and inference: Lecture notes in computer science* (pp. 335–343). Springer. https://doi.org/10.1007/978-3-540-87730-1_30

Acarturk, C., Habel, C., Cagiltay, K., & Alacam, O. (2008). Multimodal comprehension of language and graphics: Graphs with and without annotations. *Journal of Eye Movement Research, 1*(3), 1–15. https://doi.org/10.16910/jemr.1.3.2

Acartürk, C. (2015). *Referring expressions in communication through line graphs: A comparative analysis of verbal descriptions* [Paper presentation]. Workshop on Discourse Structure in Turkic Languages, Ankara, Türkiye.

Acartürk, C. (2014). Towards a systematic understanding of graphical cues in communication through statistical graphs. *Journal of Visual Languages & Computing, 25*(2), 76–88. https://doi.org/10.1016/j.jvlc.2013.11.006

Acartürk, C., & Habel, C. (2012, July 2). *Eye tracking in multimodal comprehension of graphs* [Paper presentation]. The Workshop on Technology-Enhanced Diagrams Research. Canterbury, United Kingdom.

Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in active graphing. *International Journal of Computers for Mathematical Learning*, *5*(2), 85–114. https://discovery.ucl.ac.uk/id/eprint/1500187/1/Ainley2000Construction85.pdf

Akbarov, A., Gönen, K., & Aydoğan, H. (2018). Students' attitudes toward blended learning in EFL context. *Acta Didactica Napocensia, 11*(1), 61–68. https://doi.org/10.24193/adn.11.1.5

Al-Edaily, A., Al-Wabil, A., & Al-Ohali, Y. (2013). Interactive screening for learning difficulties: Analyzing visual patterns of reading Arabic scripts with eye tracking. In C. Stephanidis (Ed.), *International Conference on Human-Computer Interaction* (pp. 3–7). Springer.

Anıl, D., Özer-Özkan, Y., & Demir, E. (2015). *PISA 2012 araştırması ulusal nihai raporu* [*National final report of the PISA 2012 survey*]. Millî Eğitim Bakanlığı.

Arif, M. M., & Hashim, F. (2010). Young learners' second language visual literacy practices: Does gender make a difference? *International Journal of Learning, 17*(4), 503–510. https://doi.org/10.18848/1447-9494/CGP/v17i04/46975

Ay, S., & Bartan, Ö. Ş. (2012). The effect of topic interest and gender on reading test types in a second language. *Reading*, *12*(1), 62–79. https://readingmatrix.com/articles/april_2012/ay_bartan.pdf

Ayhan, E. (2019). *Yabancı dil olarak Türkçe öğretiminde okuma metinlerinin seviyelerinin tespiti*: *Okuma sırasında göz izleme bulguları* [*Benchmarking the levels of the reading passages used in teaching Turkish as a foreign language*: *Eye tracking findings during reading*] [Doctoral dissertation, Hacettepe University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Baştuğ, M., Keskin, K., & Şimşek, İ. (2019). Aloud and silent reading in eye movements: An eyetracking study. *Eskişehir Osmangazi University Social Sciences Journal, 20*, 327–337. https://doi.org/10.17494/ogusbd.548375

Bayrak-Karslı, M., Demirel, T., & Kurşun, E. (2020). Examination of different reading strategies in paragraph questions with eye tracking metrics. *Hacettepe University Journal of Education Faculty, 35*(1), 91–106. https://doi.org/10.16986/HUJE.2019051160

Berzak, Y., Nakamura, C., Flynn, S., & Katz, B. (2017). Predicting Native Language from Gaze. *The 55th Annual Meeting of the Association for Computational Linguistics*, *1*, 541–551. https://arxiv.org/pdf/1704.07398.pdf.

Biedert, R., Buscher, G., & Dengel, A. (2010). The eyebook–using eye tracking to enhance the reading experience. *Informatik-Spektrum, 33*(3), 272–281. https://doi.org/10.1007/s00287-009-0381-2

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld.

Boylu, E. (2019). *Yabancılara Türkçe öğretiminde ölçme değerlendirme uygulamaları ve standart oluşturma* [*Measurement assessment practices and standardization in teaching Turkish to foreigners*] [Doctoral dissertation, Çanakkale Onsekiz Mart University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Bruski, D. J. (2011). *Do they get the picture*?: *Visual literacy and low-literacy adult ESL learners* [Unpublished Capstone Project, Hamline University]. https://digitalcommons.hamline.edu/hse_all/468

Büyüköztürk, S., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, S., & Demirel, F. (2015). *Bilimsel araştırma yöntemleri* (Geliştirilmiş 19. Baskı). [*Scientific research methods* (Enhanced 19th ed.)]. Pegem Academy.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75–100. https://doi.org/10.1037/1076-898X.4.2.75

Choo, H., & Koh, S. (2019). *Age differences in eye movements during reading* [Paper presentation]. The 17th Conference of the Japanese Cognitive Psychology Association, Kyoto, Japan.

Christopherson, J. T. (1997). *The growing need for visual literacy at the university*. Paper presented at the Visionquest: Journeys Towards Visual Literacy 28th Annual Conference of the International Visual Literacy Association, Cheyenne, Wyoming. https://files.eric.ed.gov/fulltext/ED408963.pdf

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences, 102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Chung, I., & Kakizawa, T. (2002). Intra-individual variations in reading at different levels of interest: Analysis of eye movements of children who are hearing-impaired. *The Japanese Journal of Special Education, 39*(6), 31–39. https://doi.org/10.6033/tokkyou.39.31_3

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531–554. https://doi.org/10.2307/2288400

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453–467. https://doi.org/10.1177/0267658316637401

Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education, 92*(5), 848–867. https://doi.org/10.1002/sce.20262

Council of Europe. (2020). *Common European framework of reference for languages*: *Learning*, *teaching*, *assessment companion volume with new descriptors*. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Dehaene, S. (2017). *Reading in the brain*: *The new science of how we read*. (2nd Ed., O. Karakaş, Trans.). [*Beyin nasıl okur? Okumanın bilimi ve evrimi* (2. Basım)]. Alfa Bilim.

Dolgunsöz, E. (2015). Measuring attention in second language reading using eye-tracking: The case of the noticing hypothesis. *Journal of Eye Movement Research, 8*(5), 1–18. https://doi.org/10.16910/jemr.8.5.4

Duchowski, A. T. (2007). *Eye tracking methodology theory and practise* (2nd ed.). Springer.

Duffy, S. A. (1992). Eye movements and complex comprehension processes. In K. Rayner (Ed.), *Eye movements and visual cognition scene perception and reading* (pp. 462–471). Springer-Verlag.

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30*, 149–166. https://doi.org/10.1017/S026719051000005X

Escalante-Talavera, J. M. (2018). *ESL students' reading behaviors on multiple-choice items at differing proficiency levels: An eye-tracking study* [Master's thesis, Brigham Young University]. BYU Scholars Archive. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=8424&context=etd

Gedik, E. (2017). *Yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirme* [*An assessment and evaluation in teaching Turkish as a foreign language*] [Unpublished master's thesis, Istanbul Arel University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Goldberg, J. H., & Helfman, J. I. (2010). *Comparing information graphics: A critical look at eye tracking*. Proceedings of the 3rd BELIV'10 Workshop: Beyond time and errors: Novel evaluation methods for information visualization, 71–78. https://doi.org/10.1145/2110192.2110203

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization, 10*(3), 182–195. https://doi.org/10.1177/1473871611406623

Gültekin, C. (2009). *Ortaöğretim 9. sınıf öğrencilerinin çözeltiler ve özellikleri konusu ile ilgili grafik çizme okuma ve yorumlama becerilerinin incelenmesi* [*Examining 9th grade students' abilities on drawing reading and interpreting of graphs about solutions and their properties*] [Master's thesis, Balıkesir University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Gültekin, C. (2014). *Ortaöğretim öğrencileri ile üniversite öğrencilerinin hal değişimi, çözeltiler ve çözünürlük konuları ile ilgili grafik çizme okuma ve yorumlama becerilerinin karşılaştırılması* [*Comparison of abilities on drawing, reading and interpreting of graphs of the secondary education students and university students in change of state, solutions and solubility subjects*] [Doctoral dissertation, Balıkesir University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Habel, C., & Acartürk, C. (2007). *On reciprocal improvement in multimodal generation: Coreference by text and information graphics* [Paper presentation]. MOG 2007: The Workshop on Multimodal Output Generation, Aberdeen, United Kingdom.

Habel, C., & Acartürk, C. (2009). *Eye-tracking evidence for multimodal language-graphics comprehension: The role of integrated conceptual representations* [Paper presentation]. NODALIDA: The Workshop on Multimodal Communication - from Human Behaviour to Computational Models, Odense, Denmark.

Habel, C., & Acartürk, C. (2011). Causal inference in graph-text constellations: Designing verbally annotated graphs. *Tsinghua Science and Technology, 16*(1), 7–12. https://doi.org/10.1016/S1007-0214(11)70002-5

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge. https://psycnet.apa.org/record/2004-95198-000

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5

Hallowell, B., & Heuer, S. (2005). *Eye tracking evaluation of Clipart versus carefully designed images in multiple-choice image displays for comprehension assessment* [Paper presentation]. Clinical Aphasiology Conference, Sanibel Island, USA.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary Educational Psychology, 24*(2), 95–123. https://doi.org/10.1006/ceps.1998.0987

Harada, F. K. (1988). *The effect of different orthoraphical presenrtation of a text upon the reading behaviours of native and non-native readers of Japanese: An eye-tracking study.* [Doctoral dissertation, Ohio State University]. ProQuest. https://www.proquest.com/docview/303582891?pq-origsite=gscholar&fromopenview=true

Hayashi, H., Maeno, T., & Kunifuji, S. (2002). *Eye movement tracking to support the second language learners' self-learning in comprehension of English texts* [Paper presentation]. The IEEE International Conference on Systems, Man and Cybernetics, Yasmine Hammamet, Tunisia.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University.

Işıkoğlu, M. (2015). Yabancı dil olarak Türkçe öğretiminde kullanılan yeterlik sınavlarının madde yazımı bakımından incelenmesi (Mersin ve Sakarya üniversiteleri örneği) [*Analysis of proficiency exams developed for teaching Turkish as a foreign language in terms of item writing (Samples of Mersin and Sakarya Universities)*] [Master's thesis, Atatürk University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Jian, Y. C. (2017). Eye-movement patterns and reader characteristics of students with good and poor performance when reading scientific text with diagrams. *Reading and Writing, 30*(7), 1447–1472. https://doi.org/10.1007/s11145-017-9732-6

Jian, Y. C., & Wu, C. J. (2015). Using eye tracking to investigate semantic and spatial representations of scientific diagrams during text-diagram integration. *Journal of Science Education and Technology, 24*(1), 43–55. https://doi.org/10.1007/s10956-014-9519-3

Jiang, X. (2012). Effects of discourse structure graphic organizer on EFL reading comprehension. *Reading in a Foreign Language, 24*(1), 84–105. http://hdl.handle.net/10125/66669

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://psycnet.apa.org/doi/10.1037/0033-295X.87.4.329

Karaoğlan-Yılmaz, F. G., & Yılmaz, R. (2019). *Examining the questions asked in the 2018 KPSS educational sciences exam within the scope of instructional technology and material design by eye tracking method* [Paper presentation]. 3rd International Congress on Science and Education, Afyon, Türkiye.

Kıranda, S., & Akpınar, M. (2020). Students' views about difficulties they are experienced with graphic reading and drawing. *Hacettepe University Journal of Education, 35*(2), 415–427. https://doi.org/10.16986/HUJE.2019050634

Kim, S., & Lombardino, L. J. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research, 8*(3), 1–17. https://doi.org/10.16910/jemr.8.3.2

Klein, P., Becker, S., Küchemann, S., & Kuhn, J. (2021). Test of understanding graphs in kinematics: Item objectives confirmed by clustering eye movement transitions. *Physical Review Physics Education Research, 17*(1), 1–6. https://doi.org/10.1103/PhysRevPhysEducRes.17.013102

Körner, C. (2004). Sequential processing in comprehension of hierarchical graphs. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 18*(4), 467–480. https://doi.org/10.1002/acp.997

Körner, C. (2005). Concepts and misconceptions in comprehension of hierarchical graphs. *Learning and Instruction, 15*(4), 281–296. https://doi.org/10.1016/j.learninstruc.2005.07.003

Lee, H., & Choi, W. (2019). Predictability effects modulated by age during sentence reading: An eye-tracking study. *The Korean Journal of Cognitive and Biological Psychology, 31*(1), 17–38. https://doi.org/10.22172/cogbio.2019.31.1.002

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Lim, H. J. (2017). Exploring test takers' cognition in a high-stakes reading test: An eyetracking study. *Journal of Asia TEFL, 14*(3), 482–500. https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE08971069

Lindner, M. A., Eitel, A., Thoma, G. B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision-making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology, 28*(5), 738–752. https://doi.org/10.1002/acp.3060

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction, 8*(4), 353–388. https://doi.org/10.1207/s15327051hci0804_3

Maden, S., & Altunbay, M. (2016). The use of graphs and tables as tools of visual presentation and visual reading in Turkish language education. *International Journal of Turkish Literature Culture Education, 5*(4), 1971–1983. https://doi.org/10.7884/teke.671

Mason, L., Pluchino, P., & Tornatora, M. C. (2013a). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly, 48*(2), 199–214. https://doi.org/10.1002/rrq.41

Mason, L., Tornatora, M. C., & Pluchino, P. (2013b). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education, 60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mason, L., Tornatora, M. C., & Pluchino, P. (2015). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: An eye-movement study. *Reading and Writing, 28*(6), 851–872. https://doi.org/10.1007/s11145-015-9552-5

McKenna, P. (2019). Multiple choice questions: Answering correctly and knowing the answer. *Interactive Technology and Smart Education, 16*(1), 59–73. https://doi.org/10.1108/ITSE-09-2018-0071

Olmsted-Hawala, E., Holand, T., & Quach, V. (2014). Usability testing. In J. Romano Bergstrom & A. J. Schall (Eds.), *Eye tracking in user experience design* (pp. 49–80). Elsevier.

Oruç, S., & Akgün, İ. H. (2010). The level of gaining graphic reading skills of primary school social studies 7th grade students. *International Journal of Eurasian Social Sciences, 1*(1), 51–58. https://dergipark.org.tr/en/pub/ijoess/issue/8541/105988

Özçelik, D. A. (2010). *Test hazırlama kılavuzu* (4. Baskı). [*Test preparation guide* (4th Ed.)]. Pegem Academy.

Özer, E. (2019). *Yetkin ve zayıf okurların okuma becerileri ve göz hareketleri arasındaki ilişkinin karşılaştırılarak incelenmesi* [*Comparative examination of the relationship between reading skills and eye movements of good and poor readers*] [Doctoral dissertation, Gazi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Özer, E., Özdemir, S., & Kara, M. (2020). Examining reading skills with eye tracking technique. *Journal of Turkish Educational Sciences, 18*(1), 437–455. https://doi.org/10.37217/tebd.714158

Parodi, G., Julio, C., & Recio, I. (2018). When words and graphs move the eyes: The processing of multimodal causal relations. *Journal of Eye Movement Research, 11*(1), 1–18. https://doi.org/10.16910/jemr.11.1.5

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Erlbaum. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8b751ec40e1a5721a6e745eaaacccdbde79ed977

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acqusition. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467–488). Cambridge University.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

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

Reed, S. K. (2010). *Thinking visually*. Psychology Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Sajka, M. (2017). *Visual attention while reading a multiple choice task by academics and students: A comparative eye-tracking approach* [Paper presentation]. CERME 10 European Society for Research in Mathematics Education, Dublin, Ireland.

Sajka, M., & Rosiek, R. (2015). *Discussion of visual attention for solving multiple choice science problem: An eye-tracking analysis* [Paper presentation]. The 1st International Baltic Symposium on Science and Technology Education (BalticSTE2015), Šiauliai, Lithuania.

Schall, A. J., & Romano Bergstrom, J. (2014). Introduction to eye tracking. In J. Romano Bergstrom & A. J. Schall (Eds.), *Eye tracking in user experience design* (pp. 3–26). Elsevier.

Schiff, M. (1998). *Designing graphic presentations from first principles.* [Doctoral dissertation, University of California]. ProQuest. https://www.proquest.com/docview/304422988?pq-origsite=gscholar&fromopenview=true

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology, 106*(4), 974–989. https://doi.org/10.1037/a0037054

Sezgin-Memnun, D. (2013). Examining of line graphic reading and drawing skills of secondary school seventh grade students. *Turkish Studies-International Periodical for the Languages, Literature and History of Turkish or Turkic, 8*(12), 1153–1167. https://doi.org/10.7827/TurkishStudies.6026

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association, 82*(398), 454–465. https://doi.org/10.2307/2289447

Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics, 30*(4), 353–368. https://doi.org/10.3102/10769986030004353

Stoerger, S. (2018). Writing without words: Designing for a visual learning experience. *Education for Information, 34*(1), 7–13. https://doi.org/10.3233/EFI-189002

Strobel, B., Saß, S., Lindner, M. A., & Köller, O. (2016). Do graph readers prefer the graph type most suited to a given task? Insights from eye tracking. *Journal of Eye Movement Research, 9*(4), 1–15. https://doi.org/10.16910/jemr.9.4.4

Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research & Method in Education, 29*(2), 185–208. https://doi.org/10.1080/17437270600891614

Tang, H., & Pienta, N. (2012). Eye-tracking study of complexity in gas law problems. *Journal of Chemical Education, 89*(8), 988–994. https://doi.org/10.1021/ed200644k

Tepgeç, M., & Seferoğlu, S. S. (2019). *A content analysis study on the use of eye tracking method in the evaluation of learning-teaching processes* [Paper presentation]. 6th International Eurasian Educational Research Congress, Ankara, Türkiye.

Tobii studio User's Manual Version 3.4.8. (2017). Tobii AB.

Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2011). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education, 58*(1), 375–385. https://doi.org/10.1016/j.compedu.2011.07.012

Tufte, E. R. (2006). *The visual diplay of quantitative information* (2nd ed.). Graphics.

Tuninetti, A., Warren, T., & Tokowicz, N. (2015). Cue strength in second-language processing: An eye-tracking study. *The Quarterly Journal of Experimental Psychology, 68*(3), 568–584. https://doi.org/10.1080/17470218.2014.961934

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eyemovement recordings in second language research. *Studies in Second Language Acquisition, 35*(2), 205–212. https://doi.org/10.1017/S027226311200085X

Wong, W., Barthels, M., & Chrobot, N. (2014). E-commerce websites. In J. Romano Bergstrom & A. J. Schall (Eds.), *Eye tracking in user experience design* (pp. 187–216). Elsevier.

Yurtman-Kaçar, H. (2018). *Validating multiple-text reading tasks in foreign language proficiency tests through verbal protocols and eye tracking* [Master's thesis, Boğaziçi University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Zhan, Z., Wu, J., Mei, H., Wu, Q., & Fong, P. S. (2020). Individual difference on reading ability tested by eye-tracking: From perspective of gender. *Interactive Technology and Smart Education, 17*(3), 267–283. https://doi.org/10.1108/ITSE-12-2019-0082

Authors and Affiliations

Mustafa Armut[1] · **Mehmet Kara**[2]

✉ Mustafa Armut
mustafaarmut@ahievran.edu.tr

Mehmet Kara
mehkara@gazi.edu.tr

[1] Faculty of Education, Kırşehir Ahi Evran University, Bağbaşı, Şehit Necdet Yağız Street No: 143 / E Central District, Campus of Bağbaşı, Kırşehir, Turkey

[2] Gazi University Gazi Faculty of Education Emniyet Neighbourhood, Bandırma Street, Hersek No: 6/32 06500 Beşevler, Yenimahalle, Ankara, Turkey