



# Investigation of Eighth-Grade Students' processes of solving skill- based science questions by eye tracking technique

Şeyma Özdemir<sup>1</sup> · Cemal Tosun<sup>2</sup>

Received: 12 December 2023 / Accepted: 4 June 2024  
© The Author(s) 2024

## Abstract

The aim of this study was to determine the visual measurement results related to the behavior/processes of solving skill-based science questions of eighth grade students by eye tracking technique. Non-experimental quantitative research method was used in the research and visual measurement results were supported by heat maps and eye splash movements. Nine questions, with difficulty and discrimination indexes calculated, were applied to 56 eighth grade students. Data were collected with a skill-based test, eye tracking device, think-aloud protocols and a perception scale towards next generation science questions. The data obtained from the eye tracking device was examined using Gaze Viewer software and the results were shown as images. The results revealed that visual measurement results differed according to gender and 2023 high schools entrance exam scores. Additionally, negative relationships were found between visual measurement results and students' practice test scores and their perceptions towards solving next generation science questions. It was determined that average duration of fixation had a significant predictive effect on students' self-efficacy levels for solving next generation science questions.

**Keywords** Eye tracking technique · Perception · Skill-based questions and think aloud protocols

---

✉ Cemal Tosun  
ctosun@bartin.edu.tr

Şeyma Özdemir  
seymaoz9627@gmail.com

<sup>1</sup> Karapınar Abdullah Güngüpoğlu Middle School, Zonguldak, Turkey

<sup>2</sup> Faculty of Education, Department of Science Education, Bartın University, Bartın, Turkey

## 1 Introduction

There are different applications such as central exams, school-based entrance exams, placement with school success scores, address-based placement during the transition to secondary education in the world. In the United States, the transition to secondary education is based on address, while central entrance exams are applied in some states (Gür et al., 2013). In the Netherlands, students may only transition to secondary education through a central exam (Aykaç & Atar, 2014). In Italy and Russia, students are placed in academic and vocational education schools at the high school level with the scores they receive from the middle school graduation exam. In Japan and Hungary, students are placed in academic and vocational education schools at the high school level with central entrance exam, school-based entrance exam and school scores. In Finland, one of the implementers of a successful education system, only school success scores are taken into account when placing students in secondary education, and there is no central exam (Eurypedia, 2013).

Various central exams are applied in Turkey to monitor, measure and evaluate the effectiveness of education systems and the knowledge and skills of students. A central exam called high school entrance exam has been applied since the 2017–2018 academic year, and the purpose of this exam is to determine the acquisition levels of students' knowledge and skills in the eighth grade curriculum. The main purpose of high school entrance exam in Turkey is to distinguish students with high-level cognitive skills and place these students in qualified schools such as science high school, social science high school, project schools (Yalçın, 2019).

Central exams and international large-scale exams (i.e.: TIMSS) have been implemented for the last few decades to determine whether students of a certain age group or grade level can acquire the expected knowledge and skills. In the twenty-first century, the nature of the knowledge and skills that students must acquire has changed. This change has brought about a change in the content of national central exams and international large-scale exams. In recent years, it observed that there is a tendency to measure high-level thinking skills with these exams (i.e.: national central exams such as high school entrance exam in Turkey and international large-scale exams such as TIMSS).

### 1.1 Skill-based science questions

It is difficult to explain the term high-level thinking skills with a single definition. According to Zoller (2000), these skills include the steps of asking questions, solving problems, analyzing variables, synthesizing new ideas and products, and evaluating the process. These skills begin with observation and include the skills of presenting the problem, developing a hypothesis, testing the hypothesis, and reaching generalizations from the results. High-level thinking skills refer to Bloom's Taxonomy, critical thinking, creative thinking, problem solving, decision making and metacognitive thinking skills (Yen & Halili, 2015).

The questions that measure high-level cognitive skills are called life-based, next-generation or skill-based questions in the relevant literature (Çepni, 2019; Şan & İlhan, 2022). In this study, the questions that measure high-level cognitive skills were called skill-based questions. Miller et al. (2009) defined skill-based questions as interpretation questions that contain written and visual elements such as graphics, tables, texts, maps, pictures, shapes, diagrams and require high-level skills such as analysis, interpretation, problem solving, reasoning, reading comprehension.

Science is one of the fields where students are tested whether they have acquired these skills through skill-based questions in national central exams and international large-scale exams. The main purpose of science education is to educate a science literate individual. Science aims to provide students with mental process skills that will enable them to solve daily life problems (Regis et al., 1996). One of the main components of science literacy is to acquire the scientific process skills necessary at the stage of solving daily life problems (Hodson, 1992). Scientific process skills are one of the main goals of science education (Germann et al., 1996). Science process skills are very important for teaching science (Myers et al., 2004) and are used in decision-making processes (National research council -NRC, 1996). Scientific process skills are classified as basic and high-level scientific process skills. Çepni et al. (1997) divide these skills into three as basic (making observations, measuring, classifying, recording data, number and space relations and communicating), causal (prediction, determining variables, operational identification and inference) and experimental (making hypotheses, setting up experiments, changing and controlling variables, using data and building models, and making decisions) scientific process skills. In the current research, skill-based science questions asked in high school entrance exams in Turkey were categorized according to the classification of Çepni et al. (1997) and the processes of solving these questions of eighth grade students were focused.

This research revealed visual measurement results of students' processes/behaviors in solving skill-based science questions. According to our research, no research has been found in the literature that reveals where students have difficulty in solving skill-based science questions using the eye tracking technique. Students' question-solving processes/behaviors were supported by heat maps and eye splash movements, providing concrete evidence to the literature. Additionally, students are not familiar with the skill-based science questions and may have prejudices about the questions (Yiğit et al., 2022). Student perceptions have an impact on course success. Therefore, it is important to reveal the relationship between visual measurement results and student perceptions of skill-based science questions.

## 1.2 Eye tracking technique

The study of understanding human cognitive processes is always one of the research topics of science. For this purpose, various technological tools developed and continue to be developed. One of them is eye tracking devices. The technique of monitoring eye movements known for a century (Özdoğan, 2008) and four periods are important in the historical development of this technique (Özer & Özdemir,

2022). These periods are: the early years, the period of pause due to the behavioral approach, the period when it is on the rise again, and the golden age. Javal's mention of the French concept of saccade in his study published in 1879 was a big step in terms of eye tracking research (Wade et al., 2003). An eye tracking device was developed by Huey in these years (Reed & Meyer, 2007), and after his experiments, he reached important evidence supporting Javal's views on the saccade. The efforts of this technique to explain cognitive processes were overshadowed by the behavioral approach in the mid-1910s and came to a standstill. Tinker (1958) states that what can be learned about reading from eye movements has been learned and future studies are not promising. According to Rayner, the 1970s were a period in which investigation of cognitive processes through eye tracking gained momentum (Rayner, 1998). The 2000s, when eye tracking devices were produced that provide access to more useful and reliable data with technological advances, are considered the golden age of this technique (Özer & Özdemir, 2022). For the last 10 years, research has been carried out using eye tracking devices with high sampling characteristics such as 1000–2000 Hz and highly reliable data has been obtained (Özer & Özdemir, 2022).

Today, with modern eye tracking devices, detailed information is obtained using eye-movement parameters related to the temporal, spatial and count of fixation and saccade eye movements (Lai et al., 2013). In the use of this technique, the area of interest is first determined. The area of interest may be a word, sentence, paragraph, question-answer, graphic or figure according to the purpose of the study. Visual measurement results of eye-movement parameters such as gaze duration, first fixation duration or saccade duration may be defined for the area of interest (Lai et al., 2013). The gaze duration refers to the total duration of the fixations (Jacob & Karn, 2003) and the first fixation duration refers to the duration until the first focus (Liversedge et al., 1998). Saccade duration is defined as the total saccade duration spent in the area of interest (Lai et al., 2013).

### 1.2.1 Eye tracking technique in educational research

In recent years, studies using the eye tracking technique have increased with the developing technology. The eye tracking technique, used in health, architecture, engineering fields, has attracted the attention of educators and psychologists in different parts of the world. The eye tracking technique is successfully applied in educational research on human-computer interactions (Jacob & Karn, 2003), information processing (Radach & Kennedy, 2004), arithmetic problem solving (Malcı, 2021; Özdemir, 2013; Verschaffel et al., 1992) and examination of reading skills (Rayner et al., 2006). Verschaffel et al. (1992) examined the processes of solving mathematical problems using eye tracking technique and suggested that it is necessary to recognize, select and process relevant information in order to successfully solve mathematical problems. Özdemir (2013) recorded students' problem-solving processes on three math problems of similar difficulty with an eye tracking device. Similarly, Malcı (2021) recorded students' problem-solving processes on six geometry questions containing a short explanation and a diagram with an eye-tracking device. Özdemir (2013) reported that eye movements

are important for the management of attention resources in the problem-solving process, and Malcı (2021) reported that students use different strategies in solving different problems. In the study focusing on reading comprehension processes, researchers used participants' eye movement data and reported that processing time increased in difficult text (Rayner et al., 2006). He reported that when there is an inconsistency in the text, readers focus more on that area. Jacob and Karn (2003) investigated how researchers can apply eye tracking technique to human-computer interaction issues as the personal computer increases and they reported that this technique was useful for answering questions about how commands are searched in computer menus. With this technique, computer interfaces and the usability of the computer as an interaction tool were also investigated. After the 1990s, technological developments such as the internet, e-mail and video conferencing turned into information sharing tools, and researchers investigated their availability using eye tracking techniques (Cowen et al., 2002). Additionally, the usefulness of educational software was examined using the eye tracking technique (e.g.: Avcı, 2010; Tonbuloğlu, 2010). The researchers randomly selected one of the educational software used in fifth grade science lesson (Avcı, 2010) and seventh grade mathematics lesson (Tonbuloğlu, 2010) and they distributed various task cards to the students and asked the students to perform these tasks respectively. According to the results, the usefulness of the teaching software was evaluated, user problems were revealed and various suggestions were made to improve the software.

### 1.2.2 Eye tracking technique in science education

Studies using the eye tracking technique in science education focused on the following topics: (i) whether the participants' science problem solving behaviors differ according to their cognitive levels, (ii) the effect of the problem or text consisting of images/graphics on their behavior, (iii) The effect of the experimental (visually supported) on some dependent variables.

Some studies examined students' behavior in solving science problems according to their cognitive levels (Rodemer et al., 2020; Tsai et al., 2012). Tai et al. (2006) investigated where students with physics, chemistry and biology backgrounds focused while solving science problems using the eye tracking technique. Rodemer et al. (2020) examined undergraduate chemistry students' behavior in visual decoding of problems using eye tracking technique and reported that students in the advanced class were faster in their decision-making processes. In another study, the researchers examined the visual attention of students solving a multiple-choice science problem using eye tracking technique and it was found that students with insufficient pre-knowledge focused on more specific features, while students with sufficient pre-knowledge focused on more thematic content (Tsai et al., 2012).

One of the study topics in which the eye tracking technique was used in science education was researches examining the effect of a problem or text consisting of different images/graphs on the student's visual attention and behavior. Researchers monitored students' visual behavior while reading a text on the relationship between the greenhouse effect and global climate change (Ho et al., 2014). They reported

that students spent more time reading textual information than graphical information. In another study, researchers investigated the impact of visuals on sixth grade students' understanding of science texts (Wu et al., 2021). Results showed that there were no significant differences between the illustrated/non-illustrated text groups in terms of reading comprehension and total duration of fixation. Teo and Peh (2023) examined the eye movements of doctoral (expert) and undergraduate (novice) students while solving five multiple-choice science questions and it was reported that experts focused more on the question stem and novices focused more on the graph.

Another study topic in which the eye tracking technique was used in science education was the researches examining the effect of the experimental on some dependent variables. Researchers examined the effects of animations and visual feedback on micro-scale explanations of experimental events was examined using eye tracking technique and they found increases in chemical imaging and critique of relevant features after-experimental (Hansen et al., 2019). Researchers examined the impact of simulation or screenshots on students' conceptual understanding of collision theory and it was determined that the screenshot and simulation met different cognitive demands (Sweeder et al., 2019). Jian (2021) investigated the effect of reading and practicing easy and difficult science articles on students' learning outcomes using the eye tracking technique and it was found that the experimental benefited students who read the difficult article.

In sum, studies using eye tracking techniques in science education are limited. In the current research, we are interested in how students solve skill-based science questions. The current research is intended to expand research on solving science problems in several aspects. First, we test how students solve skill-based questions on matter and industry. Second, we examine how students solve questions at different skill levels.

### 1.3 Theoretical framework

The eye-mind hypothesis is often mentioned in studies using the eye tracking technique in educational research (e.g.: Sprenger & Benz, 2020). Due to differences in the cognitive structures examined and the focus of research, there was no standard theoretical framework for studies using the eye tracking technique in educational research (Ke et al., 2024). Ke et al. (2024) systematically examined studies using the eye tracking technique in educational research and stated that the theoretical foundations were not stated in most of the articles. The theories and models referred to in the limited number of studies were divided into three categories. These are: cognitive learning theories, attention theories and other psychological theories.

With the current study, we aim to analyze in depth students' solving processes of science questions at different skill levels. Skill-based questions generally contain written and visual elements such as graphs, tables, text, maps, pictures, figures and diagrams (Miller et al., 2009). The logic of the skill-based questions used in this research is based on the cognitive theory of multimedia learning (Pellicer-Sánchez et al., 2020), cognitive load theory (Clinton et al., 2017) and dual coding theory. Paivio's (1986) dual coding theory explains that text-diagram representations are

stored in different cognitive systems. Questions in which image and text are questioned together contain more cognitive details than the information provided by text or visual alone. In the current study, we aimed to reveal students' levels of learning cognitive details by supporting skill-based science questions with visuals.

The basic idea of the cognitive theory of multimedia learning and the cognitive load theory is based on the limitation of the processing of information presented to the auditory, linguistic and visual sensory processing channels. The cognitive theory of multimedia learning argues that visual and verbal information (text or speech) are processed in different ways, and the theory emphasizes the need to integrate this information (Mayer, 2014). Reading comprehension may be improved by integrating text and pictures (Levie & Lentz, 1982). According to the cognitive load theory, the amount of information that a person can process on working memory at a given time is limited (Chandler & Sweller, 1991). If a student has too much information to process or the information is difficult to understand, their limited working memory capacity may be overloaded, resulting in poor comprehension and reduced learning (Sweller, 1994). Cognitive load may be specific to the learning goal or external (Sweller et al., 2011). External cognitive load consists of information in working memory that is unrelated to the instructional task and it is necessary to reduce the external cognitive load so that students can understand the course content (Mayer & Moreno, 2003). The skill-based questions used in the current study were created by integrating text and visuals. Some of the questions were fictionalized from situations encountered in everyday life. The cognitive load of solving some questions was low and some was high. The solution of some questions had the potential to increase the external cognitive load.

#### **1.4 The present study and research questions**

The increasing young population in Turkey, the low number of qualified schools and the fact that these qualified schools accept students with scores have made it mandatory for the central exam system. Due to these reasons, students who have passed from primary to secondary education participated in various exams. The question styles of these exams have changed over time. The content of today's high school entrance exam consisted of questions that measured high-level cognitive skills such as applying, analyzing, synthesizing and evaluating (Çakır, 2019). MoNE's first reports stated that most students did not answer these questions (MoNE, 2018, 2019). According to the results of TIMSS 2019, Turkey's science literacy score has increased (MoNE, 2020). The emphasis on the role of knowledge in daily life in science curricula and including skill-based questions in central exams (high school entrance exam) contributed to this success. For this reason, in the current study, we aim to provide an in-depth analysis of students' solving processes of science questions at different skill levels, which are asked to measure students' high-level cognitive skill levels in national central exams and international large-scale exams (e.g. TIMSS). Revealing students' processes of solving skill-based science questions will contribute to teachers' detailed understanding of why students have difficulty in these questions and how much time and cognitive effort they spend. It will provide



teachers with the opportunity to eliminate the knowledge and skill deficiencies necessary to solve the problem that their students are struggling with. The results will provide an opportunity for revision by drawing attention to skill-based question preparers, the issues that students have difficulty understanding and the cognitive load of the problem. This research has the potential to make significant contributions to the relevant literature, as it has demonstrated with concrete evidence that visual measurement results should be taken into account in the processes of developing skill-based tests and solving test questions. Item analysis of skill-based questions is generally done according to Classical Test Theory-CTT (Özer-Özkan, 2014). Item analyzes based on lower and upper groups and correlation analyzes based on item-total correlation are performed. The purpose of these analyzes is to reveal whether the test questions distinguish between those who know and those who do not know. In addition, the difficulty and discrimination levels of the questions and the preference of the options are determined by these analyzes. However, CTT cannot reveal where a student focuses most while solving a question, where she spends the most time, where she has difficulty reading questions with long paragraphs and images, and what kind of behavior she exhibits when eliminating distractors. This study revealed visual measurement results of students' behavior towards solving skill-based science questions. In this study, students' processes of solving skill-based science questions were recorded using some eye-movement parameters such as, time to first fixation, duration of first fixation, number of fixations, average duration of fixation, total duration of fixation, number of visits, average duration of visit and total duration of visit.

There were many studies in the literature revealing the opinions of teachers, students and parents about high school entrance exams (i.e: Kızıkan & Nacaroğlu, 2019; Erden, 2020). There were studies on the use of eye tracking technique in educational research (i.e.: Jian, 2021; Tsai et al., 2012). However, no study was found that revealed visual measurement results of eighth grade students' processes/behaviors towards solving skill-based science questions. In this respect, it is thought that the research will make significant contributions to the relevant literature. Students are not familiar with the skill-based science questions and may have prejudices about the questions (Yiğit et al., 2022). Student perceptions have an impact on course success. Therefore, another focus of the research was to reveal the relationship between visual measurement results and student' perceptions of skill-based science questions. In previous studies, students' science problem solving behaviors were examined according to the participants' grade level (Rodemer et al., 2020) and prior knowledge levels (Tsai et al., 2012). In this research, the processes of solving skill-based science questions of successful and unsuccessful students were examined.

In the context of the importance of the above-mentioned research, the aim of this study was to reveal the visual measurement results of eight grade students' behavior/processes in solving skill-based science questions by eye tracking technique. The four research questions are as follows:



## Research questions.

1. What were the visual measurement results of eighth grade students' processes in solving skill-based science questions?
2. Did eighth grade students' the fixation and visit duration or number on skill-based science questions differ according to gender and 2023 high schools entrance exam scores?
3. Was there a statistical relationship between students' practice test scores and their perceptions towards solving new generation science questions and visual measurement results?
4. What was the predictive effect of visual measurement results on practice test scores and students' perceptions towards solving new generation science questions?

## 2 Methodology

Non-experimental quantitative research methods were used in the research. Non-experimental research methods describe things that have occurred and examine relationships between things without any direct manipulation of conditions (McMillan & Schumacher, 2006, p.24). Descriptive research design was used for the first research question, comparative research design was used for the second research question, and correlational research designs were used for the third and fourth research questions. Descriptive research is used to summarize the current or past status of something. Comparative research examines the differences between two or more groups on a variable while correlational research examines the relationship between two or more variables without interfering with the variables (McMillan & Schumacher, 2006, p.219). Visual measurement results revealed by non-experimental research methods were supported by heat maps and eye splash movements. The research methods are presented in Fig. 1.

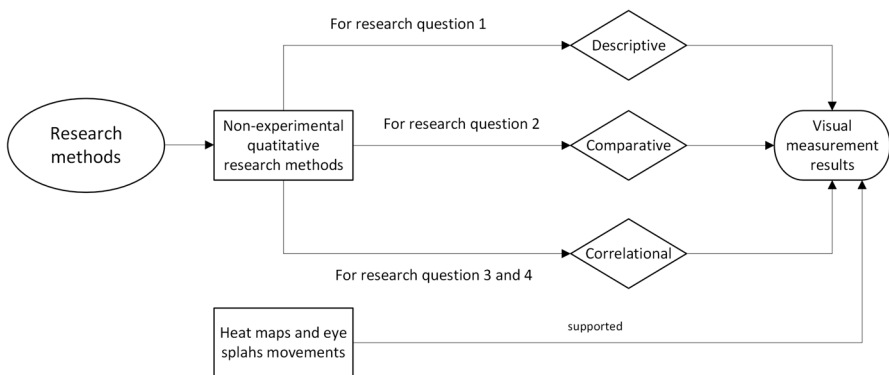


Fig. 1 Research methods

## 2.1 Research participants

Fifty-six volunteer students participated in the eye tracking application. These students were eighth graders. 46.4% of the participants were female and 53.6% were male students. The students were studying at four different schools. These schools were located within the borders of a province in the north-west of Turkey. These four middle schools were in the center of the city and there were approximately 700 students in the eighth grade of these schools. It was approximately one third of the 8th grade students studying in this province. Convenience and purposive sampling techniques were used for participant selection. Limitations in terms of time and labor were taken into account and the participants were selected from easily accessible and applicable units. Additionally, participants who were rich in information in the context of the purpose of the study were selected for in-depth research. According to Patton (1990), there are different strategies regarding purposive sampling. While determining the participants, maximum diversity strategy, one of the purposive sampling methods, was used. The first criterion in the selection of participants was their high school entrance exam performance. For this, participants were divided into three categories according to their high school's entrance exam scores and participants were selected from very good, good and average performers. The other selection criterion was willingness to participate in the study. Another criterion was the approval of the parents. The application was carried out in the Human Computer Interaction Laboratory of a university. This laboratory was approximately 10 km from the city center and parental permission was required for students to come to the university. Ethics committee approval and legal approvals were received. Additionally, voluntary participation consent was obtained from them and their parents.

## 2.2 Data collection tools

### 2.2.1 Skill-based science test

A test containing skill-based science questions was prepared. This test was within the scope of the "Matter and Industry" unit. It is a unit for the fall semester of eighth grade. Many questions are asked from this unit in the high school's entrance exams. Initially, 91 questions were determined for the item pool. Twenty-five of these questions were asked in the high school's entrance exams between 2018–2022. Sixty-six of the questions were sample questions published by MoNE. Researchers evaluated these questions according to paragraph length, presence of unfamiliar vocabulary, contains visual/graphic, targeted skill level and cognitive domain level and they determined 30 questions appropriate for the purpose of the research. These questions were about the periodic table, physical and chemical changes, chemical reactions, acids and bases, and the interaction of matter and heat. Six questions were determined for each topic.

After, expert opinions were received about Bloom's cognitive domain level (remembering, understanding, applying, analyzing, evaluating and creating),

targeted scientific process skills (basic, causal and experimental) and cognitive load of the questions. For each topic, five science teachers expressed their opinions about the skill level, cognitive domain level and cognitive load of the questions. Five questions were removed from the test with science teacher's consultation. 25-item test was applied to 252 ninth grade students and difficulty and discrimination indexes of the questions were calculated. These students were studying in three high schools, two of which accepted students with examination and one without examination. Convenience and purposive sampling techniques from non-random sampling technique was used at this stage. Participants were selected from volunteers who were easier accessible. It was prioritized that the participant group of the 25-item test and the participant group of the eye tracking technique showed similar characteristics. The sample group for the 25-item test was detailed in Table 1.

Item analyses were performed on the answers of students who had prior knowledge about the "Matter and Industrial" unit. The difficulty and discrimination indexes of the questions were calculated. For this purpose, each of the lower and upper groups was determined as 72 participants. After the analysis, statistically significant differences were found between the lower and upper groups in most of the questions [except questions thirteenth ( $t=-1.346$ ;  $p < .05$ ) and twenty-fourth ( $t=-1.549$ ;  $p < .05$ )]. The difficulty and discrimination indexes of each question are presented in Table 2.

Nine questions were selected from the 25-question test to be used in the eye tracking technique application. Two questions were selected from each topic field (the periodic table, physical and chemical changes, chemical reactions and acids and bases). A question was selected from the topic of the interaction of heat and matter. These questions were determined according to the difficulty and discrimination indexes, paragraph length, presence of unfamiliar vocabulary, contains visual/graphic, targeted skill level and cognitive domain level. Questions' targeted skill levels and cognitive domain levels are presented Table 3 (The questions were renumbered according to the order in Table 2). Three questions each measuring basic, causal and experimental scientific process skills were included for the eye tracking technique application. According to the revised Bloom' cognitive domain level,

**Table 1** Sample group for the 25-question test

Gender	<i>N</i>	Percentage
Female	158	62.7
Male	94	37.3
Total	252	100
High schools		
Science high school (A high school that accepts students with high school's entrance exam score)	84	33.3
Anatolian high school (A high school that accepts students with high school's entrance exam score)	82	32.5
Anatolian high school (A high school that accepts students without exams)	86	34.2
Total	252	100

**Table 2** The difficulty and discrimination indexes

Questions	Difficulty indexes	Discrimination indexes
Question 1	0.84	0.23
Question 2*	0.72	0.43
Question 3	0.74	0.43
Question 4	0.48	0.66
Question 5*	0.58	0.72
Question 6	0.76	0.41
Question 7	0.87	0.25
Question 8	0.36	0.48
Question 9	0.61	0.62
Question 10*	0.46	0.51
Question 11	0.45	0.69
Question 12*	0.68	0.52
Question 13	0.22	0.09
Question 14*	0.64	0.62
Question 15	0.56	0.50
Question 16	0.42	0.54
Question 17	0.56	0.80
Question 18*	0.65	0.66
Question 19*	0.31	0.20
Question 20	0.55	0.50
Question 21*	0.47	0.68
Question 22	0.52	0.62
Question 23	0.50	0.51
Question 24	0.17	0.09
Question 25*	0.26	0.16

\* Questions used in eye tracking technique

there were two questions for each of the applying, analyzing and evaluating levels. In addition, three questions were included for the level of understanding. Thus, the nine-question skill-based science test was prepared for the eye tracking application.

The reliability coefficient of the nine-question test was calculated as 0.63 using the KR-20 formula. The average difficulty level of the test was 0.53. The optimal difficulty level for four-choice items is approximately 0.62 (Kaplan & Saccuzzo, 1997). According to these results, it can be said that the nine-question test is a little difficult for eighth grade students. Sample questions are presented in the appendix.

### 2.2.2 Eye tracking device

The researchers aimed to reveal the visual measurement results related to eighth grade students' process/behaviors of solving skill-based science questions. A pilot implementation was carried out by the first author with a different sample group to gain experience in the use of the eye tracking device before the implementation.

**Table 3** Questions' targeted skill levels and cognitive domain levels

Questions	Acquisitions	Bloom's cognitive domain level	Targeted science process skills
Question 1	It classifies elements as metals, semi-metals and non-metals on the periodic table.	Applying	B-SPS (classification)
Question 2	Explains how groups and periods are created in the periodic table.	Analyzing	C-SPS (inference)
Question 3	He/she knows that compounds are formed as a result of a chemical reaction.	Understanding	E-SPS (to decide)
Question 4	Explains the differences between physical and chemical change by observing various events.	Understanding	C-SPS (inference)
Question 5	He/she knows that compounds are formed as a result of a chemical reaction.	Applying	B-SPS (to observe)
Question 6	He/she knows that compounds are formed as a result of a chemical reaction.	Evaluating	E-SPS (to decide)
Question 7	Observes the effects of acids and bases on various substances.	Evaluating	E-SPS (designing an experiment)
Question 8	He gives examples of acids and bases from everyday life.	Analyzing	C-SPS (inference)
Question 9	It interprets the state change of substances and the heating graph by drawing.	Understanding	B-SPS (number and space relations)

*B-SPS* basic scientific process skills, *C-SPS* causal scientific process skills, *E-SPS* experimental scientific process skills

Before the implementation of current research, participants were informed about the purpose of the research. After that calibration was performed for each student. During the calibration process, participants were asked to follow the points on the computer screen. Thus, the sensitivity of the participants' eye movements was revealed with the eye tracking devices. At this stage, an expert with experience in using the eye tracking device assisted. Students' the fixation and visit duration or number (time to first fixation, duration of first fixation, number of fixations, average duration of fixation, total duration of fixation, number of visits, average duration of visit, total duration of visit) on science questions were detected through sensors on eye tracking devices. Heat maps and eye splash movements were revealed.

### 2.2.3 Think aloud protocol

Students were asked to solve the questions by thinking aloud. The first author recorded the steps/process that the participants expressed aloud with a voice recorder. Voice recordings and researcher's notes were transcribed together.

### 2.2.4 Perception scale towards solving new generation science questions

This scale was developed by Yiğit et al. (2022). The scale included three subscales: self-efficacy, attitude and willingness. The self-efficacy sub-dimension includes items that measure the student's belief that they can solve new generation science questions. The attitude sub-dimension measures the student's attitude towards new generation science questions. The willingness sub-dimension measures the student's willingness to solve new generation science questions. Sample items for each sub-dimension are given below:

- I think that I will fail the high schools entrance exam because of the new generation science questions (for self-efficacy sub dimension).
- Solving new generation science questions activates my sense of curiosity (for attitude sub dimension).
- I hate new generation science questions (for willingness sub dimension).

The scale was a five-point *Likert* type. Permission was obtained from the developers of the scale. In the current study, this perception scale was used to determine the participants' perception levels towards solving new generation science questions. Thus, the relationship between the participants' perception levels and visual measurement results was revealed.

## 2.3 Data analysis

SPSS 22 statistical program was used to analyze quantitative data. Visual measurement data were analyzed with quantitative descriptive analysis techniques and the results were presented in tables. It was determined by t-tests, one-way ANOVA, Mann Whitney U test and Kruskal Wallis test whether visual measurements results

differed according to gender and high schools entrance exam score. The relationships between visual measurement results and practice test scores and perception levels were revealed using correlation analysis techniques. Finally, the predictive effects of visual measurement results on the practice test scores and perception levels were determined using regression analysis techniques. Eye tracking device data were examined using Gaze Viewer software and the results were shown as heat maps and eye splash movements. Think-aloud protocol data were analyzed with qualitative descriptive analysis techniques. Qualitative data of heat maps, visuals of eye movement, and think-aloud protocols were analyzed together by the researchers. In the current study, quantitative data regarding the visual measurement results obtained from the eye tracking device were supported with images of heat maps and eye splash movements and qualitative descriptive quotes obtained from think-aloud protocols. Thus, data triangulation was achieved by supporting quantitative data with qualitative data. The research process is summarized in Fig. 2.

### 3 Results

#### 3.1 Results for the first sub-problem

The first research question of the study: What are the descriptive results of students' fixation and visit duration and number on questions? Descriptive results were found regarding the students' fixation duration (time to first fixation, duration of first fixation, average/total duration of fixation) and number of fixations for each question. Additionally, the duration (average/total duration of visit) and number of visits of the students to each question were revealed. Descriptive results of visual measurement are presented in Table 4.

It was observed that time to first fixation increased from the first question to the last question. The first question was about the periodic table, the last question was

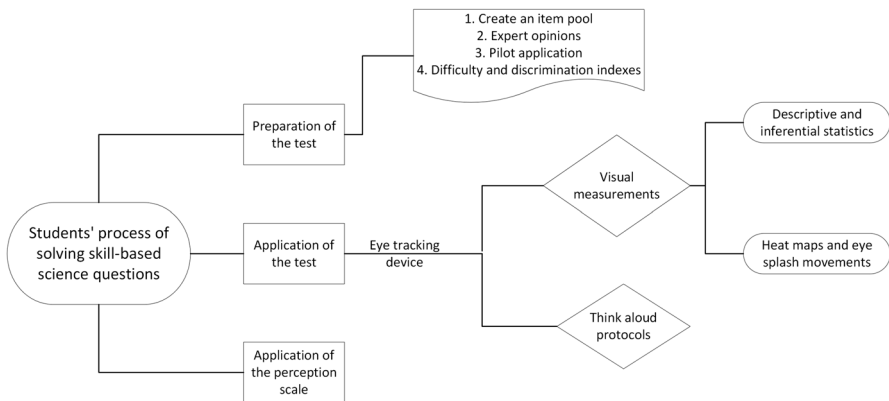


Fig. 2 The research process



**Table 4** Descriptive results for visual measurement

Questions*		TFF	DFF	NF	ADF	TDF	NV	ADV	TDV
Q1	Mean	95.83	0.17	125.73	0.32	42.21	2.30	24.01	48.89
	SD	20.72	0.09	60.92	0.08	27.55	0.93	16.19	30.47
Q2	Mean	145.93	0.12	266.46	0.30	78.57	2.84	40.94	92.19
	SD	38.94	0.08	136.70	0.07	41.33	1.81	24.99	47.01
Q3	Mean	239.99	0.21	179.21	0.28	51.21	2.48	31.11	61.63
	SD	73.69	0.01	93.73	0.09	31.66	1.48	23.64	35.99
Q4	Mean	303.16	0.18	267.39	0.30	78.26	2.61	45.08	93.29
	SD	99.31	0.09	108.90	0.08	34.99	1.81	25.16	39.75
Q5	Mean	398.12	0.18	86.30	0.26	22.53	1.86	16.71	27.18
	SD	129.88	0.08	42.64	0.05	13.85	0.77	12.99	15.78
Q6	Mean	426.21	0.20	252.25	0.28	68.68	2.18	47.14	83.50
	SD	139.85	0.14	129.44	0.07	33.88	1.19	28.72	39.44
Q7	Mean	511.05	0.19	270.09	0.31	81.51	2.28	50.78	97.05
	SD	165.13	0.11	133.60	0.09	42.60	1.45	28.52	46.75
Q8	Mean	609.55	0.20	199.02	0.28	57.40	1.89	42.61	69.89
	SD	200.99	0.11	96.94	0.06	34.75	0.73	27.77	38.59
Q9	Mean	680.42	0.18	247.11	0.33	81.01	2.39	51.04	96.66
	SD	231.18	0.10	181.37	0.08	57.67	2.84	36.84	68.04

\* The questions were renumbered according to the sequence in Table 2

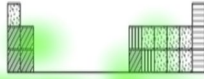
Time to first fixation: *TFF*; Duration of first fixation: *DFF*; Number of fixations: *NF*; Average duration of fixation: *ADF*; Total duration of fixation: *TDF*; Number of visits: *NV*; Average duration of visit: *ADV*; Total duration of visit: *TDV*

about interaction of matter and heat. Red areas on the heat maps indicate that the fixation time is too long. Accordingly, when the heat map for the first question was examined (see Fig. 3), it was seen that the student (student's code is S<sub>21</sub>) fixations more on option A (correct answer is D) by making an elimination among the options after reading the question text. For the ninth question, it was observed that after reading the question text, the student fixation on the table and the options and chose option B among the options (correct answer is C).

It was found that duration of first fixation were similar for all questions. The duration of first fixation was minimum for the second question (Mean = 0.12; SD = 0.08) and maximum for the third question (Mean = 0.21; SD = 0.01). Solving the third question required the use of experimental process skills. When the heat maps were examined (see Fig. 4), it was found that the students (student's code is S<sub>46</sub>) had difficulty eliminating the options I, II and III in the third question.

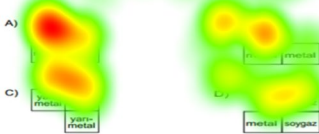
The number of fixations for the fifth question was minimum (Mean = 86.30; SD = 42.64). This question is a question prepared within the scope of the chemical reaction topic field, at the applying stage of the cognitive field according to the Bloom taxonomy, and aims to measure students' basic scientific process skills (observation skills).

S1- Şekilde bir kısmı verilen periyodik tabloda elementler; metal, ametal, yarı metal ve soygaz olma durumlarına göre farklı desenlerle taranarak gösterilmiştir.

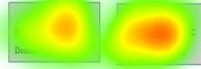


Bu periyodik tabloda verilen bir element ve bu elementle aynı grup ve periyotta yer alan komşu iki elementle üçlü gruplar oluşturuluyor.

Buna göre aşağıdaki üçlü gruplardan hangisi bu koşulu sağlamaz?

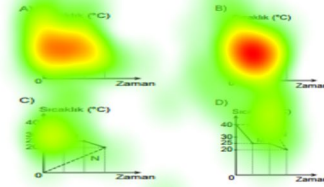


S9- Saf M sıvısı ile saf N katısına ait bilgiler verilmiştir.



İçinde M sıvısı olan bir kaba bu sıvıda çözünmeyen N katısı bırakılıyor. Isı alışverişi tamamlandıktan sonra sıcaklıkları 20 °C oluyor.

Bu durumda sıcaklık değişimini gösteren grafik aşağıdakilerden hangisi olabilir? Sıvıların M ve N maddeleri arasında olduğu düşünülüyor.



(a) Min. TFF for  $Q_1$ : 64.83 ms  
( $S_{21}$ )

(b) Max. TFF for  $Q_9$ : 1611.26 ms  
( $S_{11}$ )

Fig. 3 Heat maps for  $Q_1$  and  $Q_9$

The number of fixations for the seventh (Mean=270.09; SD=133.60) and fourth questions (Mean=267.39; SD=108.90) was maximum. Eye splash movements show the movements between the fixation points of the eye, respectively. Accordingly, the fixation numbers and eye movements are presented in Fig. 5. The fourth question was about chemical change and in the context of Covid-19. Compared to other questions, it contained many unknown concepts and pictures for the students. Solving the seventh question required designing experiments using dependent, independent and control variables. In other words, it required the use of experimental process skills.

The average duration of fixation (Mean=0.26 ms; SD=0.05) and total duration of fixation (Mean=22.53 ms; SD=13.85) for the fifth question was minimum. The average duration of fixation for the ninth question (Mean=0.33 ms; SD=0.08) was maximum. The ninth question was about interaction of matter and heat. In the think-aloud protocols, "some students stated that they could not answer this question because it would not appear in the high-school entrance exam and they left it blank" (student's code are  $S_5$  and  $S_{16}$ ). The total duration of fixation for the seventh question (Mean=81.51 ms; SD=42.60) was maximum. Solving this question required the use of experimental process skills. Additionally, it was observed that the student was undecided between the options in this question (e.g.: student with code  $S_6$ ) (Fig. 6).

The number of visits for the fifth question was minimum (Mean=1.86; SD=0.77). The number of visits for the second question (Mean=2.84; SD=1.81) was maximum. This question was about the periodic table and solving this question required mathematical operation. Therefore, it was found that students made

many transition movements between pictures and options. According to think-aloud protocols, *it was noted that some students made collection for one of the options.* This question is at the analysis stage of the cognitive field according to Bloom's taxonomy and aims to measure students' causal scientific process skills (inference).

S2- A grubu elementleri için; nötr haldaki bir element atomunun elektron dağılımındaki katman sayısı o elementin periyot numarasını, son katmandaki elektron sayısı ise grup numarasını verir.

Bir öğrenci proje ödevi olarak elementlerin periyodik tablodaki yerini gösteren şekildedeki gibi bir afiş tasarlıyor.

- Modelleri oluşturdukları bir periyodik tabloyu göstermektedir.
- Boncukların kullanılarak atomun yapısal modelini oluşturduğunu göstermektedir.

1. Katman

Hazırlanan bu proje ödevinden hareketle aşağıdaki çıkarımlardan hangileri doğrudur?

A) Kibrit alevinin sıcaklığı, bu deney ile ölçülebilir.

B) Alevin rengi, kibrit alevi arasındaki ilişkiyi gösterir. (Sarı)

C) Yeşil alevi, bu deneyde kullanılan elementler arasında sıradaki bulunur.

D) Mavi renkli elementler periyodik tablonun en sağında bulunur.

(a) Min. DFF for Q<sub>2</sub>: .01 ms (S<sub>36</sub>)

Bu deneyde kati haldeki bir bileşen deney tüpüne konuluyor. Tüpün ağzını kapatıp bir kibrit Şekil 1' deki gibi yaklaştırdığında alevin parlaklığında değişim olmadığı gözleniyor.

Bu deney için Şekil 1'deki gibi bir süre istildikten sonra içindeki bileşenin parlaklığında değişim olmadığı ve yaklaştırdıkları kibrit alevinin parlaklığının arttığı gözleniyor.

Bu deneyde alevin parlaklığında değişim olduğu gözleniyor. Bu deney ile ilgili;

A) Kibrit alevinin sıcaklığı, bu deney ile ölçülebilir.

B) Alevin rengi, kibrit alevi arasındaki ilişkiyi gösterir. (Sarı)

C) Yeşil alevi, bu deneyde kullanılan elementler arasında sıradaki bulunur.

D) Mavi renkli elementler periyodik tablonun en sağında bulunur.

(b) Max. DFF for Q<sub>3</sub>: .48 ms (S<sub>46</sub>)

Fig. 4 Heat maps for Q<sub>2</sub> and Q<sub>3</sub>

3- Kimyasal tepkime sürecinde atom ya da moleküller arasında yeni bağlar kırılır veya yeni bağlar kırılır.

5- 300 g A maddesi verildiğinde, aşağıdaki verden modellerden hangisi kimyasal tepkimeyi gösterir?

a)  $2A + 3B \rightarrow 4C + 5D$

b)  $3A + 2B \rightarrow 4C + 5D$

c)  $4A + 3B \rightarrow 5C + 2D$

d)  $5A + 4B \rightarrow 2C + 3D$

(a) Min. NF for Q<sub>5</sub>: 30 (S<sub>55</sub>)


(b) Max. NF for Q<sub>7</sub>: 725 (S<sub>6</sub>)


(c) Max. NF for Q<sub>4</sub>: 693 (S<sub>6</sub>)


Fig. 5 Eye movements for Q<sub>5</sub>, Q<sub>7</sub> and Q<sub>4</sub>


S5- Kimyasal tepkime sürecinde atom ya da moleküller arasında yeni bağlar oluşur veya var olan bağlar kırılır.

Görebildiğiniz kadar, aşağıda verilen modellerden hangisinin kimyasal tepkimeyi gösterdiğini seçiniz?

A) 

B) 

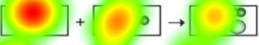
C) 


D) 

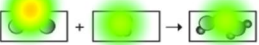
(a) Min. ADF for Q<sub>5</sub>: .17 ms  
(S<sub>45</sub>)


S6- Kimyasal tepkime sürecinde atom ya da moleküller arasında yeni bağlar oluşur ya da var olan bağlar kırılır.

Görebildiğiniz kadar, aşağıda verilen modellerden hangisinin kimyasal tepkimeyi gösterdiğini seçiniz?

A) 

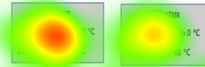
B) 

C) 

D) 

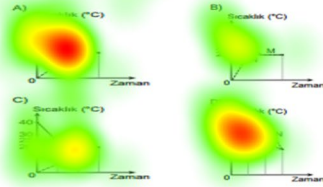
(c) Min. TDF for Q<sub>5</sub>: 6.93 ms  
(S<sub>25</sub>)

S9- Saf M sıvısı ile saf N katısına ait bilgiler verilmiştir.



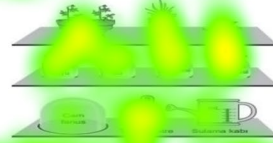
İçinde M sıvısı olan bir kaba bu sıvıda çözünmeyen N katısı birleştiriliyor. Isı alışverişi tamamlandıktan sonra son sıcaklıkları 20 °C oluyor.

Bu olay sırasında M ve N maddeleri arasındaki sıcaklık değişimini gösteren grafik aşağıdakilerden hangisidir? (Isı alışverişinin sadece M ve N maddeleri arasında olduğu düşünülecektir.)



(b) Max. ADF for Q<sub>9</sub>: .54 ms  
(S<sub>10</sub>)

S7- Bir deney yapılarak bitki büyümesini etkileyen bir maddeyi araştırılıyor.



Bitki büyüme hızını ölçmek için bitki ve tohumlardan oluşan bir deney düzeni hazırlanıyor. Seçilen bitki türü ve tohum türü aşağıdaki gibidir.

Deneyin amacı bitki büyümesini etkileyen maddelerin karabotanyıldızını belirlemektir.

Deneyin amacı bitki büyümesini etkileyen maddelerin karabotanyıldızını belirlemektir.

Deneyin amacı bitki büyümesini etkileyen maddelerin karabotanyıldızını belirlemektir.

D) Bitki büyüme hızını ölçmek için bitki ve tohumlardan oluşan bir deney düzeni hazırlanıyor. Seçilen bitki türü ve tohum türü aşağıdaki gibidir.

(d) Max. TDF for Q<sub>7</sub>: 204.03 ms  
(S<sub>6</sub>)

Fig. 6 Heat maps for Q<sub>5</sub>, Q<sub>9</sub> and Q<sub>7</sub>

Finally, the average duration of visit (Mean=16.71 ms; SD=12.99) and total duration of visit (Mean=27.18 ms; SD=15.78) for the fifth question was minimum. The average duration of visits for the ninth question (Mean=51.04 ms; SD=36.84) was maximum. The total duration of visits for the seventh question (Mean=97.05 ms; SD=89.43) was maximum. This result was consistent with the average duration of fixation and total duration of fixation (Fig. 7).

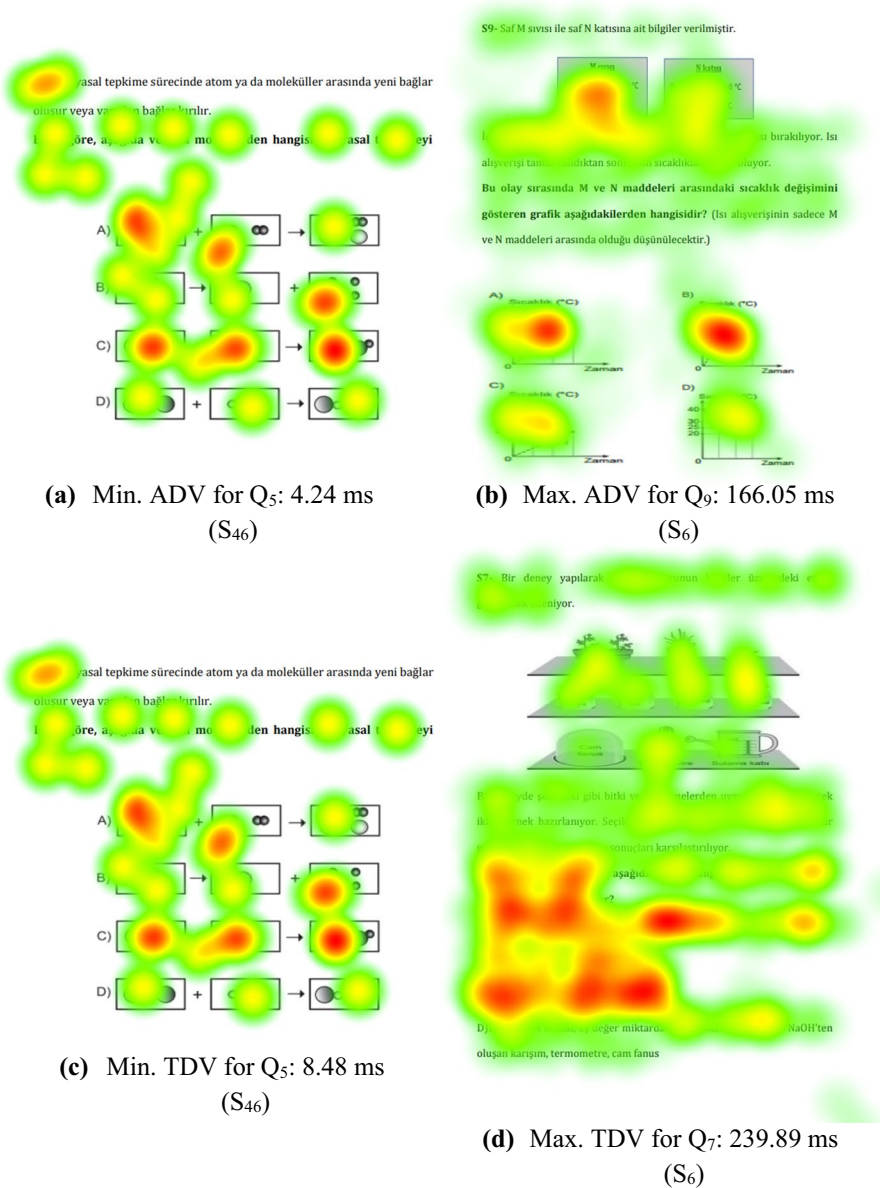


Fig. 7 Heat maps for Q<sub>5</sub>, Q<sub>9</sub> and Q<sub>7</sub>



### 3.2 Results for the second sub-problem

The second research question of the study: Do the results of visual measurements differ according to gender and 2023 high school entrance exam scores? In this context, the research revealed whether the results of visual measurements differ according to gender and 2023 high schools entrance exam scores.

#### 3.2.1 Results for visual measurements by gender

First, the skewness and kurtosis values were examined for each item. An independent sample t-test was performed for items with normal distribution, and Mann Whitney U test was performed for items that did not show normal distribution. In order to decide on the items with normal distribution, criteria were taken as skewness value 2 and kurtosis value 7 (i.e.: George & Mallery, 2020; Coşguner, 2022). The questions that provide the assumption of normality are presented in Table 5.

After t-test analysis, no differences were found according to gender. According to Mann Whitney U test results, for the number of visits variable, a difference was found according to gender in the second question (See Table 6). It was found that males (Mean = 3.40; SD = 2.22) visited the second question more than females (Mean = 2.19; SD = 0.80). According to this finding, it can be said that male visit more frequently to solve the second problem prepared within the scope of the periodic table.

**Table 5** Questions that provide the assumption of normality

Variables	Questions
Time to first fixation	Q1, Q2 and Q3
Duration of first fixation	All questions
Number of fixations	Q1, Q3, Q4, Q5, Q7 and Q8
Average duration of fixation	Q1, Q2, Q5, Q6, Q8 and Q9
Total duration of fixation	Q4, Q6, Q7 and Q8
Number of visits	Q1, Q5, Q6 and Q8
Average duration of visit	Q1, Q2, Q4, Q6, Q7, Q8 and Q9
Total duration of visit	Q3, Q4, Q5, Q6 and Q7

Q = Question

**Table 6** Mann Whitney U test results

Questions	Group	N	The average of rank	The sum of ranks	U	p
The number of visits for second question	Female	26	23.40	608.50	257.500	0.023*
	Male	30	32.92	987.50		

\* $p < .05$

### 3.2.2 Results for visual measurements by 2023 high schools entrance exam scores

Participants were divided into three categories according to their high school's entrance exam scores. In the first group, there were students ( $f=19$ ) with a score of 450–500. In the second group, there were students ( $f=19$ ) with scores of 400–449. In the third group, there were students ( $f=18$ ) with scores below 400. The significant differences that emerged after one-way ANOVA are presented in Table 7.

Differences were found in the visual measurement results of participants' behaviors/processes of solving skill-based science questions according to their high school's entrance exam scores. For the eighth question, significant differences were found in the students' fixation number, total duration of fixation and average duration of visit. According to Tukey results from Post Hoc tests, this difference was in favor of high-scoring students. High-scoring students had less numbers of fixations (Mean = 167.52; SD = 57.83) than low-scoring students (Mean = 250.83; SD = 122.13), and the difference was statistically significant ( $F_{(2-53)} = 4.362$ ;  $p < .05$ ). Additionally, high-scoring students had less total duration of fixation (Mean = 45.25 ms; SD = 18.66) than low-scoring students (Mean = 74.81 ms; SD = 35.05), and the difference was statistically significant ( $F_{(2-53)} = 3.954$ ;  $p < .05$ ). Similarly, high-scoring students had less average duration of visit (Mean = 33.64 ms; SD = 17.17) than low-scoring students (Mean = 59.12 ms; SD = 36.90), and the difference was statistically significant ( $F_{(2-53)} = 5.494$ ;  $p < .05$ ). Comparing the fixation number of students with high scores (482) and students with low scores (350) in high school entrance exam for the 8th question, it was seen that high-scoring student' number of fixations was 35 (student's code is  $S_{48}$ ), while low-scoring student' number of fixations was 282 (student's code is  $S_{16}$ ) (See Fig. 8). Heat maps for total duration of fixation and average duration of fixation showed that the result was in favor of high-scoring students.

Questions that did not meet the normality assumption were analyzed with the Kruskal Wallis test and results are presented in Table 8. Statistically significant

**Table 7** One-way ANOVA results

		Sum of squares	df	Mean square	F	<i>p</i>
Number of fixations for eight question	Between groups	73053.114	2	36526.557	4.362	0.018*
	Within groups	443807.868	53	8373.733		
	Total	516860.982	55			
Total duration of fixation for eight questions	Between groups	8621.427	2	4310.714	3.954	0.025*
	Within groups	57785.124	53	1090.285		
	Total	66406.551	55			
Average duration of visit for eight question	Between groups	7285.343	2	3642.672	5.494	0.007*
	Within groups	35141.896	53	663.055		
	Total	42427.239	55			

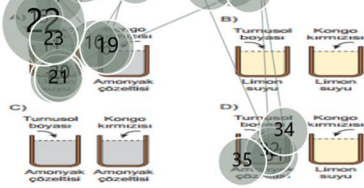
\* $p < .05$



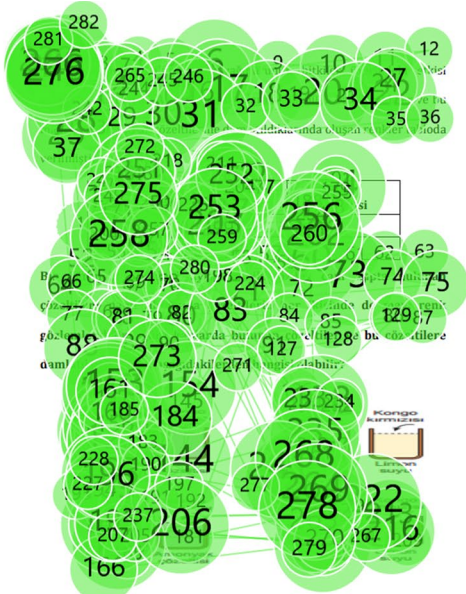
S8- Bir deney yapılarak asit yağmurunun bitkiler üzerindeki etkisi gözlenmek isteniyor. Asit-baz ayırıcı olarak kullanılan bazı maddeler ve bu maddeler asit ve baz çözeltilerine damlatıldıklarında oluşan renkler tabloda verilmiştir.

Asit	Asit	Baz
Çözeltisi	Çözeltisi	Çözeltisi
Turmusul boyası	Mavi	Mavi
Kongo kırmızısı	Mavi	Kırmızı

Bir öğrenci, tablodaki ayraçları iki özdeş cam kaptaki bulunan çözeltilere damlatıldığında çözeltilerin her ikisinde de mavi renk gözlemlendiğine göre kaplarda bulunan çözeltiler ve bu çözeltilere damlatılan ayraçları aşağıdaki gibi hangisi olabilir?



(a) High-scoring student' number of fixation for Q8: 35 (S<sub>48</sub>)

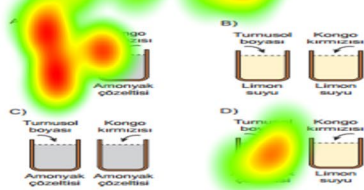


(b) Low-scoring student' number of fixation for Q8: 282 (S<sub>16</sub>)

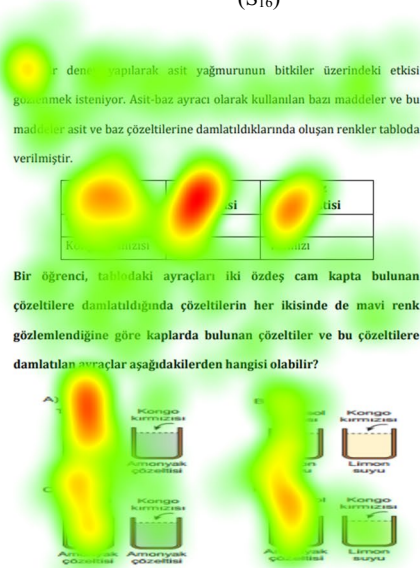
S8- Bir deney yapılarak asit yağmurunun bitkiler üzerindeki etkisi gözlenmek isteniyor. Asit-baz ayırıcı olarak kullanılan bazı maddeler ve bu maddeler asit ve baz çözeltilerine damlatıldıklarında oluşan renkler tabloda verilmiştir.

Asit	Asit	Baz
Çözeltisi	Çözeltisi	Çözeltisi
Turmusul boyası	Mavi	Mavi
Kongo kırmızısı	Mavi	Kırmızı

Bir öğrenci, tablodaki ayraçları iki özdeş cam kaptaki bulunan çözeltilere damlatıldığında çözeltilerin her ikisinde de mavi renk gözlemlendiğine göre kaplarda bulunan çözeltiler ve bu çözeltilere damlatılan ayraçları aşağıdaki gibi hangisi olabilir?



(c) High-scoring student' average duration of fixation for Q8: .18 ms (S<sub>48</sub>)



(d) Low-scoring student' average duration of fixation for Q8: .34 ms (S<sub>16</sub>)

Fig. 8 Eye movements and heat maps according to high school's entrance exam scores

**Table 8** Kruskal Wallis test results

	<i>N</i>	Mean	The average of rank	df	$\chi^2$	<i>p</i>
Total duration of fixation for three question	19	39.83	20.74	2	6.540	0.038*
	19	55.59	32.89			
	18	58.59	32.06			
Total duration of fixation for five question	19	17.10	21.53	2	7.340	0.025*
	19	24.69	28.32			
	18	25.98	36.06			
Total duration of visit for eight question	19	56.64	23.11	2	8.170	0.017*
	19	66.54	25.42			
	18	87.41	37.44			

\* $p < .05$

differences were found between the students' total duration of fixation for the third and fifth questions. The difference was in favor of students with high scores. Students with high scores had less total duration of fixation. Additionally, high-scoring students' total duration of visit for the eighth question was less. When the heat maps and eye splash movements of the high-scoring student were examined, it was seen that the student with a high score was undecided between two similar distractors, but the low-scoring student looked at all distractors.

### 3.3 Results for the third sub-problem

The research revealed the relationship between visual measurement results, students' practice test scores and their perception levels towards solving skill-based science questions.

#### 3.3.1 Relationship between visual measurement results and practice test scores

First, scatter plot matrixes were examined for binary normality assumptions of each variable (Kline, 2011, p.65; Tabacnick & Fidell, 2015, p.83). Since the distributions were close to ellipse, it was seen that the binary normality assumptions were met. When the descriptive statistics of the test scores were examined, it was seen that the test scores exhibited a normal distribution (Skewness:  $-0.62$ ; Kurtosis:  $-0.37$ ). The relationship between visual measurement results and practice test scores is presented in Table 9.

No correlation was found between time to first fixation and duration of first fixation and practice test scores for all questions. Significant negative relationships were found between number of fixation and practice test scores for the first ( $r_{NF}=-0.272$ ;  $p < .05$ ) and eighth questions ( $r_{NF}=-0.386$ ;  $p < .01$ ). Significant negative relationships were detected between average duration of fixation and practice test scores for the third ( $r_{ADF}=-0.274$ ;  $p < .05$ ), fourth ( $r_{ADF}=-0.356$ ;  $p < .01$ ), sixth ( $r_{ADF}=-0.321$ ;

**Table 9** The relationship between visual measurement results and total test scores

Questions	TFF	DFD	NF	ADF	TDF	NV	ADV	TDV
Q1	0.068	0.033	-0.272*	-0.213	-0.324*	-0.300*	-0.032	-0.320*
Q2	-0.220	-0.102	0.133	-0.239	-0.001	0.182	-0.046	0.030
Q3	-0.094	-0.112	-0.096	-0.274*	-0.212	0.114	-0.211	-0.189
Q4	-0.137	-0.042	0.151	-0.356**	-0.055	-0.041	-0.008	-0.018
Q5	-0.109	-0.016	-0.127	-0.239	-0.171	-0.208	0.011	-0.167
Q6	-0.121	-0.070	-0.146	-0.321*	-0.280*	-0.009	-0.145	-0.241
Q7	-0.164	-0.232	-0.082	-0.313*	-0.214	-0.233	0.014	-0.164
Q8	-0.175	-0.153	-0.386**	-0.366**	-0.440**	-0.041	-0.330*	-0.386**
Q9	-0.217	0.101	-0.227	-0.352**	-0.323*	-0.371**	0.069	-0.283*

$p^* < 0.05$ ;  $p^{**} < 0.01$

$p < .05$ ), seventh ( $r_{ADF} = -0.313$ ;  $p < .05$ ), eighth ( $r_{ADF} = -0.366$ ;  $p < .01$ ) and ninth questions ( $r_{ADF} = -0.352$ ;  $p < .01$ ). These relationships were moderate (there was a weak relationship for the third question). Additionally, significant negative relationships were determined between total duration of fixation and practice test scores in the first, sixth, eighth and ninth questions. On the other hand, Table 9 showed that there were negative significant relationships between number of visits and total duration of visit and practice test scores for the first and ninth questions. Negative significant relationships were determined between average duration of visit, total duration of visit and practice test scores for the eighth question.

In summary, no relationship was found between the practice test score and visual measurement results for the second and fifth questions. The questions that had a relationship between at least one visual measurement (average duration of fixation) and the practice test score were the third, fourth and seventh questions. For the sixth question, there was a negative relationship between at least two visual measurements (average duration of fixation and total duration of fixation) and the practice test score. A correlation was found between at least four visual measurements and practice test scores for the first, eighth and ninth questions. All significant relationships were negative.

### 3.3.2 Relationship between visual measurement results and perception levels

First, scatter plot matrixes were examined for binary normality assumptions of each variable (Kline, 2011, p.65; Tabacnick & Fidell, 2015, p.83). Since the distributions were close to ellipse, it was seen that the binary normality assumptions were met. When the descriptive statistics of students' perception levels were examined, it was seen that the perception levels exhibited a normal distribution (Skewness:  $-0.69$ ; Kurtosis: 1.479). The relationship between visual measurement results and perception levels is presented in Table 10.

Table 10 showed that there was no relationship between duration of first fixation and average duration of visit and perception levels for all questions. Negative relationships were found between time to first fixation and perception levels for the

**Table 10** The relationship between visual measurement results and perception levels

Questions	TFF	DFE	NF	ADF	TDF	NV	ADV	TDV
Q1	-0.077	-0.248	-0.284*	-0.430**	-0.434**	-0.177	-0.198	-0.423**
Q2	-0.377**	-0.161	0.073	-0.402**	-0.139	0.187	-0.250	-0.093
Q3	-0.255	-0.073	0.038	-0.492**	-0.244	0.008	-0.148	-0.207
Q4	-0.263	-0.181	-0.097	-0.418**	-0.351**	-0.323*	-0.022	-0.326*
Q5	-0.308*	0.057	-0.266*	-0.172	-0.325*	-0.061	-0.254	-0.327*
Q6	-0.323*	0.019	0.074	-0.396**	-0.113	-0.106	-0.088	-0.062
Q7	-0.294*	-0.148	0.035	-0.390**	-0.155	0.025	-0.158	-0.120
Q8	-0.270*	-0.188	-0.168	-0.381**	-0.338*	-0.074	-0.177	-0.308*
Q9	-0.287*	-0.063	-0.015	-0.296*	-0.112	-0.130	0.050	-0.076

$p^* < 0.05$ ;  $p^{**} < 0.01$

second, fifth, sixth, seventh, eighth and ninth questions. The Table 10 showed that there were significant negative relationships between number of fixation and perception levels for the first and fifth questions. Additionally, significant negative relationships were found between average duration of fixation and perception levels for all questions except the fifth question. There were significant negative relationships between total duration of fixation and students' perception levels for the first, fourth, fifth, and eighth questions. On the other hand, Table 10 showed that there were moderately negative significant relationships between the number of visits and students' perception levels for the fourth question ( $r_{NV} = -0.323$ ;  $p < .05$ ). Negative significant relationships were found between total duration of visit and perception levels for the first, fourth, fifth and eighth questions. In summary, significant negative relationships were found between the perception level and at least one visual measurement results for all questions.

In the study, a three-dimensional scale (self-efficacy, attitude and willingness) was used to determine eighth grade students' perceptions towards solving new generation science questions. It was revealed in the relationships between the participants' perception levels in these sub-dimensions and the visual measurement results regarding the solving of the questions. It was observed that there were almost no significant relationships between duration of first fixation, the number of fixations, the number of visits, and the sub-dimensions of perception. Negative relationships were found between self-efficacy and at least one visual measurement results for all questions. There were negative significant relationships between average duration of fixation and self-efficacy (for all questions), attitude (for the first, third and fourth questions) and willingness (for the first, second, third, fourth and seventh questions) sub-dimensions. Additionally, negative significant relationships were found between average duration of visit and total duration of visit and the sub-dimensions of perception. However, these relationships were not at the frequency and level of average duration of fixation and total duration of fixation. In summary, it was observed that there were more statistically significant negative relationships between visual measurement results and self-efficacy, one of the sub-dimensions of perception towards solving new generation science questions, than attitude and willingness.

### 3.4 Results for the fourth sub-problem

Researchers revealed the predictive effects of visual measurement results on practice test scores and perception towards solving next-generation science questions. First, the assumptions regarding multiple linear regression analysis were tested. Since the measurement results of duration of first fixation, average duration of fixation, number of visits and average duration of visits met the assumptions of normality, multiple linear analyzes were carried out on these variables. In addition, it was seen that they met the assumption of normality in the descriptive statistical results of practice test scores and perception levels. First, the correlations between the predictive variables are presented in Table 11. The highest correlation was between the number of visits and the average duration of visits ( $r = -.359$ ;  $p < .01$ ). This relationship was moderate, negative and significant. The lowest relationship was between duration of first fixation and number of visits ( $r = .010$ ;  $p > .05$ ). This relationship was weak, positive and not statistically significant. It was observed that the VIF values of visual measurements were below 10 and the relationship between the predictive variables was below 0.90. Therefore, it was decided that there was no obstacle to multiple linear regression.

#### 3.4.1 Predictive effect of visual measurement results on practice test scores

Multiple linear regression analysis results are presented in Table 12. Visual measurement results together did not show a significant relationship ( $R = .393$ ,  $R^2 = 0.155$ ) with practice test scores ( $F_{(4-51)} = 2.331$ ;  $p > .05$ ). Four variables explained 15.5% of

**Table 11** Correlations between visual measurement results

	Duration of first fixation	Average duration of fixation	Number of visits	Average duration of visit
Duration of first fixation		0.332*	0.010	-0.059
Average duration of fixation			0.112	0.234
Number of visits				-0.359**
Average duration of visit				

$p^* < 0.05$ ;  $p^{**} < 0.01$

**Table 12** Predictive effect of visual measurement results on practice test scores

Variables	B	Std. Error	$\beta$	t	p	Zero-order (r)	Partial (r)
Constant	11.537	1.856		6.216	0.000		
Duration of first fixation	-4.939	7.185	-0.095	-0.687	0.495	-0.096	-0.089
Average duration of fixation	-7.607	4.555	-0.244	-1.670	0.101	-0.228	-0.215
Number of visits	-0.538	0.356	-0.215	-1.511	0.137	-0.207	-0.195
Average duration of visit	-0.017	0.018	-0.140	-0.956	0.343	-0.133	-0.123

the total variance in practice test scores. The order of relative importance of visual measure results on practice test scores was average duration of fixation, number of visits, average duration of visit and duration of first fixation.

### 3.4.2 Predictive effect of visual measurement results on perception levels

Multiple linear regression analysis was performed to determine whether variables had a significant relationship with students' perception levels towards solving new generation science questions, and the results are presented in Table 13.

**Table 13** Predictive effect of visual measurement results on perception levels

Variables	B	Std. Error	$\beta$	t	p	Zero-order (r)	Partial (r)
Constant	5.656	0.505					
Duration of first fixation	-2.383	1.956	-0.162	-1.218	0.229	-0.168	-0.150
Average duration of fixation	-2.799	1.240	-0.315	-2.257	0.028	-0.301	-0.278
Number of visits	-0.109	0.097	-0.152	-1.122	0.267	-0.155	-0.138
Average duration of visit	-0.006	0.005	-0.172	-1.227	0.226	-0.169	-0.151

**Table 14** Predictive effect of visual measurement results on sub-dimensions of perception

Sub-dimensions	B	Std. Error	$\beta$	t	p	Zero-order (r)	Partial (r)
For self-efficacy sub-dimension							
Constant	5.951	.528		11.267			
Duration of first fixation	-1.951	2.044	-.117	-.954	.345	-.132	-.109
Average duration of fixation	-4.601	1.296	-.459	-3.550	.001	-.445	-.405
Number of visits	-.061	.101	-.076	-.607	.546	-.085	-.069
Average duration of visit	-.007	.005	-.188	-1.450	.153	-.199	-.165
For attitude sub-dimension							
Constant	5.384	.800		6.732			
Duration of first fixation	-3.452	3.096	-.159	-1.115	.270	-.154	-.147
Average duration of fixation	-2.298	1.963	-.175	-1.171	.247	-.162	-.155
Number of visits	-.090	.153	-.086	-.588	.559	-.082	-.078
Average duration of visit	-.008	.008	-.153	-1.019	.313	-.141	-.135
For willingness sub-dimensions							
Constant	5.633	.507		11.104			
Duration of first fixation	-1.747	1.963	-.125	-.809	.378	-.124	-.116
Average duration of fixation	-1.499	1.245	-.178	-1.204	.234	-.166	-.157
Number of visits	-.175	.097	-.258	-1.796	.078	-.244	-.234
Average duration of visit	-.003	.005	-.082	-.550	.585	-.077	-.072

Table 13 showed that visual measurement results together showed a significant relationship with students' perceptions of solving new generation science problems ( $F_{(4-51)}=3.734$ ;  $p < .05$ ). Four variables explained 22.7% of the total variance in perception levels. Average duration of fixation ( $\beta = -0.315$ ,  $p < .05$ ) was a significant predictor of students' perceptions towards solving new generation science problems.

Additionally, multiple linear regression analysis was performed to determine whether the variables had a predictive effect on self-efficacy, attitude and willingness sub-dimensions (See Table 14). The results showed that four variables together showed a significant relationship with self-efficacy, which is the sub-dimension of perception ( $F_{(4-51)}=6.475$ ;  $p < .05$ ). Four variables explained 33.7% of the total variance on self-efficacy. Average duration of fixation was a significant predictor of self-efficacy towards solving new generation science questions ( $\beta = -0.459$ ,  $p < .05$ ). On the other hand, it was found that the four variables together did not show a significant relationship with attitude ( $F_{(4-51)}=1.564$ ;  $p > .05$ ) and willingness ( $F_{(4-51)}=2.012$ ;  $p > .05$ ), which are the sub-dimensions of perception. The four variables together explained 10.9% of attitude and 13.6% of willingness.

## 4 Conclusions and discussion

First, visual measurement results of eighth grade students' processes of solving skill-based science questions were presented. According to visual measurement results, the fifth question had the smallest values in terms of number of fixation, average/total duration of fixation, number of visits, and average/total duration of visit. This question was about chemical reactions and was a short and visual question. Additionally, this question is at the applying level according to Bloom's taxonomy and aims to measure observation skills from basic scientific process skills. The results of the research showed that students easily solving such short and visual questions (%89.2). Part of the mental model is formed by examining the picture, and the picture speeds up the understanding of the paragraph (Eitel et al., 2013). Related literature revealed that students spend more time reading textual information than graphical information (Ho et al., 2014). It was reported that young readers' reading time for images is too short (Jian, 2016; Jian & Ko, 2017). Wu et al. (2021) reported that students solved visuals questions more easily in terms of reading comprehension and total duration of fixation compared to long-text questions. The current study revealed that the seventh and ninth questions had the greatest values in terms of average/total duration of fixation and average/total duration of visit. In other words, solving these questions required the most effort. The seventh question was about acid-base, and the ninth question was about the interaction of matter with heat. The seventh question inquired the dependent, independent and control variables with an experiment. This question was at the evaluating level of the cognitive domain according to Bloom's taxonomy, and it aims to measure the ability to create experiments from experimental scientific process skills. The results of the research showed that students focused for a long time to understand the questions that required gaining experimental process skills and there were a lot of eye movements between the options and



the question text. The possible reason for the higher average and total duration of fixation may be due to the difficulty level of the question text and the presence of elements that require visual attention (Coşguner, 2022). It has been stated in many studies and in the exam reports of MoNE that the skill-based question measures different high-level skills (Erden, 2020). Karabulut et al. (2022) reported that students had difficulty understanding skill-based science questions and said that the questions were complex and long. Students explained the reason why they could not focus on skill-based questions was that the questions were long due to tables, visuals and graphs (Deveci et al., 2023). Negi and Mitra (2020) stated that operations involving complex tasks give longer measurement results. The ninth question was about the interaction of matter and heat. It was also a question involving graphics. This question was at the understanding level of the cognitive domain according to Bloom's taxonomy and aims to measure number and space relations skills from basic scientific process skills. The authorities announced that there will be no questions on this topic in the 2023 high schools entrance exam. This may be the reason why students spend a long time on this question. According to the think-aloud protocols, the students stated that *they did not remember the topic because it would not appear in 2023 high schools entrance exam*. Additionally, number of fixations for fourth and seventh questions was more than the other questions. The fourth question was a question about chemical and physical changes. This question was at the understanding level of the cognitive domain according to Bloom's taxonomy and aims to measure inference skills from causal scientific process skills. The fact that the question contains a detailed explanation may be the reason why number of fixations is too high. Another reason may be that the question is in the context of Covid-19. Jian (2021) reported that students who read a difficult article and practiced had better problem-solving performance than those who only read the same article. The difficult article contains unfamiliar academic language and many words (Snow, 2010; Wellington & Osborne, 2001). Concretizing abstract concepts requires practice (Osborne et al., 2016; Patterson et al., 2018). Solving the question in the context of Covid-19 required concretizing abstract concepts about physical and chemical change. Achieving this depends on solving a lot of practice problems. In the above studies, it was intended that the sense of touch supports the visual sense in understanding the phenomenon. Exercising with different types of research problems will contribute to concretization and learning of the topic. The fact that the participants did not solve enough research problems may explain the high number of fixations. Exposure to unfamiliar words twice increases vocabulary (Rott, 1999). Another possible explanation may be that students encounter some of the words in the question for the first time. In other words, the question may have contained many unfamiliar words for the students. On the other hand, when time to first fixation is examined, it was found that students focused on the first question for the shortest time and the ninth question for the longest time. The possible explanation was that students may have been influenced by the sounds and objects in the environment and getting bored towards the last question. One of the limitations of studies using the eye tracking technique is that the student gets excited, bored and anxious while solving questions in front of the computer (Karaođlan-Yılmaz & Yılmaz, 2019).

The question with the shortest duration of first fixation was the second question, and the question with the highest duration of first fixation was the third question. The question with the more number of visits was the second question. The second question was a question that contained a visual and long-text about the periodic table. This question was at the analyzing level of the cognitive domain according to Bloom's taxonomy and aims to measure inference skills from causal scientific process skills. This question involved numerical operations and students made many transition movements between options and figure. This may be a possible explanation for the short duration of first fixation and the more number of visits. It was reported that young readers' reading time for visuals was very short, but their referencing behavior between text and image was insufficient (Jian, 2016; Jian & Ko, 2017). Brief examination of images may not be sufficient for a partial mental model (Wu et al., 2021). In the current study, the number of visits for the second question was high, but this question was answered correctly by 57.1% of the students. Briefly examining the visual may have prevented creating a mental model to solve the problem. The first result of current research was that the cognitive loads of the questions, supporting the question with visuals, the targeted skill levels, presence of unfamiliar vocabulary and complexity of options were effective in the visual measurement results.

Another result of the research was that male students visited the second question more than female students. The difference was statistically significant. Coşguner (2022) reported that female students' average duration of visit for a verbal question was less than male students. In the same study, it was found that male students' total duration of visit was less for a numerical question. If limited time was spent examining the images in the texts, there was sufficient information in the text, the text was not containing ambiguity, or the image was not containing extra information (Wu et al., 2021). The second question was a question that included a visual and long explanation about the periodic table. Contrary to the relevant literature (i.e.: Wu et al., 2021), the visual in this question made a significant contribution to the understanding and solving of the problem. When the question solving videos were watched, *it was seen that the male students did not read the long paragraph in the text, but only operated their decision-making processes by looking at the visual. However, it was observed that they read the question several times when they realized that they could not solve the question without reading the paragraph.* This explained the reason for the many visits.

It was found that students with high scores in 2023 high schools entrance exam had less number of fixations, total duration of fixation and average duration of visit for the eighth question. Additionally, the total duration of fixation of high-scoring students for the third and fifth questions was less. It was determined that students with higher scores made shorter visits for the eighth question. Eighth question was at the analyzing level of the cognitive domain according to Bloom's taxonomy and aims to measure inference skills from causal scientific process skills. Third question was at the understanding level of the cognitive domain according to Bloom's taxonomy and aims to measure decide skills from experimental scientific process skills. When the heat maps and eye splash movements of the high-scoring student were examined, it was seen that the student with a high score was undecided between two similar distractors, but the low-scoring student

looked at all distractors. The results may be interpreted as that successful students in 2023 high school entrance exam perform better on questions that require using decide skills from experimental scientific process skills and inference skills from causal scientific process skills. It was reported that students with prior knowledge were integrated text and graphic information, and students without prior knowledge had difficulty integrating scientific diagrams with explanatory texts (Ho et al., 2014). Questions mentioned above involved figures and their solving required the use of causal and experimental process skills. High score students solve many problems and gain practicality while preparing for the exam. Jian (2021) reported that the participants who read the article and performed the practice wrote the solution of the problem in a short time. According to the theory of embodied cognition, practice can facilitate the task (Shapiro, 2011). Successful problem solvers have high-level cognitive strategies. They can recognize and focus on relevant cues when solving problems. Unsuccessful problem solvers have difficulty separating relevant factors from irrelevant, grasping purpose, and focusing on relevant factor (Tsai et al., 2012).

No correlation was found between time to first fixation and duration of first fixation and the practice test scores for all questions. However, negative significant relationships were determined between all other visual measurement results and practice test scores. According to this result, it may be said that students with high scores in the test focus and visit less to solve the questions. Additionally, no relationship was found between the practice test score and visual measurement results for the second and fifth questions. The questions that had a relationship between at least one visual measurement (average duration of fixation) and the practice test score were the third, fourth and seventh questions. For the sixth question, there was a negative relationship between at least two visual measurements (average duration of fixation and total duration of fixation) and the practice test score. A correlation was found between at least four visual measurements and practice test scores for the first, eighth and ninth questions. All significant relationships between the results of visual measurements and practice test scores are negative. These results were consistent with Coşguner (2022)'s study results. Jian (2021) reported that participants who read the easy article spent longer on the application (preparing the pulley system). Researchers found that university students studying in upper grades had better performance when making quick decisions in mental problem-solving processes (Lindner et al., 2014; Rodemer et al., 2020). It was determined that these students frequently transitioned between images related to reaction mechanisms (Rodemer et al., 2020). The results of the current study revealed that students who had prior knowledge, prepared for the exam, and solved different skill-based question types spent less time and cognitive effort when making quick decisions in problem-solving processes.

No correlation was found between duration of first fixation and average duration of visit and perception levels for all questions. However, significant negative relationships were found between the perception level and at least one visual measurement results for all questions. It was determined that as the level of perception towards solving high schools entrance exam questions increased, students' the fixation and visit duration or number decreased. Another result of the study was that negative significant relationships were found between average duration of fixation and self-efficacy, one of the sub-dimensions of perception, for all questions. A similar relationship existed between attitude and willingness sub-dimensions and

average duration of fixation. However, this relationship was for some questions, not for all questions. On the other hand, there were negative significant relationships between average/total duration of visit and the sub-dimensions of perception. However, these relationships were not at the level of average/total duration fixation. Not every student may have access to the solution codes of skill-based science questions that require more time and cognitive effort. This situation negatively affects students' self-efficacy, attitude and willingness towards solving skill-based science questions. Karabulut et al. (2022) reported that students were more willing and had higher perceptions of success when solving acquisition-based questions, and their perception of failure was higher when solving skill-based questions.

Finally, it was revealed that visual measurement results together did not have a predictive effect on practice test scores. However, it was found to have a predictive effect on perception and self-efficacy, one of perception sub-dimensions. Average duration of fixation was a significant predictor of students' perceptions towards solving skill-based science problems. Coşguner (2022) reported that time- and count-oriented measures (average duration of fixation, time to first fixation and number of fixation) were significant predictors of practice test scores. Kragten et al. (2015) found that successful/less successful students who learned process diagrams differed in various learning activities. Additionally, the results showed that understanding the process arrows and the average duration of fixation on the core field was predictive of students' understanding levels.

## 5 Implications

This study revealed visual measurement results of students' processes of solving skill-based science questions. It was revealed with concrete evidence that the cognitive loads of the questions, supporting the question with visuals, the targeted skill levels, presence of unfamiliar vocabulary and complexity of options were effective in the visual measurement results. Item analyses of skill-based questions are based on classical test theory and item response theory. The difficulty and discrimination levels of the questions and the functionality of the options are determined by these analyses. It is suggested that visual measurement results may be taken into account in addition to classical test theory and item response theory in the evaluation of test items. It is thought that the results will make significant contributions to science teachers and science educators.

The current research revealed eighth grade students' processes of solving skill-based science questions with visual measurements. Visual measurement results were compared according to gender and 2023 high schools entrance exam. However, in these comparisons, each question was considered as a whole. Future studies might examine visual measurement results for parts of each question (options, question root, explanatory text and images).

The current study found negative significant relationships between visual measurement results and practice test scores and perception levels (for self-efficacy, attitude and willingness). Solving skill-based science questions requires more time and cognitive effort than acquisition comprehension questions. The results suggested

that students should be willing to solve these questions, have a positive attitude, and have a high level of self-efficacy. It is expected from science teachers to increase the level of student perception towards solving skill-based science questions.

Research results showed that visual measurements together did not have a predictive effect on practice test scores, but had a predictive effect on perception and self-efficacy, one of the sub-dimensions of perception. This hypothesis may be tested in future studies with more participants.

The length and complexity of skill-based questions can reduce students' interest and motivation in the questions and therefore in the course. It is recommended to design questions that are free of external cognitive load, taking into account the cognitive load caused by the length and complexity of skill-based questions.

## Appendix: Sample questions

**Q1.** Elements in the periodic table, some of which are given in the figure; They are shown by scanning with different patterns according to their status as metal, non-metal, semi-metal and noble gas.



Triple groups are formed with an element selected from the periodic table and two neighboring elements in the same group and period as this element.

**Accordingly, which of the following three groups does not meet this condition?**

A)	Semi-metal	
	metal	Semi-metal

B)	nonmetal	
	metal	metal

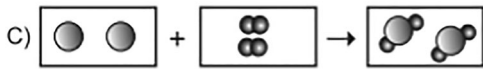
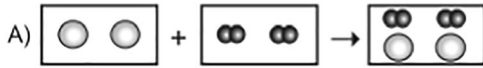
C)	Semi-metal	nonmetal
		Semi-metal

D)		noble gas
	metal	noble gas

**(2020 high school entrance exam)**

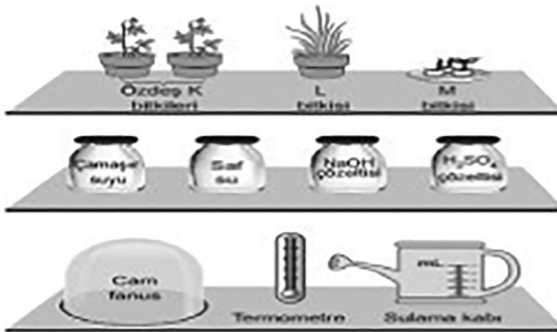
**Q5.** During a chemical reaction, new bonds are formed or existing bonds are broken between atoms or molecules.

**Accordingly, which of the models below does not show the chemical reaction?**



(2018 high school entrance exam)

**Q7.** They want to conduct an experiment to observe the effect of acid rain on plants.



In this experiment, two mechanisms are prepared by selecting suitable plants and materials as shown in the figure. The selected liquids are given to the plants from above with a watering can, like rain, and the observation results are compared.

**Accordingly, which of the following plants and materials were used in the mechanisms?**

- A) L plant and same K plants, bleach, mixture consisting of equal amounts of  $H_2SO_4$  and NaOH, glass jar
- B) K plant, M plant, NaOH solution, pure water
- C) Same K plants,  $H_2SO_4$  solution, pure water
- D) L plant, M plant, equivalent amount of  $H_2SO_4$  solution, mixture consisting of pure water and NaOH, thermometer, glass bell jar

(2018 high school entrance exam)

**Acknowledgements** This study is a part of the first author's master thesis. This study was supported by Bartın University Scientific Research Projects Coordination Unit (2023-SOS-CY-002). The authors are grateful to Bartın University. I also thank Education and Information Technologies journal' editor and reviewers and Dr F.G. Karaođlan Yılmaz, Dr B. Ekiz Kıran, Dr N. İlhan and Dr Ö. Gün for their useful suggestions and discussions.

**Funding** Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This study is a part of the first author's master thesis. This study was supported by Bartın University Scientific Research Projects Coordination Unit (2023-SOS-CY-002). The authors are grateful to Bartın University.

**Data availability** Not applicable.

**Code availability** Not applicable.

## Declarations

**Ethics approval** Data was collected from human participants in the study. All ethical standards were taken into account and followed during the research. Ethics committee approved the study, and its approval number is 2022-SBB-0435.

**Consent to participate** Not applicable.

**Consent to publication** Not applicable.

**Conflicts of interest/competing interests** The authors declare that they have no conflict of interest. The authors have no financial or proprietary interests in any material discussed in this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Avcı, B. (2010). Investigation of educational softwares usability with the methods of eye tracking and think aloud. *Master's thesis, Marmara University, İstanbul*
- Aykaç, N., & Atar, E. (2014). Geçmişten günümüze ilköğretimden ortaöğretime geçiş sisteminin değerlendirilmesi. *Akdoğan Bulut-İnsan, A. ve Yavuz-Akengin, A (Eds.). International Congress of Passing among levels and new models in education from the establishment of the republic to the present day*, 83–104.
- Çakır, Z. (2019). TEOG, LGS and PISA science questions analysis and comparison. Master's Thesis, *Uşak University, Uşak*
- Çepni, S. (2019). *PISA ve TIMSS mantığını ve sorularını anlama*. Pegem A Yayıncılık.
- Çepni, S., Ayas, A., Johnson, D., & Turgut, M. F. (1997). *Fizik öğretimi: Milli eğitimi geliştirme projesi*, Ankara.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332.



- Clinton, V., Cooper, J. L., Michaelis, J. E., Alibali, M. W., & Nathan, M. J. (2017). How revisions to mathematical visuals affect cognition: Evidence from eye tracking. *Eye-tracking technology applications in educational research* (pp. 195–218). IGI Global.
- Coşguner, T. (2022). *Investigation of the relations between the measurements obtained from the eye-tracking method and the test and item statistics*. Doctoral Thesis, Hacettepe University.
- Cowen, L., Ball, L. J., & Delin, J. (2002). An eye movement analysis of web page usability. In X. Faulkner, J. Finlay, & F. Détienne (Eds.), *People and computers XVI - memorable yet invisible*. Springer. [https://doi.org/10.1007/978-1-4471-0105-5\\_19](https://doi.org/10.1007/978-1-4471-0105-5_19)
- Deveci, D., Eroğlu, D., & Bektaş, Z. (2023). Levels of 7th and 8th grade students solving skill-based Turkish questions and affecting factors. *Medeniyet Eğitim Araştırmaları Dergisi*, 7(1), 17–32.
- Eitel, A., Scheiter, K., Schüller, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48–63.
- Erden, B. (2020). Teachers' views related to skill-based questions in Turkish, mathematics and science lessons. *Academia Eğitim Araştırmaları Dergisi*, 5(2), 270–292.
- Eurypedia (2013). *European encyclopedia on national education systems*. <https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php?title=Countries>. Accessed 2 Jan 2024.
- George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step*. Routledge.
- Germann, P. J., Aram, R., & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, 33(1), 79–99.
- Gür, B., Çelik, S. Z., & Coşkun, İ. (2013). *Türkiye'de ortaöğretimin geleceği: Hiyerarşi mi, eşitlik mi?* (vol. 69, pp. 1–28). SETA Siyaset, Ekonomi ve Toplum Araştırmaları Vakfı. [http://file.setav.org/Files/Pdf/20130802120003\\_ortaoogretim\\_analiz2.pdf](http://file.setav.org/Files/Pdf/20130802120003_ortaoogretim_analiz2.pdf). Accessed 12 Jan 2023.
- Hansen, S. J. R., Hu, B., Riedlova, D., Kelly, R. M., Akaygun, S., & Villalta-Cerdas, A. (2019). Critical consumption of chemistry visuals: Eye tracking structured variation and visual feedback of redox and precipitation reactions. *Chemistry Education Research and Practice*, 20, 837–850.
- Ho, H. N. J., Tsai, M. J., Wang, C. Y., & Tsai, C. C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International Journal of Science and Mathematics Education*, 12, 525–554.
- Hodson, D. (1992). In search of a meaningful relationship: An exploration of some issues relating to integration in science and science education. *International Journal of Science Education*, 14(5), 541–562.
- Jacob, R. J., and Karn, S. K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. H. Radach, and H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Elsevier Science.
- Jian, Y. C. (2016). Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements. *Reading Research Quarterly*, 51(1), 93–109.
- Jian, Y. C. (2021). Influence of science text reading difficulty and hands-on manipulation on science learning: An eye tracking study. *Journal of Research in Science Teaching*, 1–25. <https://doi.org/10.1002/tea.21731>
- Jian, Y. C., & Ko, H. W. (2017). Influences of text difficulty and reading ability on learning illustrated science texts for children: An eye movement study. *Computers and Education*, 113, 263–279.
- Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological testing: Principles, applications, and issues* (4th ed.). Thomson Brooks/Cole Publishing Co.
- Karabulut, H., Tosunbayraktar, G., & Kariper, İ. A. (2022). Investigation of secondary school students' opinions about skill-based (next generation) science questions. *Educacione*, 1(2), 301–320.
- Karaoğlan-Yılmaz, F. G., & ve Yılmaz, R. (2019). 2018 KPSS eğitim bilimleri sınavında öğretim teknolojisi ve materyal tasarımı kapsamında sorulan soruların göz izleme yöntemi ile incelenmesi. *III. International Congress on Science and Education*. Afyonkarahisar, Mart 2019.
- Ke, F., Liu, R., Sokolij, Z., & Dahlstrom-Hakki, I. (2024). Using eye-tracking in education: Review of empirical research and technology. *Educational Technology Research and Development*. <https://doi.org/10.1007/s11423-024-10342-4>
- Kızılcapan, O., & Nacaroğlu, O. (2019). Science teachers' opinions about central exams. *Nevşehir Hacı Bektaş Veli University Journal of SSI*, 9(2), 701–719.
- Kline, R. B. (2011). *Principals and practice of structural equation modeling*. The Guilford Press.
- Kragten, M., Admiraal, W., & Rijlaarsdam, G. (2015). Students' learning activities while studying biological process diagrams. *International Journal of Science Education*, 37(12), 1915–1937.

- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring from 2000 to 2012. *Educational Research Review*, *10*, 90–115.
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology*, *30*, 195–232.
- Lindner, M. A., Eitel, A., Thoma, G. B., Dalehefte, I. M., Ihme, J. M., & Koller, O. (2014). Tracking the decision-making process in multiple choice assessment. *Applied Cognitive Psychology*, *28*(5), 738–752.
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–76). Elsevier.
- Malcı, E. (2021). *Examining 10th grade students' problem solving processes in geometry using eye tracking technology*. Master's thesis, Middle East Technical University.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52.
- McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry* (6th ed.). Pearson.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching*. Upper Prentice Hall. Pearson Education India.
- MoNE (2018). 2018 high school transition system. *Education Analysis and Evaluation Reports Series*, *3*. [https://www.meb.gov.tr/meb\\_iys\\_dosyalar/2018\\_12/17094056\\_2018\\_jgs\\_rapor.pdf](https://www.meb.gov.tr/meb_iys_dosyalar/2018_12/17094056_2018_jgs_rapor.pdf). Accessed 6 Jan 2023.
- MoNE (2019). Central examination for secondary education institutions in 2019. *Education Analysis and Evaluation Reports Series*, *7*. [https://www.meb.gov.tr/meb\\_iys\\_dosyalar/2019\\_06/24094730\\_2019\\_ortaogretim\\_kurumlarina\\_iliskin\\_merkezi\\_sinav.pdf](https://www.meb.gov.tr/meb_iys_dosyalar/2019_06/24094730_2019_ortaogretim_kurumlarina_iliskin_merkezi_sinav.pdf). Accessed 6 Jan 2023.
- MoNE (2020). *TIMSS 2019 Türkiye ön raporu* (Eğitim Analiz ve Değerlendirme Raporları Serisi 15). MEB Yayınları. Retrieved from [https://www.meb.gov.tr/meb\\_iys\\_dosyalar/2020\\_12/08202713\\_No15\\_TIMSS\\_2019\\_Turkiye\\_On\\_Raporu.pdf](https://www.meb.gov.tr/meb_iys_dosyalar/2020_12/08202713_No15_TIMSS_2019_Turkiye_On_Raporu.pdf). Accessed 6 Jan 2023.
- Myers, B. E., Washburn, S. G., & Dyer, J. E. (2004). Assessing agriculture teachers' capacity for teaching science integrated process skills. *Journal of Southern Agricultural Education Research*, *54*(1), 74–85.
- National Research Council. (1996). *National science education standards*. National Academy Press.
- Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research*, *13*(6), 1–15.
- Osborne, J., Sedlacek, Q. C., Friend, M., & Lemmi, C. (2016). Learning to read science. *Science Scope*, *40*(3), 36–42.
- Özdemir, D. (2013). *A case study of problem solving in eye-tracking*. Master of Science, Middle East Technical University, Ankara.
- Özdoğan, F. B. (2008). A conceptual study on eye tracking and its applications in marketing abstract. *Gazi University Journal of Commerce and Tourism Education Faculty*, *2*, 134–147.
- Özer, E., & Özdemir, S. (2022). Eye tracking technique from past to present in reading research. *Ankara University Faculty of Educational Sciences Journal of Special Education*, *23*(3), 675–697.
- Özer-Özkan, Y. (2014). A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT. *International Journal of Human Sciences*, *11*(1), 20–44.
- Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford University Press.
- Patterson, A., Roman, D., Friend, M., Osborne, J., & Donovan, B. (2018). Reading for meaning: The foundational knowledge every teacher of science should have. *International Journal of Science Education*, *40*(3), 291–307.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Pub.
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*(3), 577–598.
- Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. In R. Radach, A. Kennedy, & K. Rayner (Eds.), *Eye movements and information processing during reading* (pp. 3–26). Psychology Press.
- Rayner, K. (1998). Eye movements and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension process in reading. *Scientific Studies of Reading*, *10*(3), 241–255.

- Reed, J. B., & Meyer, R. J. (2007). Edmund Burke Huey (1870–1913): A brief life with an enduring legacy. In S. E. Israel, & E. J. Monaghan (Eds.), *Shaping the reading field: The impact of early reading pioneers, scientific research, and progressive ideas* (pp. 159–175). International Reading Association.
- Regis, A. P., Albertazzi, G., & Roletto, E. (1996). Concept maps in chemistry education. *Journal of Chemical Education*, 73(11), 1084–1088.
- Rodemer, M., Eckhard, J., Graulich, N., & Bernholt, S. (2020). Decoding case comparisons in organic chemistry: Eye tracking students' visual behavior. *Journal of Chemistry Education*, 97(10), 3530–3539.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21(4), 589–619.
- Şan, S., & İlhan, N. (2022). Theoretical and conceptual framework for science course skill-based questions (next generation). *Inonu University Journal of Graduate School of Education*, 9(17), 1–20.
- Shapiro, L. (2011). *Embodied cognition*. Routledge.
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328(5977), 450–452.
- Sprengrer, P., & Benz, C. (2020). Children's perception of structures when determining cardinality of sets—results of an eye-tracking study with 5-year-old children. *ZDM*, 52(4), 753–765.
- Sweeder, R. D., Herrington, D. G., & VandenPlas, J. R. (2019). Supporting students' conceptual understanding of kinetics using screencasts and simulations outside of the classroom. *Chemistry Education Research and Practice*, 20, 685–698.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). Springer.
- Tabachnick, B. G., & Fidell, L. (2015). In M. Baloğlu, & Çev (Eds.), *Using multivariate statistics [Çok değişkenli istatistiklerin kullanımı]*. Nobel Akademik Yayıncılık.
- 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 assessment. *International Journal of Research & Method in Education*, 29(2), 185–208.
- Teo, T. W., & Peh, Z. Q. (2023). An exploratory study on eye-gaze patterns of experts and novices of science inference graph items. *STEM Education*, 3(3), 205–229.
- Tinker, M. A. (1958). Recent studies of eye movements in reading. *Psychological Bulletin*, 55(4), 215–231.
- Tonbuloğlu, İ. (2010). Usability, test of primary seventh grade mathematics instructional software with eye tracking and, video recording methods. *Master's Thesis, Marmara University, İstanbul*
- Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58, 375–385.
- Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare word problems: As eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, 84, 85–94.
- Wade, N. J., Tatler, B. W., & Heller, D. (2003). Dodge-ing the issue: Dodge, Javal, Hering, and the measurement of saccades in eye-movement research. *Perception*, 32(7), 793–804.
- Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. McGraw-Hill Education (UK).
- Wu, C. J., Liu, C. Y., Yang, C. H., & Wu, C. Y. (2021). Children's reading performances in illustrated science texts: Comprehension, eye movements, and interpretation of arrow symbols. *International Journal of Science Education*, 43(1), 105–127.
- Yalçın, E. (2019). *Analysing high school entrance examination in terms of administrators, teachers and students' parents*. Master of Science, Akdeniz University, Department of Educational Sciences, Antalya.
- Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher order thinking (HOT) in education. *The Online Journal of Distance Education and e-Learning*, 3(2), 41–47.
- Yiğit, N., Deveci, İ., & Dadandı, N. (2022). Development of the perception scale towards the next generation science questions. *YYU Journal of Education Faculty February*, 108–130.
- Zoller, U. (2000). Teaching college science towards the next millennium: Are we getting it right? *Journal of College Science Teaching*, 29, 409–414.