Investigation of the Relationship between Visual Measurements and Test and Item Statistics

Tolga COŞGUNER[1*], Burcu ATAR[2]

[1]Pamukkale University, Faculty of Education, Department of Education Measuring and Evaluation, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Education Measuring and Evaluation, Türkiye

Abstract: *The purpose of this study was to investigate the correlations between time- and count-oriented visual data collected from a wearable eye tracker and test and item statistics. To accomplish this, a wearable eye tracker was used with an achievement test that included 30 multiple-choice questions assessing eighth-grade achievements, administered to 60 high school students. The acquired data were then used to compute t-values, Pearson correlation coefficients, and regression coefficients. The results of the study demonstrated a substantial inverse relationship between time and count-oriented visual assessments and achievement scores. Furthermore, time and count-oriented measures were shown to be significant predictors of achievement test results. Additionally, substantial disparities in times to first fixation and total durations of visit were identified in favor of females, whereas significant differences in average durations of visit were discovered in favor of males. The results of this study show that there is a substantial association between time-oriented measures and test and item statistics.*

1. Introduction

When the theoretical preferences of the test development process are analyzed from a psychometric perspective, it is evident that traditional paper-and-pencil tests and self-report-based measurement methods have various structural limitations. It is stated that supporting such methods with physiological measurements in order to obtain meaningful outputs of cognitive structures and to create a complete picture offers important opportunities for the measurement process. In this sense, physiological measurements are associated with automatic responses triggered by the brain but emerging in the body, and even subtle implicit cognitive structures can create fluctuations that prepare the body for action and continue throughout the nervous system. Such outputs of implicit cognitive processes are known to contain psycho-somatic findings. Therefore, it is stated that all kinds of physiological information about cognitive features serve as objective evidence for each of the implicit cognitive structures (Becker &

Menges, 2013; Duchowski, 2002; Korkmaz, 2017; Lai *et al*., 2013; Peterson *et al*., 2015; Shayan *et al*., 2017).

Considering that it takes 150 milliseconds for a symbol to be processed by our brain and 100 milliseconds to make sense of it, it is clear that visual information plays a very important role psycho-somatic measurement (Durna & Arı, 2016b). Analyzing the cognitive and behavioral bases of eye movements within the physiological organization of the optical system reveals very strong findings in order to understand implicit cognitive structures (Duchowski, 2007). Accordingly, psycho-physical and neuro-psychological data obtained from visual measurements provide quantitative and qualitative evidence of participants' perceptual, cognitive, and behavioral responses (Duchowski, 2002; Liu *et al*., 2011; Shayan *et al*., 2017). In this context, neuro-cognitive research, on the other hand, provides objective findings regarding how visual information is processed in the brain. These findings reveal the mechanics of basic cognitive features (Galván, 2010; Goldberg & Helfman, 2011; Jacob & Karn, 2003; Lai *et al*., 2013; Liversedge & Findlay, 2000; Rayner *et al*., 2006). Therefore, recording and analyzing eye movement data is expected to provide us with strong indicators of the cognitive properties underlying visual information (Conati & Merten, 2007; Mele & Federici, 2012; Ömur & Görgülü Aydoğdu, 2017; Rayner, 1998; Rosa, 2015). At this point, eye-tracking technology offers important and versatile contributions to obtain reliable visual measurements. With this innovative method that can provide accurate measurements, detailed analytical findings on cognitive skills can be obtained (Ariasi & Mason, 2011; Conati & Merten, 2007; Just & Carpenter, 1976a, 1976b, 1980; Lai *et al*., 2013; Mele & Federici, 2012; Rayner, 1998, 2009).

The current idea underlying neuroscientific studies focused on such analytical findings in the visual context is that our cognitive acts refer to reflect the notion that we think through our bodies (Antle, 2013; Kirsh, 2013; Rodrigues & Rosa, 2017; Shayan *et al*., 2017). The objects that students create in their imagination based on their own interactions are very important for educational researchers. This is because when established automatic perceptual actions are transformed into general conceptual understandings, they can be interpreted as explanatory of the cognitive mechanism (Shayan *et al*., 2017). Temporal, spatial, and count-oriented analytical measurements obtained by the eye-tracking method are seen as a new level of research on students' cognitive and perceptual processes, and the measurements obtained provide strong visual indices in a psychometric context (Duchowski, 2007; Goldberg & Kotval, 1999; Jacob & Karn, 2003; Knight *et al*., 2014; Lai *et al*, 2013; Liversedge *et al*., 1998; Liversedge & Findlay, 2000; Tobii Pro, 2019).

With the visual indices created to examine, comprehensively and systematically define and characterize the cognitive actions performed, the regions where attention is directed and the differences in this process can be determined in detail. Thus, more sensitive, different and reliable measurements can be obtained compared to traditional assessment methods (Clinton *et al*., 2017; Knight *et al*., 2014; Lai *et al*., 2013). With the measurements obtained, findings on a wide range of learning outcomes such as information processing models, effects of instructional designs, re-examination of existing theories, individual differences, effects of learning strategies used, decision-making processes and conceptual development can be revealed. In order to obtain meaningful outcomes for cognitive skills, it is a psychometric necessity to employ consistent observations (Lai *et al*., 2013). Experimental observations conducted with the eye-tracking method provide rich psychometric evidence for the objective measurement of cognitive outcomes beyond student responses to test items (Zentall & Junglen, 2017).

1.1. Visual Measurements and Psychometric Properties

Visual measurements, which play a significant role in revealing the relationship between the situational responses of the eyes and the components of the cognitive process, are first analyzed in terms of time. It is stated that they provide detailed visual indices of the time spent on specific areas of interest and answer the questions of "when?" and "for how long?" related to cognitive processing (Clinton *et al.,* 2017; Just & Carpenter, 1976a, 1976b, 1980; Lai *et al*., 2013; Liversedge *et al*., 1998; Mele & Federici, 2012; Rayner, 1998, 2009; Rayner *et al*., 2006). On the other hand, count-oriented visual measures, which have a relatively quantitative usage area and are obtained from observations based on frequency, are reported to provide versatile visual indices regarding the significance of the visual material of interest, especially in terms of situational responses (Lai *et al*., 2013). In addition to these, spatial measures including positions, distances, directions, sequences, and spatial arrangements are reported to answer the "where?" and "how?" questions related to cognitive processing and provide sensitive visual indices regarding the spatial positions of eye movements. Therefore, it is important to measure spatial eye movements and scanning behaviors in order to reveal the control of visual search and selective processes in visual perception (Liversedge & Findlay, 2000).

1.1.1. Visual Measurements of Situational Reactions

When the temporal, spatial, and count-oriented visual measurements are contextually analyzed, it is observed that these metrics provide contextual data for the focus on and visits to the defined areas of interest, and with these measurements, detailed visual indices (measurements) can be created for the situational responses that occur in these areas (Duchowski, 2007; Goldberg & Kotval, 1999; Jacob & Karn, 2003; Lai *et al*, 2013; Liversedge *et al*., 1998; Liversedge & Findlay, 2000; Tobii Pro, 2019). Accordingly, the focusing process, which is closely related to our ability to visually encode spatially distributed information, presents multifaceted indicators for the cognitive operations performed. The area where focusing occurs provides visual evidence for the information being processed. It is stated that multiple instances of focusing on the same spatial area are an indication of indecision in understanding an encountered object (Bayazıt, 2013; Jacob & Karn, 2003; Just & Carpenter, 1976a, 1976b, 1980). Therefore, it is stated that the length of focusing time is an indicator of the difficulty experienced in processing the objects encountered, and at the same time, these measurements constitute visual evidence of the difficulty of obtaining information from areas of interest. Hence, while the length of the focusing time indicates that the information obtained from these areas is difficult to understand and/or requires more cognitive effort, the short duration of this time indicates that the information obtained from areas of interest is easier to understand (Akçay & Altun, 2019; Bayazıt, 2013; Just & Carpenter, 1976a, 1976b, 1980; Liversedge *et al*., 1998; Majooni *et al*., 2016; Schwonke *et al*., 2009).

While the total duration of fixation obtained from the defined areas of interest is associated with the difficulties experienced in processing the encountered objects, these measurements also provide the amount of visual attention directed to the area of interest. In this context, the average duration of fixation obtained from the specified areas of interest is associated with the difficulty level of the text of interest, while time to first fixation and duration of first fixation are associated with the selection and/or organization of the information in areas of interest. Time to first fixation and duration of first fixation are also indicators of the time spent accessing information (Akçay & Altun, 2019; Bayazıt, 2013; Just & Carpenter, 1976a; Liu, 2014; Liversedge *et al*., 1998; Majooni *et al*., 2016; Negi & Mitra, 2020; Rayner, 1998, 2009; Rayner et al., 2006; Scheiter & Eitel, 2015; Scheiter & Eitel, 2017; Schwonke *et al*., 2009; Underwood *et al*., 1990; van Meeuwen *et al.,* 2014). However, a number of fixations associated with the level of importance shown to the defined areas of interest are evaluated as an indicator of

cognitive load, and are also seen as the processing level of information acquired after attention is shown to the area of interest. Therefore, it is stated that an increase in the number of fixations in the search tasks carried out in areas of interest is also an indication of an ineffective search process in these tasks (Akçay & Altun, 2019; Bayazıt, 2013; Goldberg & Kotval, 1999; Jarodzka *et al.*, 2015; Lai *et al.*, 2013; Scheiter & Eitel, 2017; Tabbers *et al.*, 2008; Poole *et al.*, 2005).

Visits, which are defined as visual actions performed in the period from the first fixation on the defined areas of interest to the end of the last fixation on these areas, provide multifaceted indicators of the cognitive processes carried out (Tobii Pro, 2019). It is seen that the number of visits, which indicates how many times the areas of fixation are revisited and/or the intensity of attention in these areas, is associated with the level of importance given to the areas of interest. It is observed that durations of visits, defined as the time spent looking at a specific object or visual in the areas of interest defined are associated with the lack of understanding and/or comprehension in these areas. It is predicted that, in the event of such a deficiency, long and repeated visits focused specifically on these areas may occur (Lai *et al.*, 2013; Saleem *et al.*, 2021; Tobii Pro, 2019). Indeed, while the number of visits obtained from repeated visits is thought to be an indicator of the process of consciously acquiring information after involuntary fixation, the total and average values of duration of visits are generally considered an indicator of lack of understanding and/or comprehension (Akçay & Altun, 2019; Klein *et al.*, 2020; Saleem *et al.*, 2021; Scheiter & Eitel, 2017). On the other hand, it is stated that heat maps, which are generated by coloring visually concentrated areas, especially using view data, mostly provide the opportunity to observe visual trends between regions (Bayazıt, 2013; Tobii Pro, 2019).

As a result, when visual measurements are examined and obtained by focusing on time, location, and counting in a psychometric context, it is seen that these measurements can provide important information about situational reactions. With the information collected, valid estimates can be made for test and item statistics (Bayazıt, 2013; Berzak *et al.*, 2018; Lindner *et al.*, 2017; Liu *et al.*, 2017; Nugrahaningsih *et al.*, 2013; Saß *et al.*, 2017; Solheim & Uppstad, 2011; Tai *et al.*, 2006; Tsai, *et. al*, 2011). Therefore, the aim of this study is to examine the relationships between time and count-oriented visual measurements obtained from multiple-choice tests and the test and item statistics of these tests. For this purpose, the data obtained were analyzed in the light of the sub-problems identified within the scope of the research.

2. Method

2.1. Research Design

This research, which aims to examine the relationships between visual measurements obtained from multiple-choice tests and test and item statistics of these tests, is structured according to the correlational research method, one of the quantitative research approaches (Fraenkel *et al.*, 2012).

2.2. Participants

It is accepted by many researchers that the minimum acceptable sample size for correlational studies cannot be smaller than 30. It is stated that data obtained from a sample smaller than 30 may yield an incorrect relationship degree, while a sample larger than 30 is more likely to produce meaningful results (Fraenkel *et al.*, 2012). In regression analyses, which are frequently preferred in correlational studies, the ratios of cases to independent variables, in particular, should be reasonable, otherwise, the solutions obtained may turn out to be perfect but meaningless (Tabachnick & Fidell, 2015). Based on this, the working group was determined as 60 people, considering the existing limitations and the stated recommendations. 30 of these

participants, who were selected from students who study in the classroom and do not wear glasses, are girls, and 30 are male students.

2.3. Data Collection Tools

2.3.1. Eighth Grade Achievement Test

This test, which aims to measure the achievements of Turkish, mathematics and science courses in the LGS (transition system to high schools) central exam, consists of a total of 30 multiple-choice test items, 18 of which are verbal (Turkish) and 12 of which are numerical (mathematics and science). The KR-20 reliability coefficient calculated for the verbal section of this test, which was developed based on Classical Test Theory, on a total of 4848 students is .71, the KR-20 reliability coefficient calculated for the numerical section is .80, and the KR-20 reliability coefficient calculated for the entire test is .84. Table 1 includes the findings regarding the item statistics calculated during the development phase of the Eighth Grade Achievement Test.

Table 1. *Eighth Grade Achievement Test Item Statistics*

	p_j	s	r_{jx}	r_j
Ver.1.i	.61	.49	.18	.09
Ver.2.i	.79	.41	.50	.21
Ver.3.i	.50	.50	.30	.15
Ver.4.i	.57	.49	.24	.12
Ver.5.i	.91	.28	.58	.16
Ver.6.i	.86	.35	.52	.18
Ver.7.i	.72	.45	.51	.23
Ver.8.i	.81	.39	.52	.20
Ver.9.i	.88	.33	.57	.19
Ver.10.i	.60	.49	.41	.20
Ver.11.i	.51	.50	.29	.15
Ver.12.i	.90	.30	.57	.17
Ver.13.i	.86	.35	.55	.19
Ver.14.i	.80	.40	.51	.21
Ver.15.i	.60	.49	.33	.16
Ver.16.i	.65	.48	.37	.18
Ver.17.i	.54	.50	.29	.15
Ver.18.i	.81	.39	.50	.20
Num.1.i	.70	.46	.62	.28
Num.2.i	.48	.50	.54	.27
Num.3.i	.74	.44	.52	.23
Num.4.i	.58	.49	.60	.29
Num.5.i	.58	.49	.56	.28
Num.6.i	.52	.50	.54	.27
Num.7.i	.66	.47	.59	.28
Num.8.i	.43	.49	.52	.26
Num.9.i	.78	.42	.50	.21
Num.10.i	.78	.41	.50	.21
Num.11.i	.83	.37	.58	.21
Num.12.i	.81	.39	.55	.21

N=4848, p_j: item difficulty index, *s*: item standard deviation, r_{jx}: item discrimination index, r_j: item reliability index

2.3.2. Wearable Eye-Tracker

The Tobii ProGlasses 2 wearable eye-tracking device was used to collect data on the eye movements of students participating in the study. This device, which consists of two parts—the head unit (glasses) and the recording unit, can take at least 100 samples per second. The device detects eye movements using the corneal reflection and dark pupil technique. This device, which has a 4-camera system that can record both eyes, uses 4 cameras and 12 infrared illuminators to perform eye tracking.

2.4. Data Collection Procedures

In this part of the research, an application was initially planned for the final form of the Eighth Grade Achievement Test; however, this application had to be postponed due to the suspension of formal education because of the pandemic. The postponed application was then conducted on a total of 60 students selected from among the 9th-grade high school students (who were studying in the 8th grade during the planning phase) who did not wear glasses. During these 40-minute face-to-face applications in a real classroom environment, students used the Tobii Pro Glasses 2 wearable eye-tracking device. Immediately before using the eye-tracking device, a single-point calibration was performed on the students, and then video recordings of their gaze fields were taken. These collected recordings were then analyzed with the Tobii Pro Lab analysis program, and then visual data regarding the answering reactions of the students who took the exam were obtained, along with the results related to the Eighth Grade Achievement Test application.

2.5. Data Analysis

A two-stage analysis was carried out for the data obtained from different types of observations made within the scope of the research.

2.5.1. First Stage

At this stage, analyses were conducted in order to obtain data on visual responses given by students while answering the test, based on the records collected with a wearable eye-tracking device during the application of the Eighth Grade Achievement Test. For this purpose, areas of interest were specifically defined for each item in the Eighth Grade Achievement Test, and rectangles were drawn to include the item as a whole without distinguishing between the item stem and the options. In this process conducted with an item focus, duration measurements obtained from the defined area of interest (AOI) were recorded at the millisecond (ms) level. From these collected records, visual measurement values such as time to first fixation (ttff), duration of first fixation (doff), total duration of visit (tdov) including those who did not visit, average duration of visit (adov), number of visits (nov) including those who did not visit, total duration of fixation (tdof) including those who did not fixate, average duration of fixation (adof), number of fixations (nof) including those who did not fixation, total time spent in the area of interest (total time of interest duration), and finally total recording duration were obtained.

2.5.2. Second Stage

In this phase of the research, firstly, whether the visual measurements obtained based on the items showed a significant difference according to gender was examined with an independent samples t-test. In this context, examinations regarding the normality assumption were conducted through the calculated kurtosis and skewness values. Then, the relationships between the obtained visual measurements and the test and item statistics of the applied achievement test were examined using the Pearson correlation coefficient and linear regression analyses. In this context, firstly, assumptions regarding the Pearson correlation coefficient were examined,

and then basic assumptions regarding simple and multiple linear regression analyses were tested. For this purpose, firstly, it was investigated whether there was missing data in the data set, and then extreme values were examined using Cook's and Mahalanobis' distance values. Binary normality assumptions between variables were examined with scatter plot matrices, and multiple normality assumptions were examined with Henze-Zirkler and Mardia's multivariate normality test, and kurtosis and skewness values were calculated for univariate normality assumptions. Linearity and homoscedasticity of residuals were investigated with scatter diagrams, and in deciding on the sample size, the formula of Khamis and Kepler (2010), n ≥ 20+5k, was taken into account. Whether there was a multicollinearity problem among the independent variables was examined using Pearson correlation coefficients, VIF values, condition indices and tolerance values.

3. Findings

In this part of the research, the research question *"Do the duration-based visual measurements obtained from the areas of interest specifically defined for each item in the Eighth Grade Achievement Test differ according to gender?"* was addressed. At this stage, first, each of the item-based duration measurements was compared according to gender across all items, and then only the items that showed significant differences were tabulated. Following the analyses, no significant item-based differences were detected in durations of first fixation, average durations of fixation and total durations of fixation.

Table 2 below includes the findings regarding the verbal item 17, which showed significant differences between genders in times to first fixation.

Table 2. *Comparison of Times to First Fixation by Gender*

Items	Gender	n	$\bar{X}$	Ss	t	sd	p
Ver.17.i	Male	30	45.885	60.250	-3.232	58	.002
	Female	30	121.056	112.226			

According to Table 2, the time to first fixation of the verbal item 17 shows a significant difference between genders, t (58) =-3.232, $p<.05$. The time to first fixation for verbal item 17 among female students ($\bar{X}$=121.056 sec) is longer than that of the male students ($\bar{X}$=45.885 sec). This suggests that the female students have more difficulty focusing on the verbal item 17 compared to the male students, which may be due to the challenges experienced by the female students in selecting, organizing or accessing the information related to this item. The η^2 value calculated specifically for the verbal item 17 is .153. Accordingly, it can be stated that approximately 15% of the variance observed in times to first fixation of the verbal item 17 is is attributable to gender. Verbal item 17, which has a difficulty index (.54) and a discrimination index (.29), can be seen in Appendix A.

Table 3 presents the findings regarding the numerical item 9, which showed a significant difference according to gender in times to first fixation.

Table 3. *Comparison of Times to First Fixation by Gender*

Items	Gender	n	$\bar{X}$	Ss	t	sd	p
Num.9.i	Male	30	35.784	44.500	-2.066	58	.043
	Female	30	63.546	58.618			

In Table 3, it is observed that the time to first fixation of item 9 shows a significant difference between genders, t (58) =-2.066, $p<.05$. The time to first fixation for item 9 among female students ($\bar{X}$=63.546 sec) is longer than that of male students ($\bar{X}$=35.784 sec). In this case, it can be said that female students have more difficulty focusing on item 9 compared to male students.

The η^2 value calculated specifically for item 9 is .069. Accordingly, it can be stated that approximately 7% of the variance observed in times to first fixation of item 9 is related to gender. The difficulty index (.78) and discrimination index (.50) for item 9 for female students can be seen in Appendix B.

Table 4 presents the findings regarding verbal item 8, which showed a significant difference according to gender in average fixation durations.

Table 4. *Comparison of Average Durations of Visit by Gender*

Items	Gender	n	$\bar{X}$	Ss	t	sd	p
Ver.8.i	Male	30	4.927	3.015	2.198	58	.032
	Female	30	3.509	1.841			

According to Table 4, the average duration of visit of item 8 shows a significant difference between genders t (58) =2.198, p<.05. The average duration of visit for verbal item 8 among male students ($\bar{X}$=4.927 sec) is longer than that of female students ($\bar{X}$=3.509 sec). In this case, it can be said that male students' efforts to understand verbal item 8 or their time to go back and reread it and/or their decision-making processes for this item take longer on average compared to female students. The η^2 value calculated specifically for verbal item 8 is .077. Accordingly, it can be stated that approximately 8% of the observed variance in average durations of visits to verbal item 8 is attributable to gender. Verbal item 8, which has a difficulty index (.81) and a discrimination index (.52), can be seen in Appendix C.

Table 5 includes the findings regarding item 12, which shows a significant difference according to gender in total durations of visit.

Table 5. *Comparison of Total Durations of Visit by Gender*

Items	Gender	n	$\bar{X}$	Ss	t	sd	p
Num.12.i	Male	30	45.353	24.630	-2.294	58	.025
	Female	30	60.188	25.458			

In Table 5, it is observed that the total duration of visit for item 12 shows a significant difference between genders t (58) =-2.294, p<.05. Total duration of visit for numerical item 12 among female students ($\bar{X}$=60.188 sec) is longer than that of male students ($\bar{X}$=45.353 sec). In this case, it can be said that female students' efforts to understand numerical item 12 or their time to go back and reread and/or their decision-making processes for this item take longer on average compared to male students. The η^2 value calculated specifically for numerical item 12 is .083. Accordingly, it can be stated that approximately 8% of the observed variance in total durations of visit for numerical item 12 is attributable to gender. For the last item of the Eighth Grade Achievement Test, numeric item 12, female students may have gone back and re-read it to check their decisions regarding this item or they may have gone back to understand this item better. Numerical item 12, which has a difficulty index (.81) and a discrimination index (.55), can be seen in Appendix D.

In this part of the research, the research question *"What are the relationships between the visual measurements obtained from the areas of interest specifically defined for each item in the Eighth Grade Achievement Test and the total test scores obtained from this test?"* was answered. For this purpose, firstly, the skewness (-.81) and kurtosis (-.06) values of the total test scores obtained from the Eighth Grade Achievement Test application were examined, and based on these findings, it was seen that the total test scores did not deviate substantially from the normal distribution. Then, the relationships between fixation measurements of the verbal items in the Eighth Grade Achievement Test and the total test scores were examined, and the findings are presented in Table 6.

Table 6. *Relationships Between Fixation Measures on Verbal Items and Total Test Scores*

	$r_{ttff-tts}$	$r_{doff-tts}$	$r_{nof-tts}$	$r_{adof-tts}$	$r_{tdof-tts}$
Ver.1.i	.213	-.054	.049	-.352**	-.155
Ver.2.i	-.085	-.025	-.155	-.368**	-.332**
Ver.3.i	-.105	.014	.085	-.329*	-.101
Ver.4.i	-.317*	-.033	-.189	-.282*	-.323*
Ver.5.i	.053	.116	-.318*	-.336**	-.371**
Ver.6.i	-.217	-.099	-.153	-.212	-.218
Ver.7.i	-.077	-.328*	-.333**	-.466**	-.354**
Ver.8.i	-.100	-.156	-.191	-.446**	-.303*
Ver.9.i	.195	-.024	-.359**	-.375**	-.455**
Ver.10.i	-.155	-.133	.066	-.333**	-.186
Ver.11.i	-.153	.071	-.142	-.151	-.182
Ver.12.i	-.046	-.070	-.342**	-.229	-.385**
Ver.13.i	-.062	.072	.049	-.328*	-.121
Ver.14.i	.140	-.014	-.122	-.283*	-.212
Ver.15.i	-.271*	-.229	-.281*	-.239	-.341**
Ver.16.i	-.067	.031	.135	-.278*	.082
Ver.17.i	-.160	.091	-.066	-.458**	-.180
Ver.18.i	-.281*	-.160	.036	-.424**	-.089

n=60, *p<.05, **p<.01, *ttff:* times to first fixation, *doff:* durations of first fixation, *nof:* number of fixations, *adof:* average durations of fixation, *tdof:* total durations of fixation, *tts:* total test scores

According to Table 6, there appears to be a strong relationship between average and total durations of fixation and achievement. Accordingly, it is seen that no significant relationships were found between average durations of fixation and total test scores only for items 6, 11, 12, and 15. However, significant relationships were observed in all measurements of the remaining items. Similarly, it is observed that significant relationships were determined between total durations of fixation and total test scores for many items.

Table 7 shows the relationships between the visit measurements and total test scores for the verbal items in the Eighth Grade Achievement Test.

Table 7. *Relationships Between Visit Measures and Total Test Scores for Verbal Items*

	$r_{nov-tts}$	$r_{adov-tts}$	$r_{tdov-tts}$
Ver.1.i	.131	-.150	-.021
Ver.2.i	-.032	-.160	-.289*
Ver.3.i	.073	.083	.052
Ver.4.i	-.047	.059	-.284*
Ver.5.i	-.106	-.345**	-.390**
Ver.6.i	.202	-.406**	-.253
Ver.7.i	-.164	-.372**	-.293*
Ver.8.i	-.125	-.093	-.252
Ver.9.i	-.160	-.023	-.430**
Ver.10.i	.001	.054	-.027
Ver.11.i	-.008	-.113	-.168
Ver.12.i	-.051	-.265*	-.371**
Ver.13.i	.226	-.283*	.019
Ver.14.i	.118	-.035	-.159

Table 7. *Relationships Between Visit Measures and Total Test Scores for Verbal Items (Continued)*

	$r_{nov-tts}$	$r_{adov-tts}$	$r_{tdov-tts}$
Ver.15.i	-.112	-.190	-.343**
Ver.16.i	.211	.090	.288*
Ver.17.i	.169	-.340**	-.151
Ver.18.i	.013	-.086	-.067

n=60, *p<.05, **p<.01, nov: number of visits, adov: average durations of visit, tdov: total durations of visit, tts: total test scores

In Table 7, a positive relationship is observed with r_{16}=.288, $p<.05$ was determined, albeit weak, between the total durations of visits of the 16th item and the total test scores. This result differs from all the relationships observed within the scope of the research. Accordingly, it can be concluded that when the students revisited and examined this specific item, the verbal 16th item, they were more successful. When we look at the concentration and visit measurements collected specifically for the verbal items as a whole, it is seen that the relationships between these measurements and achievement scores are generally in the opposite direction. In this case, it can be predicted that when an increase in the concentration and visit measurements for items for which a relationship is determined, a decrease in achievement scores may also occur.

Table 8 shows the relationships between fixation measurements of the numerical items in the Eighth Grade Achievement Test and the total test scores.

Table 8. *Relationships Between Fixation Measures on Numerical Items and Total Test Scores*

	$r_{ttff-tts}$	$r_{doff-tts}$	$r_{nof-tts}$	$r_{adof-tts}$	$r_{tdof-tts}$
Num.1.i	.157	.048	-.276*	-.111	-.239
Num.2.i	-.092	-.025	-.282*	-.200	-.344**
Num.3.i	-.302*	-.017	-.450**	-.167	-.494**
Num.4.i	-.163	-.088	.099	-.190	.044
Num.5.i	.156	-.011	.017	-.179	-.034
Num.6.i	-.033	-.281*	-.001	-.285*	-.071
Num.7.i	-.136	-.149	-.198	-.170	-.210
Num.8.i	-.058	-.023	-.200	-.038	-.203
Num.9.i	-.096	-.153	-.083	-.067	-.066
Num.10.i	-.027	.006	.034	-.256*	-.003
Num.11.i	.001	.038	-.167	-.074	-.203
Num.12.i	-.079	-.198	-.107	-.189	-.169

n=60, *p<.05, **p<.01, ttff: times to first fixation, doff: durations of first fixation, nof: number of fixations, adof: average durations of fixation, tdof: total durations of fixation, tts: total test scores

According to Table 8, it can be concluded that the relationship between average and total fixation durations for numerical items and achievement is weaker than that for verbal items in terms of the number of items showing a relationship.

Table 9 shows the relationships between the visit measurements of the numerical items in the Eighth Grade Achievement Test and the total test scores.

Table 9. *Relationships Between Visit Measures and Total Test Scores for Numerical Items*

	$r_{\text{nov-tts}}$	$r_{\text{adov-tts}}$	$r_{\text{tdov-tts}}$
Num.1.i	-.200	-.188	-.239
Num.2.i	-.096	-.279*	-.304*
Num.3.i	-.097	-.197	-.464**
Num.4.i	.111	-.182	.122
Num.5.i	.057	-.153	-.022
Num.6.i	.006	-.054	-.084
Num.7.i	-.248	.030	-.155
Num.8.i	-.318*	.118	-.195
Num.9.i	.022	-.139	-.068
Num.10.i	-.142	.163	.075
Num.11.i	.080	-.240	-.198
Num.12.i	.120	-.225	-.109

$n=60$, *$p<.05$, **$p<.01$, *nov:* number of visits, *adov:* average durations of visit, *tdov:* total durations of visit, *tts:* total test scores

Table 9 shows that significant and negative relationships were found between number of visits for item 8 and the total test scores, between average duration of visit for item 2 and the total test scores, and between the total duration of visit for items 2 and 3 and the total test scores. When all the findings obtained are considered in general, it is observed that the number of relationships determined for item-focused items is higher in verbal items than in numerical items.

In this part of the research, the research question *"What is the rate at which the duration-focused visual measurements obtained from the areas of interest specifically defined for each item in the Eighth Grade Achievement Test explain the total test scores obtained from the same test?"* was addressed. For this purpose, multiple linear regression analysis was conducted to explore the connection between the measurements obtained and the total test scores. In this process, the basic assumptions regarding the multiple linear regression analysis were first tested and it was found that these assumptions were met in light of the results obtained. Table 10 shows the rate at which the duration measurements obtained within the scope of the research explain the total test scores.

Table 10. *The Explanation Ratio of Total Test Scores of Duration Measures to the Items in the Eighth Grade Achievement Test*

Variables	B	B_{SH}	β	t	p
Constant	28.999	2.216		13.087	.000***
Times to first fixation	-.054	.026	-.319	-2.095	.041*
Durations of first fixation	28.008	15.576	.379	1.798	.078
Average durations of fixation	-43.202	14.053	-.658	-3.074	.003**
Average durations of visit	.090	.211	.069	0.427	.671
$R=.466$	$R^2=.217$				
$F(4,55)=3.811$	$p=.008**$				

$n=60$, *$p<.05$, **$p<.01$, ***$p<.001$

According to Table 10, a moderate and significant relationship is observed between times to first fixation, durations of first fixation, average durations of fixation, and average durations of visit with the total test scores, $R=.466$, $R^2=.217$, $p<.01$. Accordingly, these time-based measurements included in the research process explain approximately 22% of the total variance in the total test scores. Based on the standardized regression coefficients (β), the relative importance order of the duration measurements on the total test scores is as follows: times to

first fixation and durations of first fixation. When the t-test results regarding the significance of the regression coefficients are examined, it is found that only average durations of fixation and times to first fixation are significant predictors of the total test scores. Durations of first fixation and average durations of visit do not have a significant effect on the total test scores.

In this part of the study, the research question *"What is the rate at which the duration-focused visual measurements obtained from the areas of interest specifically defined for each item in the Eighth Grade Achievement Test explain the subtest scores obtained from the same test?"* was sought. For this purpose, multiple linear regression analysis was performed in order to reveal the connection between the measurements obtained and the verbal and numerical subtest scores. After conducting analyses regarding the basic assumptions of the multiple linear regression analysis, it was observed that the verbal subtest scores and the duration measurements did not meet the multivariate normality assumption together, and the analyses proceeded with the numerical subtest scores, which met the multivariate normality assumption. Table 11 shows the rate at which the duration measurements obtained within the scope of the study explain the numerical subtest scores.

Table 11. *The Explanation Rate of Numerical Subtest Scores of Duration Measures to Items in the Eighth Grade Achievement Test*

Variables	B	B_{SH}	β	t	p
Constant	12.301	1.450		8.483	.000***
Times to first fixation	-.027	.017	-.257	-1.633	.108
Durations of first fixation	18.991	10.193	.405	1.863	.068
Average durations of fixation	-24.712	9.196	-.594	-2.687	.010*
Average durations of visit	.011	.138	.014	0.082	.935
R=.406	R^2=.165				
$F(4,55)$ =2.718	p=.039*				

n=60, *p<.05, **p<.01, ***p<.001

According to Table 11, a moderate and significant relationship is observed between times to first fixation, durations of first fixation, average durations of fixation, and average durations of visit with the numerical subtest scores, R=.406, R^2=.165, p<.05. Accordingly, these time-based measurements included in the research process explain approximately 17% of the total variance in the numerical subtest scores. When examining the t-test results regarding the significance of the regression coefficients, it is found that only average durations of fixation are significant predictors of numerical subtest scores. Times to first fixation, durations of first fixation and average durations of visit do not have a significant effect on numerical subtest scores.

In this part of the study, the research question *"What is the rate at which the count-focused visual measurements obtained from the areas of interest specifically defined for each item in the Eighth Grade Achievement Test explain the subtest scores obtained from the same test?"* was sought. For this purpose, multiple linear regression analysis was performed in order to reveal the connection between the obtained measurements and the verbal and numerical subtest scores. After conducting analyses regarding the basic assumptions of multiple linear regression, it was observed that the numerical subtest scores and the count measurements did not meet the multivariate normality assumption together, and the analyses proceeded with the verbal subtest scores, which met the multivariate normality assumption. Table 12 shows the rate at which the count measurements obtained within the scope of the study explain the verbal subtest scores.

Table 12. *The Explanation Rate of Verbal Subtest Scores of Count Measures to Items in the Eighth Grade Achievement Test*

Variables	B	B_{SH}	β	t	p
Constant	15.151	1.086		13.946	.000***
Number of fixations	-.021	.008	-.408	-2.740	.008**
Number of visits	.092	.056	.243	1.631	.108
$R=.342$	$R^2=.117$				
$F (2,57) =3.766$	$p=.029*$				

$n=60, *p<.05, **p<.01, ***p<.001$

According to Table 12, a moderate and significant relationship is observed between the number of fixations and the number of visits with the verbal subtest scores, $R=.342$, $R^2=.117$, $p<.05$. Accordingly, these count-oriented measurements included in the research process explain approximately 12% of the total variance in the verbal subtest scores. When examining the t-test results regarding the significance of the regression coefficients, it is found that only the number of fixations are a significant predictor of the verbal subtest scores. The number of visits does not have a significant effect on the verbal subtest scores.

4. Discussion, Conclusion, and Suggestions

In this study, which aims to examine the relationships between time and count-oriented visual measurements obtained from multiple-choice tests and the test and item statistics of these tests, correlational analyses were initially performed. After conducting these analyses between the visual measurements collected from the areas of interest specifically defined for each item and the total test scores, it was found that almost all of the relationships detected were negative. In this case, it can be predicted that when a certain increase is observed in the focus and visit measurements of the items with which the relationship is detected, a certain decrease in the total test scores may occur (Akçay & Altun, 2019; Bayazıt, 2013; Jacob & Karn, 2003; Just & Carpenter, 1976a, 1976b, 1980; Klein *et al.*, 2020; Majooni *et al.*, 2016; Schwonke *et al.*, 2009). However, a positive relationship, albeit weak, was found between total durations of visit and total test scores obtained specifically for the verbal item 16. Although this result is similar to the study of Glaholt *et al.* (2009), this relationship differs from all the negative relationships identified within the scope of the research. The fact that students revisit and review this item can be seen as a factor that increases success. However, this result may indicate that this item in particular is not sufficiently understood by the students, and it may also indicate that this item is an item aimed at measuring meta-cognitive skills. On the other hand, the fact that this item is a relatively easy ($p_j=.65$) and distinctive ($r_{jx}=.37$) item may imply that students understand this item more easily when they go back and review it.

When the rate at which time-focused visual measurements explain total test scores is examined, it is seen that these measurements together give a moderate and significant relationship with total test scores. At this stage, t-test results regarding the significance of regression coefficients were examined in particular, and it was observed that average durations of fixation and times to first fixation were significant predictors of total test scores. While average durations of fixation are associated with the difficulty level of the text of interest, times to first fixation are associated with the time required to select, organize, and/or access information from areas of interest. It is anticipated that these measurements, reflecting situational reactions, may offer important contextual information regarding students' test performances. Therefore, the rate at which these time-focused measurements explain verbal and numerical subtest scores was also examined, and it was seen that these measurements together gave a moderate and significant relationship with numerical subtest scores. When the t-test results regarding the significance of regression coefficients were examined, it was observed that only average durations of fixation

were a significant predictor of numerical subtest scores. At this stage, the rates of visual measurements obtained with a focus on counting, explaining the subtest scores, were also examined and it was found that these measurements together gave a moderate and significant relationship with the verbal subtest scores. However, it was observed that the rate of these counting-focused measurements explaining the verbal subtest scores remained relatively low, and when the t-test results regarding the significance of the regression coefficients were examined, it was seen that only the number of fixations was a significant predictor of the verbal subtest scores. All of these results obtained are similar to many studies (Akçay & Altun, 2019; Bayazıt, 2013; Goldberg & Kotval, 1999; Jarodzka *et al*., 2015; Just & Carpenter, 1976a, 1976b, 1980; Lai *et al*., 2013; Liversedge *et al*., 1998; Majooni *et al*., 2016; Poole, Ball & Phillips, 2005; Rayner, 1998, 2009; Rayner *et al*., 2006; Saleem *et al*., 2021; Scheiter & Eitel, 2015; Scheiter & Eitel, 2017; Schwonke *et al*., 2009; Underwood *et al*., 1990; van Meeuwen *et al*., 2014).

When we look at the mentioned relational results from a broad perspective, it is seen that the relationship between average durations of fixation and success becomes prominent. However, Negi and Mitra (2020) particularly challenge this observation. According to the authors, average durations of fixation do not reveal a clear pattern for complex tasks or visualization features. Upon examining the aggregate of measurements obtained in this regard, it is shown that although successful students have shorter fixation durations than unsuccessful students, no significant difference is found when compared to students with average success. More importantly, despite such intuitive results, students with average success have longer fixation durations than unsuccessful students (Negi & Mitra, 2020). Negi and Mitra (2020) state that the notion of average durations of fixation being related to various operations required by the tasks assigned to students is almost intuitive. According to them, operations belonging to more complex tasks require longer measurements. Therefore, they argue that students need to focus for longer periods in order to process all the information in the tasks given to them. However, as success increases, the demand for such operations diminishes; because successful students retrieve task-related schemas from long-term memory, thereby reducing the processing demands associated with the tasks assigned to them (Mitra *et al*., 2017).

When comparing the obtained visual measurements according to gender, it was observed that times to first fixation, average durations of visit and total durations of visit showed significant differences according to gender. Accordingly, significant differences were found in favor of girls in times to first fixation and total durations of visit, and in favor of boys in average durations of visit. It can be inferred that girls had more difficulty verbally focusing on item 17 and numerically focusing on item 9 compared to boys, and it can be predicted that girls had more difficulty selecting, organizing or accessing information related to these items (Akçay & Altun, 2019; Liversedge *et al*., 1998; Scheiter & Eitel, 2015; Scheiter & Eitel, 2017; van Meeuwen *et al*., 2014). However, it has been observed that female students' efforts to understand numerical item 12 or their time to go back and re-read and/or their decision-making processes regarding this item take longer than male students, suggesting that that female students experience a lack of understanding and/or comprehension in relation to this item (Klein *et al.,* 2020; Lindner *et al*., 2014; Liversedge *et al*., 1998; Saleem *et al*., 2021). On the other hand, it can be said that the efforts of male students to understand the verbal item 8 or the time they spent going back and rereading and/or making decisions about this item were longer on average than female students, and it can be predicted that male students experienced a lack of understanding and/or comprehension in connection with this item. (Klein *et al*., 2020; Lindner *et al*., 2014; Liversedge *et al*., 1998; Saleem *et al*., 2021). These findings are significant from a psychometric perspective. Based on these results, it can be said that the probability of answering these items correctly varies according to the gender of the students taking the test, and the fact

that the calculated effect sizes are medium and large adds additional importance to these results (Ackerman, 1992; Bolt, 2002; Büyüköztürk, 2020; Embretson & Reise, 2000; Klein *et al*., 2020). These results raise concerns about the validity regarding the relevant items (Demirus & Gelbal, 2016). Therefore, based on the significant differences determined among the visual measurements obtained, differential item functioning (DIF) studies can be conducted on these items. The results of such studies could provide objective evidence for DIF.

As a result, the relationships identified between the visual measurements obtained from multiple-choice tests and the test and item statistics of these tests have significant potential to provide valuable structural insights about multiple-choice tests; however, this potential has not yet been fully explored. This potential can be further investigated by conducting research with different study groups or through the application of various psychometric analyses. The psychometric insights gained from such research can provide a different perspective for future research.

Orcid

Tolga Coşguner https://orcid.org/0000-0002-8589-7472

Burcu Atar https://orcid.org/0000-0003-3527-686X

REFERENCES

Ackerman, T. A. (1992). Assessing construct validity using multidimensional item response theory. Paper Presented at the Annual Meeting of American Educational Research Association. San Fransisco, CA, USA.

Akçay, A., & Altun, A. (2019). Farklı kısa süreli bellek uzamlarına sahip öğrencilerin farklı dikkat tasarımına sahip öğrenme ortamlarındaki göz hareketlerinin incelenmesi. *Eğitim Teknolojisi Kuram ve Uygulama, 9*(2), 588-614.

Antle, A. N. (2013). Research opportunities: Embodied child-computer interaction. *International Journal of Child-Computer Interaction, 1*(1), 30-36.

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science, 39*(5), 581-601.

Bayazıt, A. (2013). *Farklı soru biçimlerinin göz hareketleri, başarım ve cevaplama süresine olan etkilerinin incelenmesi* [Doktora tezi, Hacettepe Üniversitesi]. Ulusal Tez Merkezi.

Becker, W. J., & Menges, J. I. (2013). Biological implicit measures in HRM and OB: A question of how not if. *Human Resource Management Review, 23,* 219-228.

Berzak, Y., Katz, B., & Levy, R. (2018). Assessing language proficiency from eye movements in reading. In M. A. Walker, H. Ji, & A. Stent (Ed.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1986-1996). Association for Computational Linguistics.

Bolt, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education, 15*(2), 113-141.

Büyüköztürk, Ş. (2020). *Sosyal bilimler için veri analizi el kitabı.* Pegem Akademi.

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. In C. Was, F. Sansosti, & B. Morris (Ed.), *Eye-tracking technology applications in educational research* (pp. 195-218). IGI Global.

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems, 20*(6), 557-574.

Demirus, K. B., & Gelbal, S. (2016). Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 7*(1),182-201.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. Behavior Research Methods, *Instruments, & Computers, 34*(4), 455-470.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice.* Springer-Verlag.

Durna, Y., & Arı, F. (2016b). Polinom fonksiyonları ile göz B-bakış yeri tespiti geliştirilmesi ve uygulaması. *Savunma Bilimleri Dergisi, 15* (2), 24-45.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Lawrence Erlbaum Associates, Publishers.

Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler*. Pegem Akademi.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education.* McGraw-Hill.

Galván, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping, 31*, 879-890.

Glaholt, M. G., Wu, M. C., & Reingold, E. M. (2010). Evidence for top-down control of eye movements during visual decision making. *Journal of Vision, 10*(5), 1-10.

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Journal of Information Visualization, 10*(3), 182-195.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631-645.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Radach, R., Hyona, J., & Deube, H. (Eds.), *The mind's eye* (pp.573-605). Elsevier.

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkens, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative? *British Journal of Educational Technology, 46*(4), 803-817.

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

Just, M. A., Carpenter, P. A. (1976b). The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation, 8*(2), 139-143.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Khamis, H., & Kepler, M. (2010). Sample size in multiple regression: 20 + 5K. Journal of Applied Statistical Science, 17(4), 505-517. Retrieved from https://www.researchgate.net/publication/285744052_Sample_size_in_multiple_regression_20_5k

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction, 20*(1), 1-30.

Klein, P., Lichtenberger, A., Küchemann, S., Becker, S., Kekule, M., Viiri, J., Baadte, C., Vaterlaus, A., & Kuhn, J. (2020). Visual attention while solving the test of understanding graphs in kinematics: An eyetracking analysis. *European Journal of Physics, 41*, 1-16.

Korkmaz, L. (2017). Tutumlarımızın ne kadar farkındayız? Örtük tutumlar ve örtük ölçüm yöntemleri. *Türk Psikoloji Yazıları, 20*(40), 109-127.

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 learning from 2000 to 2012. *Educational Research Review*, *10*, 90-115.

Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, *47*, 91-102.

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

Liu, H. C., Lai, M. L., & Chuang, H. H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior, 27*, 2410-2417.

Liu, P. L. (2014). Using eye tracking to understand learners' reading process through the concept mapping learning strategy. *Computers & Education, 78*, 237-249.

Liu, W., Yu, M., Fan, Z., Xu, J., & Tian, Y. (2017, Oct.). Visual attention-based evaluation for multiple-choice tests in e-learning applications. 2017 IEEE Frontiers in Education Conference (FIE). Indianapolis, IN, USA.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In Underwood, G. (Eds.), *Eye guidance in reading and scene perception* (pp. 55-75). Elsevier.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6-14.

Majooni, A., Masood, M., & Akhavan, A. (2016). An eye-tracking experiment on strategies to minimize the redundancy and split attention effects in scientific graphs and diagrams. In Di Bucchianico, G., & Kercher, P. (Eds.), *Advances in Design for Inclusion: Proceedings of the AHFE 2016* (pp. 529-540). Springer International Publishing.

Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive Processing, 13*(1), 261-265.

Mitra, R., McNeal, K. S., & Bondell, H. D. (2017). Pupillary response to complex interdependent tasks: A cognitive-load theory perspective. *Behavior Research Methods, 49*, 1905-1919.

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.

Nugrahaningsih, N., Porta, M., & Ricotti, S. (2013, Oct.). Gaze behavior analysis in multiple-answer tests: An eye tracking investigation. 12th International Conference on Information Technology Based Higher Education and Training (ITHET). Antalya, Turkey.

Ömur, S., & Görgülü Aydoğdu, A. (2017). Göz izleme araştırmaları ve iletişim alanında yeni yönelimler. *International Journal of Social Sciences and Education Research, 3*(4), 1296-1307.

Peterson, S. J., Reina, C. S., Waldman, D. A., & Becker, W. J. (2015). Using physiological methods to study emotions in organizations. In Härtel, C. E. J., Zerbe, W. J., & Ashkanasy, N. M. (Eds.), *New Ways of Studying Emotions in Organizations* (pp. 1-27). Emerald Group Publishing Limited.

Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. In Fincher, S., Markopoulos, P., Moore, D., & Ruddle, R. (Eds.), *People and Computers XVIII—Design for Life: Proceedings of HCI 2004* (pp. 363-378). Springer-Verlag.

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

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

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(3), 241-255.

Rodrigues, R., & Rosa, P. (2017). Eye-tracking as a research methodology in educational context. In Was, C., Sansosti, F., & Morris, B. (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 1-26). IGI Global.

Rosa, P. J. (2015). What do your eyes really say? Bridging eye movements to consumer behavior. *International Journal of Psychological Research, 8*(2), 91-104.

Saleem, M. R., Straus, A., & Napolitano, R. (2021). Interpretation of historic structure for non-invasive assessment using eye tracking. International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences-ISPRS Archives, 46*(1), 653-660.

Saß, S., Schütte, K., & Lindner, M. A. (2017). Test-takers' eye movements: Effects of integration aids and types of graphical representations. *Computers & Education*, *109*, 85-97.

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction, 36*, 11-26.

Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In Was, C., Sansosti, F., & Morris, B. (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 143-164). IGI Global.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*, 1227-1243.

Shayan, S., Abrahamson, D., Bakker, A., Duijzer, A. C. G., & van der Schaaf, M. F. (2017). Eye-tracking the emergence of attentional anchors in a mathematics learning tablet activity. In Was, C., Sansosti, F., & Morris, B. (Eds.), *Eye-tracking technology applications in educational research* (pp. 166-194). IGI Global.

Solheim, O., & Uppstad, P. (2011). Eye-tracking as a tool in process-oriented reading test validation. *International Electronic Journal of Elementary Education*, *4*(1), 153-168.

Tabachnick, B. G., & Fidell, L. (2015). *Çok değişkenli istatistiklerin kullanımı.* Baloğlu, M. (Çev. ed.). Nobel Akademik Yayıncılık.

Tabbers, H. K., Paas, F., Lankford, C., Martens, R. L., & van Merriënboer, J. J. G. (2008). Studying eye movements in multimedia learning. In Rouet, J. F., Lowe, R., & Schnotz, W. (Eds.), *Understanding Multimedia Documents* (pp. 169-184). Springer.

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

Tobii Pro. (2019, March). Learn & support. https://www.tobiipro.com/learn-and-support/

Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2011). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*, 375-385.

Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationships between reading skill, reading speed, and visual inspection. *Language and Speech, 33*(1), 69-81.

van Meeuwen, L. W., van Merriënboer, J. J. G., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., & Bock, J. J. P. R. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction, 32*, 10-21.

Zentall, S. R., & Junglen, A. G. (2017). Investigating mindsets and motivation through eye tracking and other physiological measures. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-tracking technology applications in educational research* (pp. 48-64). IGI Global.
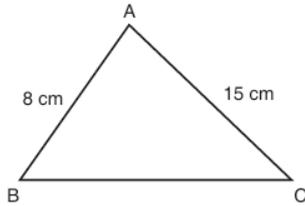
Appendix

A

17. Sakın bitti sanma her şeyi,
    Sevdiğin kadar sevileceksin.

Aşağıdaki cümlelerin hangisindeki virgül, bu dizeler-deki işleviyle kullanılmıştır?

A) Ak akçe kara gün içindir, demişler.

B) Evet, bunlar eski olanlar…

C) Al dersini, otur şimdi!

D) Sevimli, bembeyaz bir kediydi.

B

9.

Yukarıda verilen ABC üçgeninin kenar uzunlukla-rı santimetre türünden tam sayı değerleri almak-tadır.
|ABI| = 8 cm ve |ACI| = 15 cm **olduğuna göre, BC kenarının uzunluğunun çift bir tam sayı olma olasılığı kaçtır?**

A) $\frac{3}{5}$ B) $\frac{8}{15}$ C) $\frac{7}{17}$ D) $\frac{1}{5}$

C

8. Bu metinde tırnak işareti (" ") içerisinde verilen deyimle anlatılmak istenen aşağıdakilerden hangisidir?

A) Kendi gidiş ve davranışını başkasınınkine benzetmek

B) Alışılan bir yere gitmekten, alışkanlıktan kendini alamamak

C) Birini aldatmak, kandırmak için dalavere çevirmek

D) Her türlü fedakârlığı göze alıp ödün vermek

D

12.

	X maddesi	Y maddesi	Z maddesi
Tepkime öncesi kütle (g)	▲	■	0
Tepkime sonrası kütle (g)	12	0	★

X ve Y maddelerinin kapalı bir kapta gerçekleşen kimyasal tepkimesi sonucunda Z maddesi oluşmaktadır.

Buna göre ▲, ■ ve ★ sembolleri ile gösterilen kütle değerleri aşağıdakilerden hangisi gibi olabilir?

	▲	■	★
A)	60	12	60
B)	100	30	70
C)	48	27	48
D)	60	48	60