academicJournals

Vol. 12(17), pp. 839-854, 10 September, 2017
DOI: 10.5897/ERR2017.3304
Article Number: D05592865818
ISSN 1990-3839

http://www.academicjournals.org/ERR

Educational Research and Reviews

Full Length Research Paper

Examination of test and item statistics from visual and verbal mathematics questions

Cagla Alpayar* and H. Deniz Gulleroglu

Educational Measurement and Evaluation, Mathematics Education, Educational Theory, Faculty of Educational Sciences, Ankara University, Ankara, Turkey.

Received 19 June, 2017; Accepted 11 August, 2017

The aim of this research is to determine whether students' test performance and approaches to test questions change based on the type of mathematics questions (visual or verbal) administered to them. This research is based on a mixed-design model. The quantitative data are gathered from 297 seventh grade students, attending seven different middle schools in Cankaya, Ankara. Of all the students who participated in the research, qualitative data were gathered from 10 of them. In this research, seventh grade mathematics achievement test was developed by the researchers in visual and verbal forms. 10 of the students were selected and interviewed by utilizing semi-structured interview form. From the findings of the study, there was significant difference between the test scores and response time of the two forms in favor of visual form. The difference between item statistics is changeable in terms of the function of the visual. The results of the interviews showed that students have positive views toward visual questions. The students' perceptions of the use of visual in mathematics questions are examined in three main categories: preferability, comprehensiveness and responsibility of the questions.

Key words: Visual, item difficulty, item discrimination, students' perceptions.

INTRODUCTION

Today, information is transferred through different forms and visual stimuli. Visual is commonly utilized in different areas including marketing and education, and is defined differently based on field of utilization. Generally visual can be defined as the stimulus related to sight (Turkish Language Institutions (2016).

In the literature, visuals have been defined as vehicles of concretizing the mental representation of any concepts (Sharma, 1985; Beb-Chaim et al., 1989).

More specifically, some researchers define visuals as a form of geometry used for stating mathematical concepts (Habre, 2001; Zaraycki, 2004). In education, visual aids are utilized in many print education materials. In educational fields, visuals are defined as means (like symbols, graphics or photograph) of transforming information in other ways other than verbal forms (Lanzig and Stanchev, 1994).

Visuals are used in written instructional documents for many purposes. The functions of visuals in materials are classified in different ways. One of the most extensive classifications of the function of visuals in instructional documents was developed by Clark and Lyons (2004).

*Corresponding author. E-mail: c.alpayar@gmail.com.

Visuals are placed in written instructional documents mainly for two purposes:

1. Communicational and
2. Emotional.

Communicative functions generally focus on how information is transferred. With this approach, visuals are used in instructional documents for decorative purposes. They therefore add a sense of aesthetic and humor to the documents, making them more attractive.

Visuals used in measurement instruments, especially in reading comprehension tests, are similar to the ones used in other instructional documents in making them more attractive although they are not directly related to the content of the test item (Levin, 1981).

This usage is sometimes recommended for instructional activities but not for measurement instruments. Visuals are used to make specific place, object or person described in the test item more concrete and real for readers (Levin, 1981).

For transformative purposes, visuals are utilized to describe the process. To highlight changes with respect to time and place, the steps of the process are reflected by visuals. Due to this usage, congruent events could be ordered with fewer words (Yuill and Oakhill, 1997).

Both usages, related to interpretational purposes, help the reader to understand the text. Therefore, it is necessary to state some situations in a more comprehensible way and briefly (Peeck, 1993). Similarly, the basic purpose of integrating visuals in the tests items is to state complex stations verbally with less words or even without using words (Furst, 1958).

Visuals thus make complex test statements more simple and comprehensive (Peeck, 1993). In test items with scientific content, visuals may be more informative than words (Stewart et al., 1979; Crisp and Sweiry, 2003). By representing the same content with appropriate visuals, fewer words may also prevent wrong answers because visuals may reduce the difficulty in reading comprehension and contribute to item validity (Shorrocks-Taylor and Hangravesgen, 1999). Especially for students with low reading comprehension ability, using visual in test items may help to compensate for this problem (Kopriva, 2008).

Another form of visuals used for communicative purposes is mnemonic. Students can easily recall information gained in the text with visuals (Nickerson, 1965; Standing, 1973; Diamond, 2008). Similarly, visuals in the test items are effective in recalling information to answer the test item (Peeck, 1974). For the communicative purposes, visuals serve as organizers. This function provides an arrangement of the information given in the test item within the appropriate structure (Levin, 1981). The qualitative relationships between contents can be projected through visuals such as trees, charts and concept maps (Clark and Lyons, 2004). These kinds of usage are frequently encountered in science questions.

Visuals in educational materials have psychological as well as communicative functions. In line with this, visual factors primarily help individuals to focus on educational materials, and also motivate them (Sweiry et al., 2002; Clark and Lyon, 2004).

Handono (1996) argues that visual materials with texts are more effective in making abstract thoughts more concrete. Specifically in math questions, visuals motivate individuals and help them to concentrate (Murphy, 2009a). Hence, tests are recommended to include symbols that motivate students (Salend, 2009). However, there are some other researches with reverse findings. 15% of the studies focused on this issue conclude that visual materials do not have a positive influence on students' motivation (Levie and Lentz, 1982).

Visuals are utilized to make the situations in test items as close as possible to the daily life (Berberoğlu, 2012; Saß et al., 2012). This usage is one of the basic purposes of visuals in the test items (Murphy, 2009a). Visuals not only make the information in the test items more meaningful for the students but also provide an opportunity to evaluate the utility of knowledge in daily life. Based on this function of visuals, the aim is to increase its availability in the math subtests of the Programme for International Student Assessment (PISA) (Tout and Spithill, 2015).

Therefore, visuals can be incentive for students with low achievement level by making the information more meaningful (Kopriva, 2008). Therefore, they also respond to psychological functions indirectly. However, designing test items to reflect everyday life may lead to misunderstanding of the question (Ahmed and Pollitt, 2007), and it may have a negative effect on performance. In order to prevent this kind of performance decrease, the focus of mathematics questions should be mathematical content rather than everyday life relation (Brown, 1999).

Beyond psychological and communicational functions, visuals may be effective for the students' cognitive processes. Mathematicians and scientists have a consensus on the idea that visuals play a central role in cognitive processes (Phillips et al., 2010). Although the models examining the effects of visuals predominantly try to explain learning situations, theoretical assumptions make these models relevant for test situations as well (Sternberg, 1999c).

Every assessment is based on a theory or an understanding regarding how people learn, what people know and how knowledge develops over time (National Research Council, 2001). Thus, cognitive learning models are important for designing items in developing tests and important pieces of the framework used to reach conclusions about answerers depending on their test item performance (National Research Council, 2001).

In other words, cognitive theories can be used to determine the characteristics of items in the answering

process (Whitely, 1983; Haladyna et al., 2002). Based on cognitive learning theories, it can be said that mental operations during the learning and assessment have some similarities and differences (Saß et al., 2012). Therefore, one of the theories used to explain how visuals influence answerers' performance is "dual coding theory". The theory developed by Paivio (1990) explains how to learn the information coming from two different sources.

According to the theory, information can be confronted in two ways, visual and verbal. Verbal information is expressed by words whereas visual information is expressed by non-verbal forms such as picture and voice. Moreover, two different types of information are recorded as two different codes in the cognitive system through the affective memory (Pavio, 1990).

These two systems are related in terms of being convertible to each other, although they are completely independent in the area of functions and structures of verbal and visual operation process (Vekiri, 2002). In other words, visual codes have a verbal response inside the brain; meanwhile verbal codes have a visual response.

Consequently, Pavio (2013) emphasizes that visuals are used to organize information. Additionally, only verbal stimuli have less influence on activation of non-verbal memory therefore they are less remembered (Lohr, 2003). In contrast, giving visual and verbal information together provides connection while coding.

Eventually, learning is more permanent when both systems are included (Paivio, 2013; Shepard, 1967; Sweller, 2010). The mental process of visual elements in a learning procedure is similar to the process of visual elements in an assessment procedure; therefore, based on this similarity, Ahmed and Pollit (1999) developed a response model.

According to this model, individuals interpret the elements of an item over two different systems when they confront a test item with visual elements. First, individuals make mental representation, and then examine this representation whether it matches with existing information and finally makes a comparison between these two situations (Pollitt and Ahmed, 1999). These phases are named; reading, examining and matching respectively. Answerers have no control over this process (Ahmed and Pollitt, 2007); as a result, the elements of an item can differentiate the mental representation of an individual from his/her test performance.

Physical characteristics of a test can be determinant of increasing true answer possibility within the test items including both verbal and visual information. In addition, Carpenter et al. (1990) argue that one of the three factors which are effective in predicting item difficulty is to make abstract connections between variables of the item. In order to relate these variables, the information presented from different sources such as verbal and visual should be ascribed a meaning and interpreted by supporting each other. Hence, the concern is the mental process of responding while developing test items, and so potential statistical reflection of this design would be beneficial. There are some points to be concerned about due to positive reflections.

Due to the positive effect of answering performance, visual components should be salient for students and need less mental operation to ascribe meaning; that is to say answerers can executive operations automatically (Vekiri, 2002). For this, visual components should be as clear as possible. To ascribe meaning to visual information including irrelevant information to the stem of an item or repeating the information existing in the stem, needs more mental steps, therefore it may obstruct the real performance of answerers. Additionally, irrelevant interactions emerge from divided attention of answerers into various information sources on account of non-central information components (Berends and Lieshout, 2009).

Individuals are obligated to interpret more than one source in order to ascribe information, and consequently their attention is splitted. For instance, an image with its description in the stem of an item leads to the same effect, and this situation requires more mental operations. According to Ahmed and Pollit (2000) model, irrelevant details and contents cause activation of wrong concepts, and they can orient to wrong answers specifically in the stressful atmosphere of the exam.

Moreover, simplicity is a desired feature of a test in respect to the multi-choice test item writing guide identified by Haladyna (1989). Clear questions prevent distraction. The basic approach in simplicity is to allow visual components only when they support understanding of the question and help in the answers (Filippatou and Pumfrey, 1996; Crisp and Sweiry, 2003; Kopriva, 2008; Haladyna and Rodriguez, 2013). Thus, each visual component of an item must have a function for the question (Crisp and Sweiry, 2003). Kopriva (2008) states that unnecessary or non-supportive information should not be given in the visual component of an item.

A research by Rasmussen and Bisanz (2005), conducted with elementary and kindergarten students, has revealed that answerers have difficulty in ignoring irrelevant information and this leads to decrease in their performance especially mathematics performance.

Similarly, Berends and Van Lieshout (2009) argue that giving room to repeated and irrelevant information in the item content is not appropriate for exams which examine arithmetic skills. Accordingly, students who do not become automatic in operational skills should struggle more to calculate to get answer. Another way to prevent divided attention is to place related information sources as close as possible to each other (Sweller, 1994). Placing visual and verbal information sources away from each other may have negative effect on the respondents' performance.

Another point to be noticed is that test items should be

written by considering respondents' close and distant environment, and situations they are familiar with or probably they will be familiar with (Demirtaşlı, 2010). A question not understood by individuals may lead to various biases (Schiffman, 1995; Shriver, 1997; Anagnostopoulou et al., 2015).

Similarly, in the framework they developed for the right usage of visual element in a test item, (Salona-Flores and Wang (2011) emphasize that the visual component of the test item should belong to the respondents' culture. They forecast that if respondents have no idea about the topic their motivation decreases. This situation can be applied to the test item which has a visual component. In a study conducted by Ahmed and Pollitt (2000), answerers stated that they left the question blank because they had not seen the bridge in the visual before, so thought they could not have replied.

To have consistency of all the visual components within the scope of the test is another point to be considered (Haladyna et al, 2013). Accordingly, all visuals given in a test should have the same size and format. Afore-mentioned, consistency enables the test to be simplified (Osterlind, 1989). There are various suggestions in the literature for the type of visual components which will be placed in the test scope. Salona-Flores and Wang (2011) emphasize that visual should be a drawing rather than a picture or caricature. It should be a simplified representation of a real ingredient. Real pictures may take respondents' attention away from the important information of the text; therefore, to ascribe meaning and study on diagrams may take long time. As for caricatures, some findings reveal that caricatures take respondents away from the direction of scientific thinking (Mevarech and Stern, 1997). Hence, respondents try to solve the question- specifically designed with everyday life content- by using everyday life information with less abstract thinking. In line with this, a study by Ahmed and Pollitt (2000) indicates that respondents reported that the fish caricature within a biology question took them away from reality.

While writing test items in a visual-including way, to consider all these points argued under the title of formal features of a measuring instrument does not ensure item function to be executed. Accuracy of measurements can be possible when the instrument performs its function properly. This is crucial particularly for appropriate decisions that are made on the basis of this information. Thus, the features of test items should be considered carefully to obtain practical, meaningful and applicable information from measurements. The features of a test item are handled with various aspects such as relatedness, balance, competence, objectivity, specificity, difficulty, discrimination, reliability and response speed (Ebel, 1965). Two basic approaches can be assessed empirically and judicially (Wiggins, 1998).

One of the methods used to identify the qualification of test items is empiric approach. Empirical evidence is collected by test and item statistics. Anastasi (1982) states that there are features of a test item on the basis of these collected qualitative and quantitative values. Hence, formal features of an item including visuals can influence respondents' performance, and so test and item statistics can be influenced. The basic purpose of item analyses is to ensure that developed instrument consists of items which have desired features and to be informed about respondent groups on the item level (Erkuş 2003). Item statistics are not only used in developing studies but also in determining item biases (Nitko, 2004).

Item discrimination is one of the statistics that is calculated to investigate the qualification of a measuring instrument. Item discrimination is the power used to distinguish individuals who have high and low performance in the feature measured (Crocker and Algina, 1986).

In other words, item discrimination is also referred to as item validity in terms of demonstrating the degree of expediency of the item. According to Hillocks (2002), the basic problem with tests is whether it fulfills the function aimed by the test developer or not. In this sense, only a well-qualified multi-choice question is answered correctly by respondents who have high scores in total; is answered wrong by answerers who have low scores in total. Another feature providing information about item qualification is item difficulty. In the most practical sense, item difficulty is the correct response rate in the group it is applied (Crocker and Algina, 1986). This value gives information about difficulty of a question therefore it is described as easy when majority of the group answer correctly.

Various studies are conducted in order to identify the effects of visual ingredients on test and item statistics. Thinsley and Davis (1974) concluded that to measure the same achievements, two different test forms can be developed, one of which is completely configured by verbal items whereas the other one is configured by using visuals. On the basis of this conclusion, two different forms of the same item (one of them is to include visuals and the other is only verbal) are prepared and differentiation of item statistics is examined in both forms. In this frame, Washington and Godfrey (1974) and De Melo (1980) applied air force special ability test and biology test respectively; as a result they found that visual questions are more advantageous than non-visuals. Similar studies conducted in Turkey through geometry, physics and science questions revealed that test statistics do not differ by the addition of visual elements (Kaptan, 1985; Bağcı, 1998; Civelek, 1998; Duran and Balta, 2014).

In addition, judicial approaches are beneficial to determine the qualification of the test developed. For this purpose, not only the experts on the issue but also respondents are referred to when evaluating the test item. The test item can be asserted that it fulfills the aim properly identified by developers when their expectations

and answerer's responses are overlapping. In other words, if a student achieves expected mental processes while answering the question, relevant measurement can be made (Ahmed and Pollitt, 2000).

To identify this, students' views are taken. It is aimed to determine what they think and how they approach the item when they encounter a question. From these expressions, it is tried to detect if developers' expectations and answerers' mental processes are overlapping. Respondents primarily expect a test item not to be confusing (Popham, 2000).

It is emphasized in the literature that test developers attach importance to test items to be pure in order for respondents to deliver real performance. Therefore, developers and respondents' stylistic expectations from a test item are overlapping at the point that a test item needs to be pure as possible. Additionally, visual ingredients can be described as one of the stylistic features, therefore, predicted to be determinant of item functionality.

In the studies that aim to determine the impact of visual elements on item statistics, various tests developed for achievements of different courses are utilized. Those studies conclude that a general framework contribute to the visual ingredients to achieve tests of different courses. According to those findings, there are some domain specific situations in using visual ingredients. It is valid for specifically mathematics test items as well. Students frequently perceive mathematics problems as visual sets and use mathematics models to solve them (Murphy, 2009a).

Understanding abstract mathematics concepts is related to the ability of seeing how these concepts function. Hence, students naturally utilize visual models while solving a mathematics problem (Murphy, 2009b). İşler (2003) states that visuals in education materials can increase the level of understanding of verbal idea, and can take part in the focus of discussion in problem solving process. Cognitive psychological theories also support the impact of visuals on answering questions. Although some studies focus on the relationship between visualization and problem solving skills, there are few researches examining the influence of visuals on the psychometric value of tests.

Respondents tend to solve mathematics problems by using the way they are familiar with (Luchins, 1942). This situation can prevent individuals to head for different solutions (Antoinetti, 1991) whereas visual ingredients can help respondents to use diverse solutions. Moreover, according to Gestalt approach, respondents rearrange problems progressively in their mind (Wertheimer, 1960). This process keeps respondents away from thinking about the item factors separately. Presenting verbal information as a serial can prevent respondents from using holistic approach while visuals can submit a holistic framework in line with information process (Kaufmann, 1985).

The necessity of using visuals in math questions can be based on various grounds in both national and international literature. Whether miscellaneous target behaviors described in National Council of Teachers of Mathematics (NCTM) standards are achieved can be examined more effectively through questions with visual components. Those target behaviors include daily life reasoning skills and representation of mathematical ideas in diverse ways such as tables, pictures and graphics (NCTM, 1999).

Similarly, visual elements are often utilized in the pen-paper exams made for the assessment of the target of cultivating individuals who can transfer the math to daily life and who can share his mathematical ideas, emphasized in the vision of The Ministry of National Education middle school teaching program. The relevant part of the vision takes place in the program with the expression "to cultivate individuals who can use mathematics in daily life, solve problems, and share his ideas and solutions".

Moreover, it can be asserted that mathematics questions containing visual ingredients are specifically functional in order to examine the target behaviors described as "to be able to use mathematical model, to match models with verbal and mathematical expressions" in general aims of The Ministry of Education program (2009). This type of usage enables mathematics, which is accepted as having an independent language within the program, to be enriched with visual ingredients.

In line with the functional purposes mentioned earlier, visual components take place in both large scope tests and teacher made tests. Visual elements are frequently utilized in the test items developed for particularly examining the ability of transferring mathematics information given at school to daily life. In this sense, one of the leading international exams Program for International Student Assessment (PISA) often uses visual elements in mathematics questions. Visuals within the scope of PISA are utilized in order to make the question more remarkable.

Visual elements are integrated into the national exam whose scores are utilised for the decision of transition from middle to high schools in Turkey. Although this exam is named differently from time to time, it is applied every year regularly. For example, OKS (middle education institutions exam) was implemented on only eighth grade students until 2008 where it was left to SBS (placement test) which was applied to sixth, seventh and eighth grade students.

Changes in that exam are not only limited by changing its name but also including some differentiations in designing questions such as creating the questions within everyday life framework and utilizing visual elements in those questions (Berberoğlu, 2012). The exam name was transition from primary school to middle school (TEOG) with the latest regulations made in 2013. TEOG is done two times in an academic year, one is at the end of the

autumn and the other is at the end of spring term, which was done by only the senior students. It is observed in TEOG questions that usage of visual elements proceeds in terms of everyday life.

Approximately, half of the mathematics questions of the tests done for transition to high school in Turkey between 2010 to 2015 years contained visual elements. The same approach is common in middle school mathematics textbooks and teacher made tests. However, there is no research in the relevant literature that aims to identify differentiation in students' performance on verbal or visual mathematics questions and in students' approaches towards questions.

Thus, whether visual elements serve the purposes in mathematics questions, how test and item statistics differs based on visual or verbal expression of the question and students' views and approaches regarding the visual questions are unknown.

Consequently, to compare the students' performance on visually and verbally expressed mathematics questions is seen as a requirement. According to this requirement, the aim of this study is to identify whether there is a relation between visual elements in mathematics questions with test and item statistics, and students' approaches. In line with this general aim, the following questions will be answered:

1. Is there statistically significant difference between items difficulties of correspondent items in visual and verbal forms?
2. Is there statistically significant difference between items discrimination of correspondent items in visual and verbal forms?
3. Is there statistically significant difference between test reliability obtained from verbal or visual questions?
4. Is there statistically significant difference between mean scores of visual and verbal forms?
5. Is there statistically significant difference between response times of visual and verbal forms?
6. What are students' views about correlation between visual ingredients and individuals' approaches towards questions and test performance?

METHODOLOGY

In the scope of the study, first the data obtained from implementation of achievement test were collected. Then qualitative data were collected through interviews with the respondents, and finally the results of two operations were interpreted together. Fraenkel et al. (2011) described this method as explanatory mixed research. As stated in explanatory mixed studies, the researcher made a quantitative study; however, additional information was required to flesh out the results. Thus, the researcher refined and did a follow up of the quantitative findings by using the qualitative method. With the qualitative findings, the results of the quantitative phase of the study became deeper (Creswell and Clark, 2007).

Participants

The study was conducted with middle school students from Ankara, Çankaya. The study group in the research was selected through convenience sampling. Convenience sampling is defined by Fraenkel et al. (2011) as "a group of individuals who are available for a study." The study group is composed of 292 seventh grade students from 10 middle schools in the spring term of the academic year of 2015 to 2016. All the participants took one form of the mathematics achievement tests which were constructed by the researcher. After the test, 10 students were selected for the interview based on the teachers' guidance. As suggested in the literature, the number of participants was determined by "saturation rule". According to the saturation rule, when there are no new data and themes, it is not possible to replicate the study, for the data are saturated and the data collection process is finished (Guest et al., 2006). Based on this approach, when the students' statements become similar, it was then concluded that the interviews conducted were completed. Since the statements became similar and they do not provide any new data after the 10th student, the interviews had to be stopped.

Data collection tools

Quantitative data are collected through the instrument developed by the researcher. Seventh Grade Mathematics Achievement Test is constructed in two different forms as verbal and visual. The researcher developed the tool in accordance with the objectives described in the seventh grade mathematics instructional program. In order to determine the difference between students' performance in visual and verbal tests, correspondent questions measuring the same objectives were designed in both forms. The content of question in the visual questions is explained through picture, graphic and photos whereas in the verbal questions it is described through words. Due to equivalence of these two forms, experts and respondents' recommendations and feedbacks are considered, consequently required regulations are made. While constructing the measurement tool, visuals are noted to serve a particular function described in the literature (Clark and Lyons, 2004). Therefore, the visuals in the measurement tool have four basic functions. The questions in the achievement tests are divided into four groups by the researcher:

1. Regulator (visuals aim to regulate the data given in the problem in order for data process); 4, 5, 6
2. Informative (visuals that describe the stages of the process given in the problem); 10, 11
3. Descriptive (visuals that describe the situation given in the problem); 1, 2, 3, 7, 8, 14
4. Demonstrative (visuals including mathematical models); 9, 12, 15

A number of experts' views are referred to on account of content validity of the instruments including visuals to fulfill those four functions. In order to eliminate the threat for internal validity, all teachers and experts are requested to examine:

1. Whether the questions are equivalent in terms of meaning (does the verbal-only item have the same meaning with the visual form of it)?
2. In terms of the functionality of visuals, does the visual contribute to the answers or their understanding?

Additionally:

1. Two mathematics teachers in middle school examine the test questions to find out whether they are appropriate for children's reading comprehension level, existing knowledge, and assessment tools to which the students are familiar with.
2. Six measurement and evaluation experts examined the test

forms to find out whether distracters work, the problem expression is clear and it is overlapping with the features in the indicator table 3. Two mathematics education experts examined the questions to understand if there is any scientific error in it.

After making required regulations, the latest form of the test consisted of 19 constructed questions. Pilot application was done with 100 seventh grade students. In accordance with the results of the pilot application, three correspondent questions were removed from the test because they did not work. Moreover, in line with students' views, the expressions of three questions were changed. As a result of this pilot practice, the final form with 15 questions was constituted.

Semi-structured interview forms were utilized for qualitative data collection. In semi-structured interviews, questions are prepared before interview; however, they are detailed with questions asked during interview (Finn et al., 2000).

Semi-structured interviews have some advantages such as easy anlysis, allow participants to express their own ideas and provide deeper knowledge (Büyüköztürk et al, 2012). The interview form is answered by students who take the test form. The interview form is prepared based on the findings of former studies that determine the difference between students' performances in visual and verbal test questions.

The aim of these interview questions is to identify if individuals have similar approaches to the visual test question as described in the literature, and to also introduce possible new approaches that have not been mentioned in the literature yet. However, questions are conscientiously generated as open-ended form in a non-manipulative way.

Draft interview form was examined by a measurement and evaluation expert, and received feedback; after that, pilot application of the form was done with five seventh grade participants. Consequently, the final form of the interview form was prepared.

Data collection procedure

The data collection procedure was conducted in two steps. In the first step, which is quantitative data collection, 292 seventh grade students from 10 middle schools in Ankara, Çankaya were involved. All the participants take the mathematics achievement test. Before the test session, the participants were informed about the purpose of the study and the approximate response time.

In each class, half of the students apply the verbal form, and the other half applied the visual form of the test. The students were also asked to score their willingness to answer the test on 5 points (1 is the highest score). Moreover, the students reported the time when they finish the test. During data collection process, the researcher realizes that the students who take the verbal form have tendency to drop answering items. More verbal forms are being copied and applied to students to eliminate the potential imbalance between verbal and visual form applications.

In the second step which is the qualitative data collection, 10 students were interviewed. They were selected based on teachers' guidance. In this step, the aim was to provide deeper explanation for the test and item statistics. Interview is a mutual and interactive process during which the predetermined and purposive questions are answered by the participants (Stewart and Cash, 1985).

In semi-structured interview process, the researcher has the flexibility to direct new questions based on the students' answers, and ignores the pre-determined questions. The interviews are conducted as two sessions for two groups with five students in the teachers' room. Each interview session takes one hour. The statements are recorded and reported by the researcher. To make students state their real opinion confidently, it is aimed to provide a comfortable and silent environment.

To increase the validity of the interview, the questions were constructed away from directing the students to certain answers. Before the interview, the students were given the test papers to check and remember the test items and their approaches to the solution. Direct citations were made from the students' statement to increase the reliability of the study.

Analysis of the data

First, the data collected by achievement test was transferred to an electronic form; then Excel program was used to calculate the item statistics. The item difficulty indexes of each question were computed. Then the following equation 1 (the so-called z-test) was used to determine if there was a significant difference between the indicators of the difficulty of the correspondence questions in both tests separately (Akhun, 1991).

$$z = \frac{P_1 - P_2}{\sqrt{PQ(\frac{1}{N_1} + \frac{1}{N_2})}} \quad (1)$$

P1: Right answer percentage in the first sample
P2: Right answer percentage of the second sample
P: The weighted percentage of both samples
Q: 100-P
N1: Size of first sample
N2: Size of second sample

In line with another sub-research question, as the normality assumption was satisfied, "point-biserial correlation coefficient" was computed for item discrimination of each question. Then, the following equation 2 was used to transform the Pearson correlation index r into Fisher's transformation Zr (Akhun, 1991).

$$Z_r = \frac{1}{2} log_e \frac{1+r}{1-r} \quad (2)$$

Z_r: Transformation of correlation coefficient to Fischer's z Coefficient
r : Correlation coefficient

The significance of the difference between the transformed correlation coefficients was identified by the following equation (3)

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \quad (3)$$

Z: Student's t statistics of the difference between the transformed correlation coefficient
Z_{r1}: Transformed correlation coefficient from first sample
Z_{r2}: Transformed correlation coefficient from second sample
n_1: Size of first sample
n_2: Size of second sample

In line with the other sub-research question of the study, the difference between verbal and visual test statistics was investigated. As a result of the t-test, it was decided that there is a significant difference between the mean of the total test scores and the response time of each test form. Moreover, test form reliability

Table 1. Z-test results of the item difficulty indexes from visual and verbal forms.

Item no.	Item difficulty index		z values
	Visual form	Verbal form	
1	0.76	0.54	3.27*
2	0.67	0.63	0.58
3	0.70	0.46	3.09*
4	0.66	0.41	2.98*
5	0.60	0.58	0.26
6	0.53	0.46	0.77
7	0.73	0.70	0.51
8	0.82	0.78	0.93
9	0.82	0.78	0.93
10	0.51	0.31	2.02*
11	0.52	0.32	2.04*
12	0.56	0.52	0.47
13	0.53	0.45	0.88
14	0.54	0.76	-3.27*
15	0.78	0.71	1.32

scores were computed by Cronbach reliability coefficient. Statistically, significant difference between the test reliabilities was tested by Z test after Fisher's Z transformation as stated in equation 3.

Qualitative data obtained by interviews were examined by descriptive analysis. The aim of descriptive analysis is to organize and interpret the findings of the study before presenting it to the reader (Yıldırım and Şimşek, 2008). Since this study aims to constitute a general framework of students' views about visual ingredients in the test items, descriptive analysis was performed. With this aim, voice records and notes taken during interviews were transferred to an electronic form in a computer and then analyzed. Students' answers to the interview questions were read again and again. For the requirement of qualitative data analysis, participants' statements related to the students' test performance and their approach to the test item were corresponded to a code. Then the researcher examined the codes and calls the codes which are related to each other and cluster in scope based on the same concept.

Literature review and the aims of the study were considered while coding. At this point, the validity of the qualitative results is a critical point. In the study, the method of external audit was applied to eliminate the threat to the validity of results (Creswell, 2012).

In this method, an external auditor examines the study for some points such like the appropriateness of categories, and if the inferences are logical etc. (Schwandt and Halpern, 1988). An external auditor, the narrative account becomes credible. As the external auditor documents and reviews a study, the credibility of the study increases (Creswell and Miller, 2000).

In this study, coding the statement by only one researcher may be a threat to the validity of the study. To eliminate this threat, the consistency of results of the analysis is checked by other experts in the measurement and evaluation field. The experts examined the codes and categories to find out whether the codes were placed in the same category.

Therefore, no disagreement exists between the researcher and the external auditor. After reaching an agreement about the codes, the final results were reported. Categories were constituted as a consequence of coding and findings were tabulated under three main categories.

RESULTS AND DISCUSSION

Quantitative data analysis

Findings about the difference between item difficulties of verbal and visual test questions

Based on the first sub research question, the aim is to identify if there is a significant difference between item difficulty indexes of correspondent questions from both two achievement test forms. For this objective, first item difficulty indexes are calculated for each item. Then, the significance of the values obtained from each item pair is tested. Item difficulty indexes and z-values are presented in Table 1.

Table 1 illustrates that the difficulty indexes of six questions have a significant difference. In general, the difficulty indexes of the items in visual test form are lower; only item 14 had a higher difficulty index. Students' interviews reveal that this situation is related to learning outcome measured by the item. Similarly in item 14, questions 1 and 3 in the visual test form have significantly higher difficulty indexes than verbal form. This could be due to the fact that representing the question in the mind of students is easier with the help of visual element.

However, questions 2, 7 and 8 do not have a significant difference in terms of difficulty indexes. It could be that the use of visuals did not help the students to represent the question in their minds. For questions 10 and 11 consisting of both verbal and visual forms, the difference

Table 2. Z-test results of the item discrimination indexes from visual and verbal forms.

item no.	Item discrimination index				Z
	Visual form		Verbal form		
	r_{bis}	Z_r	r_{bis}	Z_r	
1	0.56	0.63	0.37	0.39	2.04*
2	0.52	0.58	0.37	0.39	1.57
3	0.52	0.45	0.32	0.58	1.07
4	0.54	0.60	0.34	0.35	2.09*
5	0.50	0.55	0.30	0.31	2.00*
6	0.56	0.63	0.53	0.59	0.36
7	0.56	0.63	0.32	0.33	2.51*
8	0.56	0.63	0.42	0.45	1.55
9	0.38	0.40	0.35	0.37	0.29
10	0.50	0.55	0.30	0.31	2.00*
11	0.59	0.68	0.39	0.41	2.22*
12	0.59	0.68	0.57	0.65	0.25
13	0.52	0.58	0.31	0.32	2.13*
14	0.46	0.50	0.48	0.52	-0.21
15	0.56	0.63	0.47	0.51	1.02

between difficulty indexes in favor of visual form may result from the functionality of informative visuals. Such informative visual questions allow students to understand the steps of finding solutions to problems more easily. It is observed that item difficulty indexes of questions 5 and 6 are close to each other. These questions are called regulators in relevant literature, and they enable one to present the data given in the question systematically like a graph. This result demonstrates that there is no difference between the item difficulty indexes of regulator visual or verbal mathematics questions.

Studies in the literature have revealed contradictory findings about the impact of visual usage on item difficulty indexes. Suh and Grant (2014) examined the history questions in National Assessment of Educational Progress (NAEP) exam applied in a particular year through descriptive method. The results of the study indicate that non-visual questions are more difficult than visual questions.

In a test developed by Vorstenbocsch et al. (2013) for the purpose of the heart anatomy course achievements, the impacts of the usage of answer list or visual on item statistics are examined. The study has demonstrated that different kinds of visuals affect item statistics in different level. However, Civelek (1998) does not reveal a significant difference in the study conducted through electrical circuits. Observing no significant difference between item difficulty indexes may be related to using decorative visuals in those questions predominantly. Unlike the literature in this study, to observe a significant difference between verbal and visual forms of some certain functions may be associated with visuals used not only for decorative purpose but also to facilitate the understanding of the question.

The difference between item discriminations of verbal and visual test questions

In relation with the second sub research question, it is determined whether item discrimination values differentiate two types of the achievement test forms. First, item discrimination indexes are calculated, and then these indexes are transformed to Fisher's Z_r values. Their two values are compared. Item discrimination indexes were obtained from two forms, and Fisher's Z_r values are presented in Table 2.

Acceptable item discrimination value is 0.30 and above (Crocker and Algina, 1986). Table 2 indicates that item discrimination values in visual test are between 0.38 and 0.59, while in verbal test they vary between 0.30 and 0.57. These values are within the critique values described in the literature. Moreover, Table 2 reveals that there is a significant difference between item discrimination indexes in favor of visual test form. It is observed that only question 14 in the verbal test is insignificantly higher.

Furthermore, Table 2 indicates that there is a significant difference between the informative test questions (10 and 11) of the two test forms in terms of item discrimination indexes in favor of visual test. Observing the same situation in the difficulty indexes is a proof that informative visual questions make the item to be more qualified. Also, item discrimination indexes of the first two of questions 4, 5 and 6 (which include regulator visuals) have a significant difference in favor of visual test. Presenting the data given in the stem of question in a

Table 3. T-test results of the difference between reliability of visual and verbal test forms.

Visual form		Verbal form		z value
KR-20	Z_r	KR-20	Z_r	
0.78	1.05	0.72	0.91	1.15

Table 4. T-test results of the difference between total score of visual and verbal test forms.

Variable	N	Mean	t	p
Visual	167	10.14	3.48*	0.00
Verbal	125	8.52		

*p<0.01.

more regular way influences the performance of answerers in a positive manner.

Questions 9, 12 and 15 that include mathematical models in the tests have no significant difference between item discrimination indexes. In order to interpret this situation correctly, answer sheets are examined and it has been observed that students could answer those questions by making the necessary drawings. Moreover, the familiarity of students with using models such as Venn diagram may be the determinant of the item discrimination of these questions. Students can answer the questions they are familiar with without any visuals.

There are various studies that examine the difference between item statistics of verbal and visual test questions. One of those studies was carried out by Civelek (1998) through geometry questions. In the study, there were two separate test forms. One of those tests explains a triangle measure of angles in a figure and the other explains the same content with words. Two types of tests were given to the students. As a consequence, it is observed that there is no significant difference between item discrimination indexes of the tests.

Bağcı (1998) utilized questions prepared for examining the achievements of the topic of electric circuits. Electric circuits are expressed by figures in one of the forms and by words in another form. Results reveal that there is no significant difference of item discrimination indexes between verbal and visual tests. Researchers explain this insignificance with students' ability to make drawings when they need them.

Although the findings of this study are similar with those of Civelek (1998) and Bağcı (1998) study in one aspect, to conclude a general deduction is not possible about the difference in item statistics between visual and non-visual questions. Instead, according to the results of this study, it can be concluded that there are differences in item discrimination indexes between informative and regulative visuals that the interpretation of descriptive and demonstrative various visuals depends on test type.

The difference between test reliability of verbal and visual test questions

In relation with the third sub research question, KR-20 formula is used to calculate item reliability regarding the results of the application of two different types of test (visual and non-visual test). The reliability of the visual test form is 0.78 whereas verbal test reliability is 0.72. Consequently, it can be stated that the reliability of the verbal test is relatively lower than visual test. The significance of this difference is tested. For this, Fisher's Z transformation is done and the significance of Z_r values is examined. Those values are presented in Table 3. As Table 3 illustrates, there is no significant difference between verbal and visual test form in item reliability. This finding is consistent with other studies in the literature. For example, Civelek (1993) and Bağcı (1998) also found that there is no significant difference in the reliability of the two tests done.

The difference between mean scores of verbal and visual test

T-test is performed in line with the second sub research question in order to observe if test scores means differ based on test forms. The results of t-test are presented in Table 4. The means of visual test scores are significantly higher than the means of verbal test scores. The studies in relation with this purpose reveal contradictory results in the literature. For example, Washington and Godfrey (1974) examined the visual questions of American Air Force Specialty Exam; De Melo (1980) examined the visual questions of biology test: both studies indicated that visual questions are more advantageous than non-visuals. Moreover, Duran and Balta (2014) conducted a study through SBS science questions, and concluded that the mean of the visual test scores is significantly higher than the mean of the verbal test scores. Accordingly, the number of questions left blank in the verbal form is higher than that in the visual form. Hall et al. (1997) state that students have higher performance in visual test forms because visual components make scientific contents to be more understandable.

The difference between response time of verbal and visual test

Depending on the second sub research question, t-test is performed to understand if response times differ significantly according to the test forms. The results are shown in Table 5. Table 5 demonstrates that the difference between answer times of students is significant

Table 5. T-test results of the difference between total response time in visual and verbal test forms.

Variable	n	Mean	t	p
Visual	167	28.89	5.28*	0.00
Verbal	125	36.68		

*p<0.01.

in favor of visual form. This result is consistent with the literature. Saß et al. (2012) which asserts that test questions consisting only of visual elements in the stem lead to different answers. From this finding, the results of the interviews with students are determinant.

The analysis of qualitative data

To present the students' expressions, those who answered the verbal form questions are notated as "verbal" while those who answer the visual form questions are notated as "visual" in the following part of the report.

Students' views about the difference in preference to answering verbal or visual mathematics questions

In order to determine the level of willingness of students to answer the questions during the test, they are requested to rate their willingness between1 to 5 (1 is the lowest and 5 is the highest). The mean of the ratings of visual test answers is 3.9 while the mean of the ratings of verbal test answers is 3.0.

Although, this finding cannot be interpreted properly as there is a significant difference in willingness in favor of visual test answers as this test may be influenced by a number of factors, visual test answerers can be stated as more willing to answer. In support of this situation, a study by Peeck (1993) emphasized that educational materials containing visuals increase the willingness of students to answer questions. Table 6 contains the codes related to preferableness of visual questions.

Nine respondents state that individuals answering visual test form are more advantageous. Visuals in test questions create an impression on nine answers as they are easier. All respondents specify that visual test questions can be described shorter and in parallel with this, they can be answered in a shorter time. Additionally, nine students conceive that answering visual questions is more practical.

Most of the respondents expressed that they think visual test is easier. Thus, respondents are more willing to solve visual test problems and therefore the probability of giving correct answer increases. The willingness of students to answer questions, and the persistence of students to think in a detailed way in order to answer the question instead of superficial thinking increase the probability of giving correct answer (Whimley and Lochhead, 1999). Accordingly, Shepard (1967) stated that visuals included in problems have more positive impact than words and increase the students' willingness to answer. Abedi et al. (2003) also conceive that visual components make questions easier for respondents. In support of these opinions aforementioned, students who think that visual questions are easier to answer make more effort to answer. Hence, it can be said that there are differences in the answering behaviors of visual and verbal test forms. Only one respondent specifically stated that he does not prefer to answer visual test form. According to this respondent, asking the same content using visuals or words does not differentiate the difficulty of the question.

Verbal 1: There was no visual in my test but everything about the question was described. The same questions are given to my friends with visuals in the other test. Eventually, both two tests requested for the same thing but the question was longer in mine whereas it was shorter in theirs. Therefore, both of them are equally difficult.

Students' views about the difference in comprehensibility of items between verbal and visual questions

Eight respondents who prefer to answer the visual test form express that those kinds of questions are more understandable. Codes related to the category of comprehensibility of the visual questions are presented in Table 7.

Table 7 illustrates that six respondents think that visual questions are more understandable because those visuals make it easy to represent the problem in the mind. Similarly, four respondents conceive visual questions are more understandable because they do not need to execute logical reasoning to understand the questions. Some respondents provided the following reasons:

Visual 5: I understand more easily with a picture. Sometimes I can even solve the problem without reading the whole question by just looking at the picture. I do not bother reading the questions. I do not want to read the question if it is long. I do not answer those long questions in other exams just because I am too lazy to read.

Visual 3: I am bored reading the question when it is long. In fact, I take notes to summarize the question and understand better. But the question can be expressed shortly when it contains visual.

Verbal 2: The visual presence in the question definitely makes me to understand. There was no visual in the test I answered, therefore, I had difficulty representing it in my mind. Hence, I think the visual test respondents are more

Table 6. The codes related to the category of "preferableness of visual questions".

Code	Frequency
Visual test form is easier	9
Questions in visual form are shorter	10
Solutions of those question in visual form are more practical	9
Replying time is shorter in visual form	10
It is easier to understand the visual form is easier	8
The reasons of those who do not prefer visual form	The number of individuals gave a reason
The same content expressed differently in both test forms	1

Table 7. Codes related to the category of "comprehensibility of the visual questions".

Code	Frequency
It reduces the burden of reading	10
It makes it easy to represent in the mind	6
We do not have to execute logic to understand the question	4
Trying to understand the picture is easier than understanding the sentences in the question	8

advantageous.

Verbal 3: I feel bored reading the question when it is long. I did not read the questions in this test because the questions are long. The questions are shorter with a picture, and it makes it easier to read. Eight respondents mentioned that trying to understand pictures is easier than trying to understand words. Therefore, they mentioned visuals in questions make them more understandable. Those expressions of the respondents reveal the students have tendencies to be bored of reading.

Thus, opinions about shorter questions are preferable and more understandable by means of containing visuals illustrate similarities. Respondents stated that they could reach all information they need for solution by just looking at the visual. This situation mentioned by students takes part in the literature as "mistake in reading" which is one of the error sources in problem solving process (Whimley and Lochhead, 1999).

According to this, error sources which respondents encounter at the stage of understanding the question are put in the following order: reading the question without focusing enough, skipping some words while reading or not being able to focus on the meaning while trying to read fast because of not paying enough attention. The students' expressions support these as well. Although respondents do not feel entitled, they state that they prefer to interpret visuals rather than words as visual questions reduce the burden of reading by reducing the number of words located in the stem of the question.

The respondents of the visual test were asked if they prefer to visualize differently any one of the visual questions in order to determine whether mental representations and existing visuals are overlapping or not. Students reflect to change only the third question visual, in addition they express that they need more examples to understand the rules of the pattern. This situation is parallel with the other students' views who answered verbal test form. Some students' views about this question are as follows:

Verbal 1: For example, I cannot understand the third question. I looked at the pictures on the class board in order to solve it.

Visual 5: I would draw more pictures for the third question as well; there should be 5 or 6 pictures at least.

As a consequence of the influence of the differentiation in the test item on the test statistics, it is observed that visual form has more acceptable values. It is interpreted as a proof of mental representation, and the test visuals are overlapping that students do not need to visualize the test questions differently. Therefore, it is concluded that visual mathematics questions are more understandable than verbal ones.

Students' views about the difference in responsiveness of questions expressed visually or verbally

Nine respondents think that the solutions of the visual questions are more practical. Six respondents express making drawings and transactions on the given visual are enough to solve the problem. Table 8 demonstrates the codes related to category of responsiveness of visual questions. Table 8 indicates that all of the respondents say reading and solving the visual problems take less time. Seven respondents reflect that questions can be solved with the information given in the visual while eight

Table 8. Codes related to category of "responsiveness of visual questions".

Code	Frequency
Questions can be solved with the information given in the visual so there is no need to read the question	7
The questions can be solved by making drawings and transactions on the visual	8
Less time is consumed to understand and solve the problem	10

respondents express that they get the solution by making transactions on the visual. One expression of a respondent regarding the issue is:

Verbal 2: I can answer the question without reading if there is a picture. Everything is already given in the picture for the solution. I make transactions on it and this makes me faster.

Another respondent states that visual questions can prevent possible errors and mistakes as there is no need to make drawings for the solution.

Visual 5: The visual presence in the question makes it definitely more practical. Because we can make mistakes while drawing for solution and so we cannot solve the question. However, when the drawing is given in the question, we can make transactions on it and solve it more easily. Particularly in coordinate plane questions, we can solve the problem without dealing with drawing.

The expressions of students are in parallel with the literature. The mistakes made in visualization of situations and relations described in the question are one of the obstacles that make students not to give correct answer due to "inaccuracy in thinking" (Whimbey et al., 1999). The respondents of this study stated that visuals reduce the possibility of making mistake. All the students accept that they can focus on the visual test more easily.

According to both qualitative and quantitative results, it would be more appropriate to make inferences specific to the functions of visuals instead of concluding a general outcome regarding the differentiation in performance of verbal and visual mathematics questions. Positive expressions of students related with the visual questions and answering the visual questions in a significantly shorter time support the claim that those components make positive contribution to willingness to answer the test.

In detail, using visuals to describe staggered issues expressed in the question stem turns test statistics in favor of visual form as a result of reducing the burden of reading. Accordingly, providing the data given in the question stem in a regulated way by graphs does not increase the correct answer possibility but it has a positive impact on the indexes of item discrimination.

However, it would not be possible to make generalizations for the differences in item statistics of the questions which include words or visuals to figure out the problem situation. As one of the results of the study, there is no significant difference between the reliability of visual form and verbal test form. However, the students' expressions reveal that they have tendency to prefer the visual form. This is also reflected in the data collection process. More students took the verbal form cancel test application without even reading the question; however, the students who take the visual form generally pay attention till the last question. In order to balance the number of students who respond to the visual and verbal forms, the latter is applied more by students. Therefore, after excluding the missing data related to loss of participants, the number of students who answered each form is balanced. Therefore, it may be concluded that the existence of visual in test item may reduce the error from the instrument and answerer by making the students more willing and less reluctant to answer the test item.

Conclusion

In line with the aim of this study, different approaches of individuals to questions and different item statistics between verbal and visual mathematics questions are examined. In general, item difficulty indexes are found lower in visual questions. The difference between item statistics is changeable in terms of the function of the visual.

In those questions which describe the situation expressed in the question stem, item difficulty indexes and item discrimination indexes differ depending on the content of the problem. However, the difference between item difficulty indexes and item discrimination indexes is significant in those questions, which describe the stages of a process given in the stem by using visuals. In the questions including the visuals such as graphics and table which have regulation functioning, the difference between item difficulty indexes is not significant. In addition, item discrimination indexes are computed higher in visual form questions. This is valid for the mean of the test scores and response time as well. However, thereis no significant difference between test reliabilities.

Overall, from the students' expressions, it is concluded that they have more positive attitude towards visual questions, and the visual in question makes them to perceive the questions as easier. Moreover, the respondents approach visual items positively as they reduce the burden of reading. Also, they express that visuals make questions both understandable and speed up response process. They describe visuals as more perceptible due to the fact that they get solution less

logically, visuals reduce the burden of reading and that visuals facilitate mental representation. They also say it is easier to understand visual than words. Finally, the respondents state that visual questions are more answerable on account of being able to solve the problem with only given information in visual, of being able to reach the conclusion by making transactions on the picture 2 and of answering in a shorter time.

The test and item statistics are parallel to the students' expressions. The students tend to prefer visual form. Therefore, their test performance is better in the visual form than the verbal form as can be seen from the significant difference between the test score mean of both forms. There is significant difference in the response time and it is supported by the students' expressions. The students state that the visual questions speed up the understanding and answering process.

Similar to the students' expressions about the easiness of understanding the visual question, there is a significant difference between the item difficulty and item discrimination where the visual describes a process in the items. This is because in such question, the visuals make it easy to concretize the process in the test situation. This is also valid for the visual questions in which the data are organized by tables, graphics etc. As the students do not use effort to organize the data, the item discrimination increases. All in all, to observe a significant difference between verbal and visual forms of some certain functions may be associated with visuals used not only for decorative aims but also to facilitate understanding of questions.

Recommendations

Based on the results of this study, some recommendations can be made for future researchers. The participants of this study were 7th grade students. Another study can be conducted with students from different grade level. Moreover, the learning area can be changed. Mathematics questions are examined in this study; science or social sciences questions can be examined in another study. Another point worth examining in future studies is the differentiation of statistics in terms of the students' achievement level or level of their spatial intelligence. In contrast with the literature, it is concluded that there are differentiations in the performance of students, and in test and item statistics of the visual questions. This issue is explained by effective usage of visuals apart from decorative aims. Therefore, it may be recommended that the operators integrate purposive and age-appropriate visuals into mathematics problems to increase their willingness to answer them.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGMENT

This study was summarized from the master's thesis prepared by Cagla ALPAYAR at Ankara University, Institute of Educational Sciences under the advisory of Assist. Prof. Dr. H. Deniz GULLEROGLU, a part of this study was presented at the 4th International Eurasian Educational Research Congress (May 11-14, 2017)

REFERENCES

Abedi J, Courtney M, Leon S (2003). Effectiveness and validity of accommodations for English language learners in large-scale assessments. CSE Report 608. University of California, Los Angeles.

Ahmed A, Pollitt A (2000). Observing context in action. A paper presented in IAEA Conference, Jerusalem. Retrieved from http://www.cambridgeassessment.org.uk/Images/109669-observing-context-in-action.pdf.

Ahmed A, Pollitt A (2007). Improving the quality of contextualized questions: An empirical investigation of focus. Asses. Educ. 14(2):201-232.

Akhun I (1991). Statistical significance and sample [İstatistiklerin manidarlığı ve örneklem]. Ankara: Kendi Yayını.

Anastasi A (1982). Psychological testing, U.S.A.: 7th edition Macmillan.

Antoinetti A (1991). Why does mental visualization facilitate problem-solving? Adv. Psychol. 80:211-227.

Anagnostopoulou K, Hatzinikita V, Chrisdou V (2015). Comparing international and national science assessment: what we learn about the use of visual representations. Educ. J. Univ. Patras UNESCO Chair 2(1):96-110.

Bağcı M (1998). The effect of formal characteristics of item root on the item validity and reliability of test in geometry test [Geometri testinde madde kökünün biçimsel özelliklerinin madde geçerliğine ve testin güvenirliğine etkisi] (Unpublished master dissertation). Hacettepe University, Ankara.

Beb-Chaim D, Lappan G, Houang RT (1989). The role of visualization in middle school mathematics curriculum. Focus Learn. Problems Math. 11(1-2):49-60.

Berends IE, van Lieshout EC (2009). The Effect of illustrations in arithmetic problem-solving: Effect of increased cognitive load. Learn. Instruction 19(4):345-353.

Berberoğlu G (2012). Content validity [Kapsam geçerliği]. Cito Eğitim: Kuram ve Uygulama, 15:9-16.

Brown M (1999). One Mathematics for All. In C. Hoyles, M. Morgan & G. Woodhouse (Eds.). Rethinking the Mathematics Curriculum. Falmer Press: London. pp. 78-89.

Büyüköztürk Ş, Çakmak EK, Akgün ÖE, Karadeniz Ş, Demirel F (2012). Scientific research methods [Bilimsel araştırma yöntemleri] Ankara: Pegem Akademi Yayimcilik.

Carpenter PA, Marcel AJ, Peter S (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. Psychol. Rev. 97(3):404-431.

Civelek M (1998). The effect of using shapes in multiple choice test items on the item and test properties [Çoktan seçmeli test maddelerinde şekil kullanmanın madde ve test özelliklerine etkisi] (Unpublished master dissertation). Hacettepe University, Ankara.

Clark RC, Lyons C (2004). Graphics for learning: proven guidelines for planning, designing and evaluating visuals in training materials. San Franciso: Pfeiffe.

Creswell JW, Clark VLP (2007). Designing and conducting mixed methods research. London: SAGE Publication.

Creswell JW, Miller DM (2000). Determining validity in qualitative inquiry. Theor. Pract. 39(3):124-130.

Creswell JW (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Pearson Education: Boston.

Crisp V, Sweiry E (2003). Can a picture ruin a thousand words? Physical aspects of the way exam questions are laid out and the impact of changing them. A paper presented in British Educational

Research Association Annual Conference, Edinburgh. Retrieved from at www.ucles-red.cam.ac.uk.
Crocker L, Algina J (1982). Introduction to the classical and modern test theory. New York: Wadsworth Thompson Learning.
De Melo HT (1980). Visual self-paced ınstruction and visual testing ın biological science at the secondary level (Unpublished doctoral dissertation). Pennsylvania State University.
Diamond MS (2008). The impact of text-picture relationships on reader recall and inference making: a study of fourth graders' responses to narrative picture books. (Unpublished doctoral dissertation). Temple University, Philadelphia.
Demirtaşlı NÇ (2010). The element of everyday life in measurement of high-level thinking skills [Üst düzey düşünme becerilerinin ölçülmesinde günlük yaşam unsure]. CITO: Kuram ve Uygulama 7:9-26.
Duran M, Balta N (2014). The influence of figured and non-figured questions on secondary students' success at science exams. Pak. J. Stat. 30(6):1279-1288.
Ebel RL (1965). Measuring educational achievement. New Jersey: Prentice-Hall.
Erkuş A (2003). Writing on psychometry [Psikometri üzerine yazılar] Ankara: Türk Psikologlar Derneği Yayınları.
Finn M, White ME, Walton M (2000). Tourism and leisure research methods: Data collection, analysis, and interpretation. Essex: Pearson Education Limited.
Filippatou D, Pumfrey PD (1996). Pictures, titles accuracy and reading comprehension: a research review (1973-1995). Educ. Res. 38(3):259-291.
Fraenkel JR, Wallen NE, Hyun H (2011). How to design and evaluate research in education (8th ed.). New York: McGraw-Hill Education.
Guest G, Bunce A, Johnson L (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods 18(1):59-82.
Furst EJ (1958). Constructing evaluation instruments. London: Longman group.
Habre S (2001). Visualization enhanced by technology in the learning of multivariable calculus. Int. J. Comput. Algebra Math. Educ. 8(2):115-130.
Haladyna TM (1997). Writing test ıtems to evaluate higher order thinking. Boston: Allyn & Bacon.
Haladyna TM, Downing SM, Rodriquez MC (2002). A review of multiple choice item writing guidelines for classroom assessment. Appl. Meas. Educ. 15(3):309-334.
Hall VC, Bailey J, Tillman C (1997). Can student-generated illustrations be worth ten thousand words? J. Educ. Psychol. 89(4):677-681.
Handono EB (1996). Effective and visualization for learning materials for girls and woman, LRC Training Workshop in Phnom Penh: Cambodia. Retrieved from http://www.accu.or.jp/litdbase/pub/dlperson/98LRC/98LRC_05.pdf.
Hillocks G (2002). The testing trap: How state writing assessments control learning. New York: Teachers College Press.
İşler ŞA (2003). Place and importance of illustration use in written course materials [Yazılı ders materyallerinde illüstrasyon kullanımının yeri ve önemi]. Milli Eğitim/ Eğitim- Kültür- Sanat Dergisi 157:55-63.
Kaptan F (1985). A qualitative research on SSPE (Student Selection and Placement Exam) physics questions [ÖSYS fizik soruları üzerinde bir nitelik araştırması] (Unpublished master dissertation). Hacettepe University, Ankara.
Kaufmann G (1985). A theory of symbolic representation in problem solving. J. Mental Imagery 9(2):51-70.
Kopriva R (2008). Improving testing for English language learners. New York: Taylor & Francis Grouping: Routledge.
Lanzig JWA, Stanchev I (1994). Visual aspects of courseware engineering. J. Comput. Assisted Learn. 10(2):69-80.
Levie WH, Lentz R (1982). Effects of text illustrations: a review of research. Educ. Commun. Technol. J. 30(4):195-232.
Levin JR (1981). On the functions of pictures in prose. In: Pirozzolo, F. J., and Wittrock, M. C. (Eds.) Neuropsychological and Cognitive Processes in Reading: New York: Academic Press. pp. 203-228.
Lohr LL (2003). Creating graphics for learning and performance: Lessons in visual literacy (2nd ed.). Upper Saddle River, New Jersy: Merrill.
Luchins AS (1942). Mechanization in problem-solving: The effect of Einstellung. Psychological Monographs, Washington: Am. Psychol. Assoc. 54(6).
Mevarech ZR, Stern E (1997) Interaction between knowledge and contexts on understanding abstract mathematical concepts. J. Exp. Child Psychol. 65(1):68-95.
Ministry of Education program (2009). Primary mathematics lesson 6-8 classes' curriculum and instruction. [İlköğretim matematik dersi 6-8. sınıflar öğretim programı ve kılavuzu]. Ankara: MEB Basım".
Murphy SJ (2009a). The power of visual learning in secondary mathematics education. Retrieved from https://assets.pearsonschool.com/asset_mgr/legacy/200916/MatMon092291HS2011StuMur_LR_20702 _1.pdf.
Murphy SJ (2009b). Visual learning in elementary mathematics research ınto practice mathematics: how does visual learning help students perform better in mathematics?. Retrieved from http://mathematicsuniversity.com/research/visual.pdf.
National Council of Teachers of Mathematics (NCTM) (1999). Principles and standards for school mathematics. NCTM Publications.
National Research Council (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.
Nickerson RS (1965). Short-term memory for complex meaningful visual configuration: A Demonstration of Capacity. Can. J. Psychol. 19(2):155-160.
Nitko AJ (2004). Educational assessment of students. Englewood Cliffs, New Jersey: Pearson Education.
Osterlind SJ (1989). Constructing test items: Multiple-choice, constructed response, performance and other formats. (2nd ed.). Boston/Dordrecht/London: Kluwer Academic Publisher.
Paivio A (2013). Imagery and verbal processes. Psychology Press.
Paivio A (1990). Mental representations: A dual coding approach. Oxford: Oxford Psychology Series.
Peeck J (1974). Retention of pictorial and verbal content of a text with illustrations. J. Educ. Psychol. 66(6):880-888.
Peeck J (1993). Increasing picture effect in learning from illustrated text. Learn. Instruction 3(3):227-238.
Phillips LM, Norris SP, Macnab JS (2010). Visualization in mathematics, reading and science education. Dordrecht: Springer.
Pollitt A, Ahmed A (1999). A new model of the question answering process. A paper presented in International Association for Educational Assessment Conference, Slovenia. Retrieved from http://www.cambridgeassessment.org.uk/Images/109651--a-new-model-of-the-question-answering-process.pdf.
Popham WJ (2000). Modern educational measurement: Practical Guidelines for Educational Leaders (3rd edition). Needham, MA: Allyn & Bacon.
Rasmussen C, Bisanz, J (2005). Representation and working memory in early arithmetic. J. Exp. Child Psychol. 91(2):137-157.
Rodriguez MC, Haladyna TM (2013). Writing selected-response items for classroom assessment. In: J. H. McMillan (Ed.) Sage Handbook on Research on classroom assessment, Thousand Oaks: SAGE Publications, Inc. pp. 293-312.
Salona-Flores G, Wang C (2011, April). Conceptual framework for analyzing and designing illustrations in science assessment: Development and use in the testing of linguistically and culturally diverse populations. A paper presented in Congress of National Measurement and Evaluation Commission New Orleans University: L.A.
Salend SJ (2009). Classroom testing and assessment for all students: Beyond standardization. Thousand Oaks: Corwin Press.
Saß S, Wittwer J, Senkbeil M, Köller O (2012). Pictures in test items: Effects on response time and response correctness. Appl. Cogn. Psychol. 26(1):70-81.
Schwandt TA, Halpern ES (1988). Linking auditing and meta-evaluation: Enhancing quality in applied research. Newbury Park, CA: Sage.
Schiffman P (1995). Low grade metamorphism of mafic rocks. Wiley Online Library 33(1):81-86.
Sharma MC (1985). Visualization. Math Notebook 4(5-6):1-2.
Shepard RN (1967). Recognition memory for words, sentences, and pictures. J. Verbal Learn. Verbal Behav. 6(1):156-163.
Shorrocks-Taylor D, Hargreaves M (1999). Making it clear: A review of

language issues in testing with special reference to the national curriculum mathematics tests at key stage 2. Educ. Res. 41(2):123-136.
Shriver KA (1997). Dynamics in document design: Creating texts for readers. New York: Wiley Publisher.
Standing L (1973). Learning 10,000 pictures. Q. J. Exp. Psychol. 25(2):207-222.
Stewart CJ, Cash WB (1985). Interviewing: Principles and Practices. Dubuque, Iowa: W. C. Brown Pub.
Stewart J, Van Kirk J, Rowell R (1979). Concept maps: A tool for use in biology teaching. Am. Biol. Teacher 41(3):171-175.
Sternberg RJ (1999c). The theory of successful intelligence. Rev. General Psychol. 3:292-316
Suh Y, Grant LW (2014). Assessing ways of seeing the past: Analysis of the use of historical images and student performance in the NAEP U.S.A. History Assessment, History Teacher 48(1):71-90.
Sweiry E, Crisp V, Ahmed A, Pollitt A (2002). Tales of the expected: The influence of students' expectations on exam validity. A paper presented in British Educational Research Association Conference, Exeter. Retrieved from http://www.cambridgeassessment.org.uk/Images/109695-tales-of-the-expected-the-influence-of-students-expectations-on-exam-validity.pdf
Sweller J (1994). Cognitive load theory, learning difficulty, and instructional design. Learn. Instruction 4(4):295-312.
Sweller J (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. Educ. Psychol. Rev. 22(2):123-138.
Thinsley HEA, Dawis RV (1974). The equivalence of semantic and figural presentation of the same test items. Educ. Psychol. Measure. 34(3):607-615.
Tout D, Spithill J (2015). The challenges and complexities of writing items to test mathematical literacy. In: Stacey, K. & Turner R (Eds.). Assessing mathematical literacy: the PISA experience. Australia: Springer. pp. 145-171.
Turkish Language Institutions (2016). Retrieved from http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.599b45efd3bfe8.83114184i.
Washington WN, Godfrey RR (1974). The effectiveness of illustrated items. J. Educ. Measure. 11(2): 121-124.
Whimbey A, Lochhead J, Narode R (2013). Problem solving and comprehension. New York: Routledge.
Whitely SE (1983). Construct validity: Construct representation vs nomothetic span. Psychol. Bull. 93(1):179-197.
Wiggins G (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey - Bass Publishers.
Vorstenbocsch MATM, Klaassen TPFM, Kooloos JGM, Bolhuis SM Laan RFJM (2013). Do images influence assessment in anatomy? Exploring the effect of images on item difficulty and Item discrimination. Anatomical Sci. Educ. 6(1):29-41.
Vekiri I (2002). What is the value of graphical displays in learning? Educ. Psychol. Rev. 14(3):261-312.
Yıldırım A, Şimşek H (2008). Qualitative research methods in the social sciences [Sosyal bilimlerde nitel araştırma yöntemleri]. Ankara: Seçkin Yayıncılık.
Yuill N, Oakhill J (1991) Children's problems in text comprehension: An experimental investigation. Cambridge: Cambridge University Press.
Zaraycki P (2004). From visualizing to proving. Teach. Math. Applicat. 23(3):108-118.