TÜRK FEN EĞİTİMİ DERGİSİ
Yıl 17, Sayı 4, Aralık 2020

Journal of
TURKISH SCIENCE EDUCATION
Volume 17, Issue 4, December 2020

http://www.tused.org

Student Achievement on the Concepts of Light and Shadow in Different Assessment Formats: Students' Learning Styles and Gender

Uygar KANLI[1], Ömer ILICAN [2]

[1] Assoc. Prof. Dr., Gazi University, Ankara-TURKEY, ORCID ID: 0000-0003-3076-475X
[2] Physics Teacher, Min. of National Educ., Ankara-TURKEY, ORCID ID: 0000-0003-1698-3796

Received: 05.10.2020 **Revised:** 26.11.2020 **Accepted:** 05.12.2020

The original language of article is English (v.17, n.4, December 2020, pp.468-486, doi: 10.36681/tused.2020.39)

Reference: Kanlı, U. & Ilıcan, Ö. (2020). Student Achievement on the Concepts of Light and Shadow in Different Assessment Formats: Students' Learning Styles and Gender. *Journal of Turkish Science Education,* 17 (4), 468 - 486.

ABSTRACT

This study was conducted to examine the achievements of students in the concepts of light and shadow measured in different assessment formats according to the learning style and gender. In this study, correlational survey model was used. The sample consisted of 10th grade (16 years) high school students (n=815) from different types of six high schools (public general and vocational) in Turkey who were selected by using stratified random sampling method. The students' achievement was determined using three different assessment formats, which contains open-ended test, multiple-choice test and structured communication grid test. The results show that there were significant differences between the students' achievement in light and shadow concepts when assessed by different assessment formats. While the achievement of boys was statistically higher in open-ended tests, girls were more successful in the structured communication grid tests. In addition, the test scores obtained by students in different test formats vary significantly according to their learning styles. For example, the mean scores of converger and accommodator-style students in open-ended tests were significantly different from the diverger-style students.

Keywords: Assessment formats, gender, learning styles, light and shadow, science achievement.

INTRODUCTION

How learning should be objectively assessed and measured has attracted the attention of people studying in this subject for so long because the objective assessment and evaluation of learning give information about not only to what extent a student learns but also about the sufficiency of educational programs, schools, administrators, and teachers (De Houwer, Barnes-Holmes & Moors, 2013). Therefore, the learning process is affected by the assessment and evaluation results, and the assessment and evaluation are affected by the features of learning. There is a relationship between the methods and formats used to assess the

Correspondence author e-mail: omerilican@gmail.com © ISSN:1304-6020

achievements, abilities, attitudes, and traits of students. Although traditional paper-pencil tests are generally preferred by teachers due to their reliability and validity, it can be observed that standardized tests are used in the teaching of mathematics and science courses. This situation actually leads teachers to using effective teaching methods less (Smith, Breakstone, & Wineburg, 2019). Hence, it is suggested that teachers use more suitable assessment methods regarding the subject to be learned by students (Herman, Ascbacher & Winters, 1992; Kulm & Malcom, 1991).

It is argued that complementary assessment methods, aimed at assessing the learning process besides traditional assessment methods, are more effective than traditional methods in terms of students' learning process. The methods such as open-ended and diagnostic tests, laboratory studies, portfolio, and structured grid can be given as examples of the complementary assessment methods (Resnick & Resnick, 1992; Wiggins, 1989). Harnisch (1994) asserted that when the complementary assessment methods are utilized in the assessment of individuals' achievements by considering individual differences, a more righteous assessment can be conducted. Also, how the students with different traits structure the knowledge in their brain can be understood when such an assessment method is used. For instance, Chen and Whitehead (2009) used structured grid tests to evaluate students' understanding of the process in their physics class, and as a result, they stated that grids can be an important tool in monitoring students' progress because many true and false answers are present in a pattern and there is little chance of guessing.

Regarding the individual differences, the frequently considered variable is gender (Anderson, 1989; Bell, 1997; Bolger, 1984; Lawrenz, Huffman & Welch, 2001; Logan & Hazel, 1999) because boys and girls differ from each other from a young age in terms of their interests and expectations. Students' perceptions of self-competence in different subjects may be affected by these gender differences and this may affect their science achievement. Murphy (1991) explained that boys tend to analyze issues separately while girls tend to analyze contextual features as an integral part of science tasks. Accordingly, girls usually define more complex multivariate studies which are more difficult to work on, but this difficulty is frequently perceived by teachers as a proof of girls' misunderstanding or inadequacy in science. Yip, Chiu, and Ho (2004) statistically analyzed the results of the HKPISA 2003 exam conducted by OECD-PISA, which included questions aimed at assessing the science literacy of 2437 students in the age group of 15. At the end of the research, it was seen that the boys were more successful in true-false type and multiple-choice questions about understanding scientific knowledge in the fields of earth, environment, and technology, whereas the girls were more successful in recognition/identification (open ended test) questions.

One other individual difference between students is their learning styles. The concept of learning style can be defined as "students' utilization of different and unique ways while they are getting ready to learn new and difficult knowledge or learning and remembering it" (Dunn & Dunn; 1993). In the relevant literature, there are many definitions about what learning is and how it happens (Kaya, Özarabacı & Tezel, 2009). According to Keefe and Ferrell (1990), learning style is a combination of cognitive, affective and physiological factors which are relatively stable parameters of how they perceive the learning environment, how they interact with it and how they react to it. In the Experimental Learning Theory (ELT), Kolb and Kolb (2005) state that the learning styles of individuals are determined in a cycle form and where the students take part within this cycle is identified by the Students' Learning Style Inventory (KLSI). In this cycle there are four learning orientations as Concrete Experience (CE), Reflective Observation (RO), Abstract Conceptualization (AC) and Active Experimentation (AE) (Figure 1).

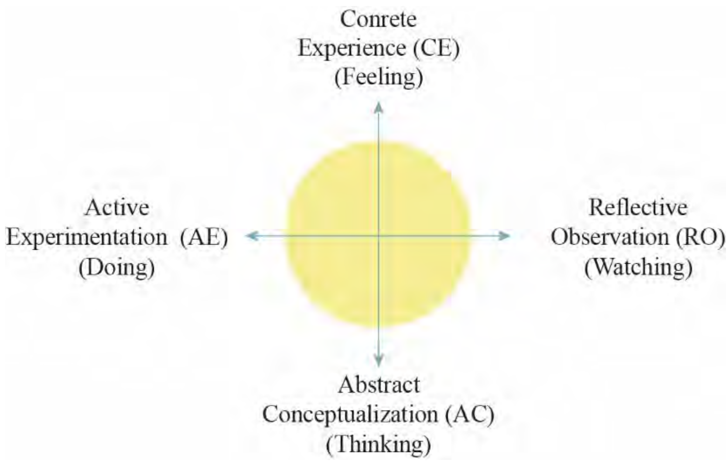


Figure 1. Kolb's Experiential Learning Cycle. (Adapted from Kolb & Kolb, 2005)

The learning style of each individual is a component of concrete experience, reflective observation, abstract conceptualization and active experimentation orientations. Two orientations that individuals possess predominantly determine the learning style of the individual. These learning styles are Diverger, Assimilator, Converger and Accomodator. Kolb's Experiential Learning Model is summarized in Figure 2.

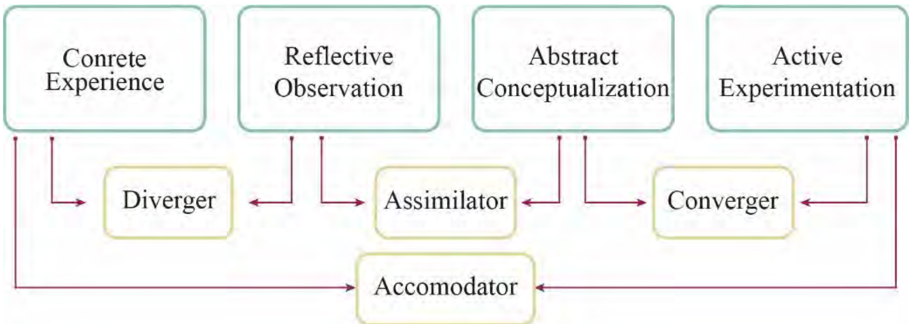


Figure 2. Kolb's Experiential Learning Model's Learning Styles (Adapted from Kolb & Kolb, 2005)

Learning styles of students are important for teachers in order to maximize their learning since students' learning styles are related to their academic achievement (Altun, 2019; Berber, 2016; Kamışlı & Özonur, 2019; Reynolds, Dallaghan, Smith, Walker & Gilliland, 2019; Villajuan, 2019). Learning styles are personal learning styles that facilitate learning, increase motivation and academic achievement, and help the student overcome difficulties in learning easily (Li & He, 2016 as cited in Cevher & Yıldırım, 2020). In addition, it has great importance not only in the individuals' learning and teaching method process, but also in assessment and evaluation processes (Pashler, McDaniel, Rohrer & Bjork, 2008; Widiastuti & Budiyanto, 2018). Accordingly, the learning styles of individuals should also be taken into consideration in the methods and formats used in assessment and evaluation processes because the use of different assessment and evaluation formats in the learning process causes differences in students' knowledge, skills, and attitude scores (Bahar & Hansell 2000; Danili & Reid, 2006; Lawrenz et al., 2001). For example, in the research by Steinberg and Sabella (1997), the first grade university students were given a Force Concept Inventory (FCI) (Hestenes, Wells & Swackhamer, 1992), in the form of multiple-choice and an open-ended test developed by the researchers with a similar quality, to measure the students' achievement in laws of motion. The students got a higher score in the open-ended scale in a statistically meaningful manner. On the other hand, in another study conducted by Lawrenz et al. (2001), multiple-choice, open-ended, laboratory-related tests, and research-oriented tests were used to assess the levels of 3500 science students studying at 13 different schools in national science education norms. As one of the results of this research, they found that the students were significantly more successful in multiple-choice tests than open-ended tests.

As can be seen in the studies within the literature, the utilization of different assessment and evaluation formats in learning process affects students' assessment scores. In the light of this fact, the aim of this study is to analyze in terms of learning styles and gender variables the topics of light and shadow concepts through the student scores obtained in different assessment formats. The fundamental reason to choose the concepts of light and shadow is the fact that most of the studies related with science education are about physics, and among all physics topics, the least examined ones are the optic concepts (Duit, 1993; Önder et al., 2013, Şenkal & Dinçer, 2016, Kaltakçı Gürel et al., 2017). Additionally, although students often observe the occurrences of light and shadow in their daily life, they still have many misconceptions about light and shadow (Haagen-Schützenhöfer & Hopf, 2014). Furthermore, there is also no study regarding the assessment of the concepts of light and shadow via different assessment formats in terms of learning style and gender variables, which are individual differences. In this context, the research question of this study is shaped as: "Do students' achievements in the concepts of light and shadow measured via different assessment formats differ significantly in terms of the learning style and gender variables?".

METHOD

In this section, the design used in the study, the sampling procedure, the data collection tools, and the analysis of the data will be presented.

a) Design of the Study

In this study, correlational survey model was used. Correlational survey model is a quantitative research method used in studies aimed at collecting data to determine certain characteristics of a group (Fraenkel, Wallen & Hyun, 2011). With this method, gender and learning style characteristics of high school students' achievement in the concepts of light and shadow measured by different assessment formats were compared.

b) Sampling Procedure

The population of research consists of 7137 students in 10th grade (age 16) who are studying in 35 different types of high schools (public general high schools, n=15, 43%; public vocational high schools, n=20, 57%) in one of the central districts of Turkey's capital. The sample of the research is 815 students (11% of the population) in six schools (public general high schools, n=2; 33%; public vocational high schools, n=4, 67%) which were chosen from this population by stratified random sampling method. School types were determined as a subgroup. Stratified random sampling is a procedure in which specific subgroups, or strata, are selected for the sample in the same extent as they exist in the population (Fraenkel et al., 2011). Also, it is ensured that the sample represents 10% of the population in accordance with survey research (Gay, Mills & Airasian, 2012).

c) Data Collection Tools

Kolb Learning Styles Inventory

In the research, Kolb Learning Styles Inventory (KLSI) was used as a data collection tool to determine students' learning styles (A. Y. Kolb and Kolb, 2005). KLSI consists of 12 scored items. The four options in each item are scored by the student by giving four points to the most suitable option and one point to the least suitable option. Each option represents a learning style. The respondents receive a score of 12 to 48 in each style. After this step, the combined points of respondents are calculated as:

AC-CE: Abstract Conceptualization-Concrete Experience (1)

AE-RO: Active Experimentation-Reflective Observation (2)

As a result of these calculations, the respondent gets a score between –36 and +36 from steps (1) and (2). The negative result from the AC-CE process shows that the learning is abstract and the positive result shows it is concrete. Similarly, the negative result from AE-RO operation shows that the learning is reflective and the positive result shows that it is active. The results obtained from the combined points are placed on the scale in Figure 3. The intersecting region of a student's every two combined scores indicates the student's learning style according to the inventory (Kolb & Kolb, 2005).

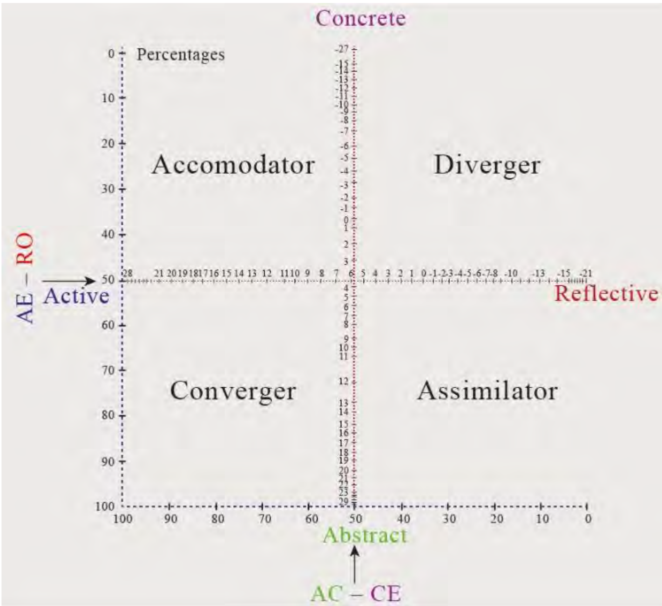


Figure 3. Kolb's Learning Style Inventory scores on AC-CE and AE-RO (Adapted from Kolb, A. Y. & Kolb, 2005)

Light and Shadow Concept Tests

The questions developed by Wosilait, Heron, Shaffer, and McDermott (1998) in order to determine the achievements of students on the concepts of light and shadow were converted into different evaluation formats for the same purpose in the form of Open-Ended Test (OET), Multiple Choice Test (MCT), and Structured Communication Grid Test (SCGT) (see Appendices A, B, and C for Light and Shadow Concept Tests). While preparing the scales in MCT and SCGT formats, the students' possible answers on this topic were chosen (Kanlı, Kartal, Aktaş, & Küçükay, 2015). These tests created in different formats were sent to two field specialists and a physics teacher, and their opinions on the items were obtained. As a result of the received opinions, improvements were made in the item statements and drawings in the tests. After the opinions of the specialists, these tests in different formats were applied as pilot study with 119 students one week apart. After the pilot study, randomly selected students (n=20) were interviewed and their opinions were taken on issues such as comprehensibility, suitability, and readability of the items in the tests. As a result of the pilot study, question statements and drawings were rearranged with the data obtained from the tests and interviews.

Reliability and Validity of Test Scores

The Cronbach-Alpha coefficients that belong to the combined scores in terms of learning styles in KLSI from the process of the pilot study are given in Table 1. The Cronbach-Alpha coefficients that were obtained from the pilot study of KLSI are compatible with the reliability coefficient obtained by Aşkar and Akkoyunlu (1993). This similarity can be given as a proof for the validity of the test. Moreover, since Cronbach's alpha values are higher than .70 is considered acceptable for the reliability (Cortina, 1993).

Table 1. *Cronbach-alpha reliability coefficients obtained from the pilot study regarding Kolb's Learning Styles Inventory*

Scale	Learning Orientations		Cronbach - Alpha
Kolb's Learning Style Inventory	Concrete Experience	(CE)	.73
	Reflective Observation	(RO)	.74
	Abstract Conceptualization	(AC)	.76
	Active Experimentation	(AE)	.81
	Abstract C.- Concrete E.	(AC - CE)	.74
	Active E. - Reflective O.	(AE - RO)	.76

The tests in different formats on the concepts of light and shadow were examined by taking the opinions of two field specialists and a teacher in terms of their face and content validity (whether they are for the same purpose, if the scientific the visuals and expressions of the questions are appropriate etc.). Following the opinions of the specialists, the scales in three different formats (OET, MCT, SCGT) were applied to the student group (n = 119) representing the sample as a pilot study. In order to reduce the recall effect in the application, the situations that may threaten internal validity were controlled by leaving a week between the tests and by informing the students about the test application methods before each application (Fraenkel et al., 2011). After the pilot study, the students (n = 25) were interviewed, and accordingly the content and visuals of the test items were adapted to the student level, and the test items were rearranged for them to be clear and understandable. In light of the preliminary results, the reliability coefficients of OET, MCT and SCGT were calculated using parallel forms method (Crocker & Algina, 1986). The correlation coefficients and significance levels between the test scores are given in Table 2.

Table 2. *The correlation coefficients and significant levels between the open-ended test, multiple choice test and structured communication grid test scores*

Assessment Formats		OET	MCT	SCGT
Open-ended Test (OET)	r	1	.92	.78
	p	-	.00	.00
Multiple Choice Test (MCT)	r	.92	1	.82
	p	.00	-	.00
Structured Communication Grid Test (SCGT)	r	.79	.85	-
	p	.00	.00	1

When Table 2 is examined, the correlation coefficients between the test scores are observed to be .92 between OET and MCT, .78 between OET and SCGT, and .82 between MCT and SCGT, and the correlation coefficients are significant at the level of p = 0.01. These correlation results show that there is a strong and positive relationship between the test mean scores in the pilot study. The positive and strong relationship between the test mean scores and the similarity of the mean scores of two tests can be interpreted that the scores obtained from these tests are reliable (Cohen, 1988).

d) Data Analysis

Outlier analysis was performed by converting students' KLSI and light and shadow concept test scores to standard Z scores, and the scores of 7 students were accepted as extreme values and excluded from the evaluation (Mertler & Vannatta, 2005). Descriptive and inferential statistics were performed to examine whether there is a significant relationship between OET, MCT and SCGT mean scores in terms of the obtained data, and the variables of students' learning style and gender.

Although the sample size is large enough (n=815), the normality assumption was not satisfied. The reason for this could be that schools differ in terms of student achievement and the test is a conceptual test. Since normality assumption was not verified (p>.05), non-parametric tests were used to examine whether there was a significant difference between the mean scores of the students in different formats in terms of learning style and gender, Kruskal-Wallis H test, one of the non-parametric tests, was used to make comparison in terms of learning styles while Mann-Whitney U test was used to examine whether there was a significant difference between the test mean scores of the students in regard of the gender variable.

FINDINGS

In this section, the mean scores of the students about the concepts of light and shadow, obtained through the same question forms in three different assessment formats (OET, MCT and SCGT) were compared.

a) Comparison of Test Scores of Students in terms of Learning Styles

The descriptive statistics of students' achievements measured in different assessment formats in terms of the learning style variable are given in Table 3.

Table 3. *The descriptive statistics related to the OET, MCT and SCGT scores of students in terms of the variables of learning style and gender*

Learning Styles	f	%	OET		MCT		SCGT	
			$\bar{X}$	sd	$\bar{X}$	sd	$\bar{X}$	sd
Convenger	446	54.7	37.2	19.4	46.5	37.4	78.1	15.0
Assimilator	213	26.1	36.5	20.7	43.7	24.2	76.6	13.5
Accomodator	101	12.4	38.6	19.5	40.4	29.2	73.3	15.9
Diverger	55	6.7	29.6	16.7	31.4	31.5	75.5	17.5

When Table 3 is examined, the highest mean of the students' OET scores belongs to the students with the Accommodator learning style, while the lowest mean OET scores belongs to the students with the Diverger learning style. The highest mean of the MCT scores belongs to the students with the Converger learning style, but the lowest mean MCT scores belongs to the students with the Diverger learning style. Whereas the highest mean of the

SCGT scores belongs to the students with the Converger learning style, the lowest mean MCT scores belongs to the students with the Diverger learning style.

The findings of whether the students' achievement scores in the sample show a significant difference according to the learning styles variable or not are given in Table 4.

Table 4. *The Kruskal-Wallis H test results of students' scores in terms of learning styles variable*

Assessment Formats	Learning Style	N	Mean Rank	df	χ^2	p	Significant Difference
Open-ended Test (OET)	Converger	446	415.9	3	9.9	.02*	Converger – Diverger
	Assimilator	213	403.9				Accomodator – Diverger
	Accomodator	101	426.8				
	Diverger	55	325.6				
Multiple Choice Test (MCT)	Converger	446	426.4	3	13.6	.03*	Converger – Diverger
	Assimilator	213	407.2				Assimilator – Diverger
	Accomodator	101	379.3				
	Diverger	55	314.4				
Structured Communication Grid Test (SCGT)	Converger	446	424.3	3	7.9	.05	----
	Assimilator	213	406.7				
	Accomodator	101	360.5				
	Diverger	55	367.9				

*p < .05

The results obtained from the Kruskal-Wallis H test in Table 4 indicate that there are significant differences between the OET and MCT scores of students with different learning styles (p <.05), while there is not a significant difference in SCGT scores. To determine the binary groups from which the differences in OET scores originated, the groups were compared in pairs with Mann-Whitney U test. Accordingly, it was observed that there were significant differences between the student groups in the Converger-Diverger and Accomodator-Diverger styles in favour of students in the Converger and Accomodator styles. In order to determine the binary groups from which the differences in MCT scores originated, the groups in pairs were compared using the Mann-Whitney U test and Bonferroni correction (Miller, 1981). Accordingly, there were significant differences between the Converger-Diverger (p =. 001) and Assimilator - Diverger (p =.005) student groups in favour of students in the Converger and Assimilator style. On the other hand, there is no significant difference in the level of SCGT scores of students with different learning styles (p>. 05).

It follows from the findings that the students' OET an MCT scores might differentiate with respect to their learning styles, but that the students' SCGT scores do not differ with respect to their learning styles.

b) Comparison of Test Scores of Students in terms of Gender

The mean scores of the girls and the boys in the sample in different formats and the standard deviations are given in Table 5 and Figure 4. When Table 5 and Figure 4 are analyzed, it is observed that the highest mean scores of both the girls and the boys are in SCGT and the lowest mean scores are in OETs.

Table 5. *Descriptive statistics of students' scores in terms of the gender variable*

Gender	Formats	N	$\bar{X}$	sd
Girls	Open-ended Test (OET)	397	35.3	19.8
	Multiple Choice Test (MCT)		45.9	29.8
	Structured Communication Grid Test (SCGT)		78.4	14.7
Boys	Open-ended Test (OET)	418	39.0	20.8
	Multiple Choice Test (MCT)		42.4	31.3
	Structured Communication Grid Test (SCGT)		75.7	16.2

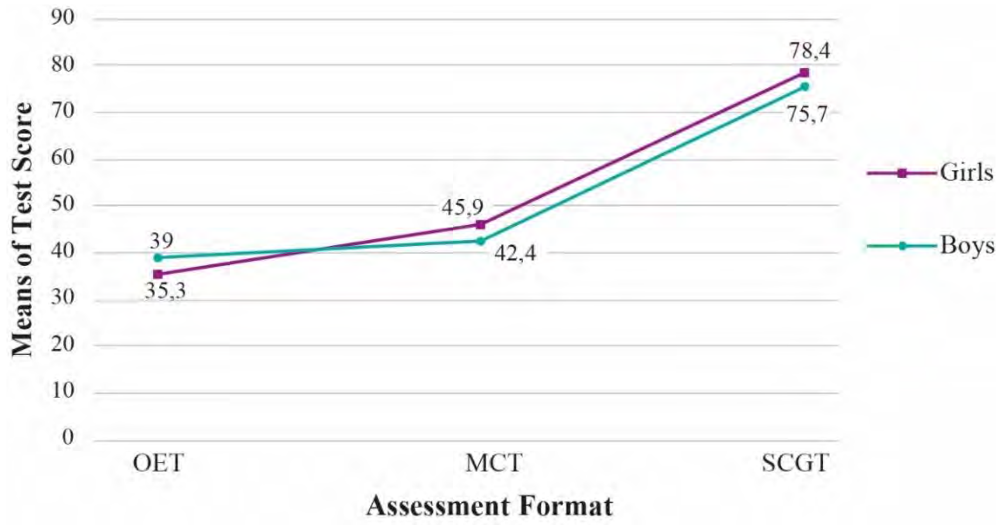


Figure 4. Students' mean scores of different test formats according to gender variable

Whether there is a significant difference between the OET, MCT and SCGT mean scores in terms of the gender variable of the students was examined with the Mann-Whitney U Test, which is a nonparametric test. The results for the Mann-Whitney U Test are given in Table 6.

Table 6. *Mann-Whitney U test results regarding students' OET, MCT and SCGT scores in terms of the gender variable.*

Assessment Format	Gender	N	Mean Ranks	Sum of Ranks	U	p	d
OET	Girls	397	391.2	155308	76305	.027*	.18
	Boys	418	423.9	177212			
MCT	Girls	397	422.6	167772	77177	.076	.11
	Boys	418	394.1	164748			
SCGT	Girls	397	427.5	169733	75216	.021*	.17
	Boys	418	389.4	162787			

*p < .05 and d=Effect size, calculated by dividing the mean difference by the pooled standard deviation of the two groups (Cohen, 1988).

There is a significant difference between the test scores obtained with OET and SCGT of the students in the subgroups of the gender variable (Table 6). This difference is in favour of the boys for OET and, it is significant in favour of the girls for SCGT (p <.05). Although Cohen's effect sizes (d) are below 0.20, they can be considered close enough, so it can be said that there is a small difference in OET and SCGT scores of students as regards the gender variable. The girls and the boys showed the largest performance gap on open-ended items (d= 0.18). No significant difference was observed between the test scores obtained with MCT of the girls and the boys (p> 0.05).

The detailed results of the mean scores of the students with different learning styles and gender taken from different assessment formats can be found in Figure 5 in Appendix D.

DISCUSSION

In the literature, there are many studies investigating the relationship between students' individual traits and their academic achievement. Some of these differences are motivation (Hancock, 2007, Stiggins, 1999), anxiety and learning strategies (Birenbaum, 2007), learning patterns and attitudes (Birenbaum & Feldman 1998), cultural groups (Lawrenz et al., 2001), cognitive styles (Bahar & Hansell, 2000), race and age (Bruschi & Anderson, 1994), assessment format (Kastner & Stangla, 2011; MacGuire & Johnstone, 1987), and language background (Logan & Hazel, 1999).

In the present study, statistically significant differences were found between the learning styles and gender of 10th grade students (age 16), and their achievement in the concepts of light and shadow assessed in different assessment formats. The results of our study are in line with those of the studies in the literature that found different assessment tools cause differences in students' achievement. For instance, Danili and Reid (2006) compared the mean test scores of high school students in Chemistry topics with different scale types (SCGT and MCT) and reported that the students had higher scores in SCGTs than in MCTs. They maintained that this may be due to the fact that MCTs restrict students in transferring their knowledge and give them less opportunity to broaden their opinions compared to SCGTs. Karaçam and Ateş (2010) used OET, MCT, and SCGTs to investigate the conceptual knowledge levels of students in high school in the concepts of force and motion. They found the SCGT mean scores of the students are significantly higher than the OET and MCT scores. The results of our research are compatible with their results. The reason for the high SCGT mean scores can be attributed to the fact that that it contains both correct and wrong answers in the question and allows points to be given to the partial correct answers.

In the present research, it is also observed that the students in the Accommodator style, which is one of the cognitive learning styles, have the highest mean score in the OETs and the lowest mean score in the SCGT, while the students in the Converger style have the highest mean score in the MCTs and SCGTs. The students in Diverger style have the lowest mean score in OETs and MCTs. Our results confirm Matthews (1996), who, based on students' school scores, emphasized that the students in Converger style were more successful in the tests that contain correct answers in the question. Çelik (2010) applied OET, MCT, and short-answer tests to compare the mean test scores of students with different learning styles and it was determined that the test scores of the students in the Converger style were higher than the mean scores of the students in the Diverger style in OETs, and the test scores of the students in the Converger and Assimilator styles were higher than the mean scores of Diverger style students in MCTs. On the other hand, Kablan and Kaya (2013) concluded that the mean scores of the students in the Converger style were significantly higher than the students in the Diverger style in their studies where the results of TIMMS, which uses OET and MCT, were examined. The reason for this situation, which is in line with the MCT and OET results of our study, may be attributed to the fact that the students in Converger style are more successful in explaining abstract concepts and in analytical thinking. In another study for students' cognitive differences, Bahar and Hansell (2000) compared the achievement of high school and university students in word association tests and structured communication grid tests. As a result, they concluded that there was a significant difference in the structured communication grid between the students with field-dependent and field-independent cognitive styles in favour of the students with field-independent cognitive style. They linked these results to the structured communication grid that forces one to choose from complex fields.

Other results of our study are that the OET mean scores were significant in favour of the boys, the SCGT mean scores were in favour of the girls, and no significant difference was observed between the MCT mean scores of boys and girls. Considering the MCT, the results of this study are similar to those of the study by Liu & Wilson (2009), who examined mathematics questions in PISA 2000 and 2003. They did not find any difference in students' performance on traditional multiple-choice items. In addition, while the results of present study are partially compatible with those of Dimitrov (1999), which used OET and MCT, it is not compatible with other studies within the literature (Bolger & Kellaghan, 1990; DeMars, 1998; Matthews 1996). Dimitrov (1999) compared the mean scores of the students in MCT and OET and, as a result, he stated that the mean scores of the boys were higher in both OET and MCT than those of the girls. Bolger and Kellaghan (1990) and DeMars (1998) stated that there was a significant difference in favour of the boys in the mean scores of MCT while the significant difference was in favour of girls in the mean scores of OET. Matthews (1996), on the other hand, explained that the mean scores of the girls was higher than those of the boys. Regarding the reason for these differences, O'Neil and Brown (1998) stated that girls use different cognitive strategies than boys in solving questions. While there is a difference between the OET and SCGT scores of the girls and the boys obtained in the present study, the reason for the lack of difference between the MCT scores may be that the expectations of the tests and the strategies used by girls and boys in answering the questions are different. In addition, students' experiences on the topic can be effective in their achievement. For example, the mean scores of the boys were higher than those of the girls in Bruschi and Anderson (1994) on the subjects of matter and energy, and also those of Engelhardt and Beichner (2004) on the subject of electric circuits, while Johnson (1987) reported that mean scores of the girls on the topics of living things, plants, heat, and temperature were higher. As a reason for these differences, Jovanovic, Solano-Flores, and Shavelson (1994) explained that these differences between boys and girls are not based on the assessment method but on students' past experiences or attitudes.

CONCLUSION

The unchanging truth in learning-teaching process is that learners are different from each other in terms of many characteristics. Additionally, the diversity in education should be reconstructed because it is affected by many variables such as individual differences, education level, course content, and physical conditions. Therefore, actions in educational programs, and relative assessment and evaluation processes should be taken within the perspectives of diversity and flexibility (Harrison & Rainer, 1992). The main result of our study is that different assessment types affect student achievement. In other words, some assessment types are in favour of students while others are not. Especially in countries such as Japan, China, France, Germany, USA, England, and Turkey, where students get into university with a central examination method, these differences have a great importance. For instance, while Japan, China, and USA apply MCT, Germany and France use OET formats (Klein, Kühn, Ackeren, & Block, 2009; Neumann, Trautwein, & Nagy, 2011; Watanabe, 2015). Also, it is also known that central exams cause exam anxiety in students (Davey, De Lian & Higgins, 2007; Karatas, Alci & Aydin, 2013). Therefore, assessment types should be prepared in different formats as much as possible in order for them not to favour any student and to compensate for the possible problems caused by the traditional assessment formats (open ended, multiple choice, etc.) that teachers generally use in their classes (Acar-Erdöl & Yıldızlı, 2018). A good assessment procedure should be planned according to the specific needs of the evaluation (Schuwirth & Van Der Vleuten, 2004). Hence, it is recommended to use different assessment formats as widely as possible in order to assess the students in the most righteous, valid, and reliable way (Race, 2009).

REFERENCES

Acar-Erdöl, T., & Yıldızlı, H. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction, 11(3)*, 587-602.

Altun, H. (2019). Investigation of high school students' geometry course achievement according to their learning styles. *Higher Education Studies, 9(1)*, 1-8.

Anderson, J. (1989). Sex-related differences on objective tests among under-graduates. *Educational Studies in Mathematics, 20(2)*, 165-177.

Aşkar, P., & Akkoyunlu, B. (1993). Kolb öğrenme stili envanteri. *Eğitim ve Bilim, 17(87)*, 37-47

Bahar, M., & Hansell, M. H. (2000). The relationship between some psychological factors and their effect on the performance of grid questions and word association tests. *Educational Psychology, 20(3)*, 349-364.

Bell, J. F. (September, 1997). Sex Differences in Performance in Double Award Science GCSE. Annual Meeting of the British Educational Research Association, York, England.

Berber, N. C. (2016) Eğitim fakültesi fizik eğitimi öğrencilerinin öğrenme biçemleri ile akademik başarılarının çeşitli yönlerden karşılaştırılması. *Manisa Celal Bayar Üniversitesi Eğitim Fakültesi Dergisi, 4(1)*, 15-29.

Birenbaum, M. (2007). Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education, 53(6)*, 749-768.

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research, 40(1)*, 90-98.

Bolger, N. (August, 1984). Gender Difference in Academic Achievement According to Method of Measurement. Annual Convention of the American Psychological Association, Toronto, Ontario.

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27(2)*, 165-174

Bruschi, B. A. & Anderson, B. T. (1994, February). Gender and ethnic differences in science achievement of nine-, thirteen-, and seventeen-year-old students. *Paper presented at the Annual Meeting of the Eastern Educational Research Association*, Florida.

Cohen, J. W. (1988) Statistical power analysis for the behavioral sciences. New Jersey: Erlbaum

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology. 78*, 98-104

Çelik, T. (2010). İlköğretim öğrencilerinin bilişsel stil ve öğrenme stillerinin farklı ölçme formatlarından aldıkları puanlara etkisi. (Unpublished master's thesis) Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü, Bolu.

Danili, E., & Reid, N. (2006). Cognitive factors that can potentially affect pupils' test performance. *Chemistry Education Research and Practice, 7(2)*, 64-83.

Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education, 31(4)*, 385-396.

Cevher, A. Y. & Yıldırım, S. (2020) Öğrenme Stilleri Konusunda Yapılmış Akademik Çalışmaların İncelenmesi: Sistematik Derleme. HAYEF: *Journal of Education, 17(1)*, 20-50

Chen, W. C., & Whitehead, R. (2009). Understanding physics in relation to working memory. *Research in Science & Technological Education, 27(2)*, 151-160.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review, 20(4)*, 631-642.

DeMars C. E. (1998) Gender differences in mathematics and science on a high school proficiency exam: the role of response format. *Applied Measurement in Education, 11(3)*, 279-299.

Dimitrov, D. M. (1999). Gender differences in science achievement: differential effect of ability, response format, and strands of learning outcomes. *School Science and Mathematics, 99(8)*, 445-450.

Dunn, R. S., & Dunn, K. J. (1993). Teaching secondary students through their individual learning styles: Practical approaches for grades 7-12. Prentice Hall.

Duit, R. (1993, August). *Research on students' conceptions–developments and trends*. Paper presented at The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Ithaca, NY

Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics, 72(1)*, 98-115.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). How to design and evaluate research in education. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Gay, L. R., Mills, G. E & Airasian, P. W. (2012). *In Educational research: Competencies analysis and applications*. America: Library of Congress Cataloging-in-Publication Data.

Haagen-Schützenhöfer, C., & Hopf, M. (2014). Development of a two-tier test-instrument for geometrical optics. In Constantinou, C.; Papadouris, N.; Hadjigeorgiou, A.(Hg.), E-Book *Proceedings of the ESERA 2013 Conference*: Science Education Research for Evidence-based Teaching and Coherence in Learning (pp. 24-30).

Hancock, A. M. (2007). Intersectionality as a normative and empirical paradigm. *Politics & Gender, 3(2)*, 248-254.

Harnisch, D. L. (1994). Performance assessment in review: New directions for assessing student understanding. *International Journal of Educational Research, 21(3)*, 341-350.

Harrison, A. W., & Rainer Jr, R. K. (1992). The influence of individual differences on skill in end-user computing. *Journal of Management Information Systems, 9(1)*, 93-111.

Herman, J. L., Aschbacher, P. R. & Winters L. (1992). A practical guide to alternative assessment. Association for Supervision and Curriculum Development: Alexandria

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30(3)*, 141-158.

Johnson, S. (1987). Gender differences in science: Parallels in interest, experience and performance. *International Journal of Science Education, 9(4)*, 467-481.

Jovanovic, J., Solano-Flores, G., & Shavelson, R. J. (1994). Performance-based assessments: Will gender differences in science achievement be eliminated?. *Education and Urban Society, 26(4)*, 352-366.

Kablan, Z., & Kaya, S. (2013). Science Achievement in TIMSS Cognitive Domains Based on Learning Styles. *Eurasian Journal of Educational Research, 53*, 97-114.

Kaltakçı Gürel, D., Ölmeztürk, A., Durmaz, B., Abul, E., Özün, H., Irak, M., ... & Baydar, Z. (2017). 1990-2016 yılları arasında Türkiye'de fizik eğitimi alanında yayınlanmış tezlerin içerik analizi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 37(3),* 1141-1172.

Kamışlı, H., & Özonur, M. (2019). Students' learning styles in vocational education. *International Journal of Curriculum and Instruction, 11(1)*, 209-220.

Kanlı, U., Kartal, Y. E., Aktaş, G. ve Küçükay, S. (2013) Işık ve gölge kavramları hakkında fen ve fizik öğretmen adaylarının kavramsal anlamaları üzerine kesitsel bir araştırma. *I. Ulusal Fizik Eğitimi Kongresi*, 12-14 Eylül 2013 Hacettepe, Türkiye.

Karaçam, S., & Ateş, S. (2010). Ölçme tekniğinin farklı bilişsel stillerdeki öğrencilerin hareket konusundaki kavramsal bilgi düzeylerine etkisi. A*bant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 10 (1),* 21-30.

Karatas, H., Alci, B., & Aydin, H. (2013). Correlation among high school senior students' test anxiety, academic performance and points of university entrance exam. *Educational Research and Reviews, 8(13)*, 919-926.

Kastner, M., & Stangla, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter?. *Procedia-Social and Behavioral Sciences, 12*, 263-273.

Kaya, F., Özarabacı, N. & Tezel, Ö. (2009). Investigating Primary School Second Grade Students' Learning Styles According to the Kolb Learning Style Model in terms of Demographic Variables. *Journal of Turkish Science Education. 6, (1)*, 11-25

Keefe, J. W., & Ferrell, B. G. (1990). Developing a defensible learning style paradigm. *Educational Leadership, 48(2)*, 57-61.

Klein, E. D., Kühn, S. M., Ackeren, I. V., & Block, R. (2009). Wie zentral sind zentrale Prüfungen? Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik, 55(4)*, 596-621.

Kolb, A. Y. & Kolb, D. A. (2005). *The Kolb learning style inventory-version 3.1 2005 technical specifications*. Boston, MA: Hay Resource Direct.

Kulm, G., & Malcom, S. M. (1991). Science Assessment in the Service of Reform. (Report No. ISBN-0-87168-426-5) Washington: AAAS

Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education, 85(3)*, 279-290.

Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education, 22(2)*, 164-184.

Logan, P., & Hazel, E. (1999). Language background and assessment in the physical sciences. *Assessment & Evaluation in Higher Education, 24(1)*, 53-65.

MacGuire, P. R. P., & Johnstone, A. H. (1987). Techniques for investigating the understanding of concepts in science. *International Journal of Science Education, 9(5)*, 565-577.

Matthews, D. B. (1996). An investigation of learning styles and perceived academic achievement for high school students. *The Clearing House, 69(4)*, 249-254.

Mertler, C. A., & Vannatta, R. A. (2005). Advanced and multivariate statistical procedures. California, USA: Taylor & Francis

Miller, R. G., Jr. (1981). Simultaneous statistical inference. New York: McGraw-Hill.

Murphy, P. (1991). *Gender differences in pupils' reactions to practical work*. In B. Woolnough (Ed.), Practical science. Milton Keynes: Open University Press.

Neumann, M., Trautwein, U., & Nagy, G. (2011). Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the University entrance qualification in Germany. *Studies in Educational Evaluation, 37(4)*, 206-217.

O'Neil Jr, H. F., & Brown, R. S. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied measurement in Education, 11(4),* 331-351.

Önder N. Ç, Özlem, O. Eraslan, F., Gülçiçek, Ç., Göksu, V., Kanlı, U., Eryılmaz, A., Güneş, B. (2013). Content Analysis of Physics Education Studies Published in Turkish Science Education Journal from 2004 to 2011. *Journal of Turkish Science Education, 10(4)*, 151-163.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. Psychological Science in The Public Interest, 9(3), 105-119.

Race, P. (2009). *Designing assessment to improve physical sciences learning*. Hull: Higher Education Academy.

Reynolds, Q. J., Dallaghan, G. L. B., Smith, K., Walker, J. A., & Gilliland, K. O. (2019). Comparison of medical student learning styles and exam performance in an integrated curriculum. *Medical Science Educator, 29(3)*, 619-623.

Resnick, L. B., & Resnick, D. P. (1992). *Assessing the thinking curriculum: New tools for educational reform*. In B. R. Gifford & M. C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 37–75). Boston: Kluwer.

Schuwirth, L. W., & Van Der Vleuten, C. P. (2004). Different written assessment methods: what can be said about their strengths and weaknesses?. *Medical education, 38(9)*, 974-979.

Smith, M., Breakstone, J., & Wineburg, S. (2019). History assessments of thinking: A validity study. *Cognition and Instruction, 37(1)*, 118-144.

Steinberg, R. N., & Sabella, M. S. (1997). Performance on multiple-choice diagnostics and complementary exam problems. *The Physics Teacher, 35(3)*, 150-155.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan, 81(3)*, 191-198.

Şenkal, O., & Dinçer, S. (2016). Türkiye'de fizik eğitimi - öğretimi ile ilgili yapılan çalışmaların eğilimi. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 25(2)*, 57-70

Watanabe, M. E. (2015). Typology of abilities tested in university entrance examinations: comparisons of the United States, Japan, Iran and France. *Comparative Sociology, 14(1)*, 79-101.

Widiastuti, I. & Budiyanto, C. W. (2018). Applying an experiential learning cycle with the aid of finite element analysis in engineering education. *Journal of Turkish Science Education, 15,* Special Issue, 97-103. Doi: 10.12973/tused.10261a

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 703–713.

Villajuan, A. M. L. (2019). Relationship between learning styles & academic achievement in mathematics of grade 8 students. *International Journal of English Literature and Social Sciences (IJELS), 4(4),* 1052-1055.

Wosilait, K., Heron, P. R., Shaffer, P. S., & McDermott, L. C. (1998). Development and assessment of a research-based tutorial on light and shadow. *American Journal of Physics, 66(10)*, 906-913.

Yip, D. Y., Chiu, M. M., & Ho, E. S. C. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science and Mathematics Education, 2(1),* 91-106.

APPENDICES

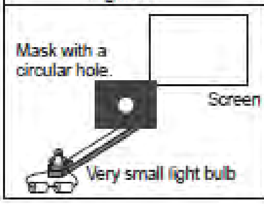
Appendix A

Name :
Gender : ☐G ☐B

IN ALL PARTS OF THIS PRETEST, ASSUME THAT THE ROOM IS VERY DARK BEFORE ANY BULBS ARE TURNED ON

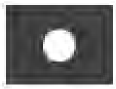
1. A very small bulb is held in front of a screen. A mask with a triangular hole is placed between the bulb and a screen as shown at Figure 1.
Sketch what you would see on the screen when the bulb is lighted.

Figure 1



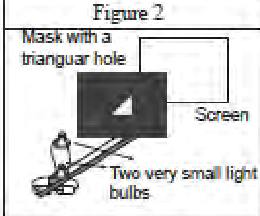
Sketch here

ANSWER



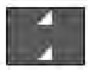
2. Two very small bulbs are held in front of a screen. A mask with a circular hole is placed between the bulbs and a screen as shown at Figure 2.
Sketch what you would see on the screen when the bulb is lighted.

Figure 2



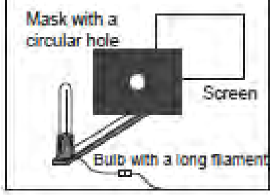
Sketch here

ANSWER



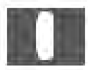
3. A bulb with long filament is held in front of a screen. A mask with a circular hole is placed between the bulb and a screen as shown at Figure 3.
Sketch what you would see on the screen when the bulb is lighted.

Figure 3



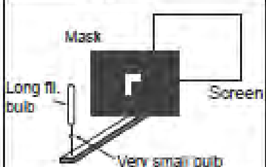
Sketch Here

ANSWER



4. A very small bulb and a bulb with long filament are held in front of a screen. A mask is placed between the bulbs and a screen as shown at Figure 4.
Sketch what you would see on the screen when the bulb is lighted.

Figure 4



Sketch Here

ANSWER

Appendix B

Name :
Gender : ☐G ☐B

IN ALL PARTS OF THIS PRETEST, ASSUME THAT THE ROOM IS VERY DARK BEFORE ANY BULBS ARE TURNED ON

1. A very small bulb is held in front of a screen. A mask with a triangular hole is placed between the bulb and a screen as shown at Figure 1.
Which of the below seen on the screen when the bulb is lighted?

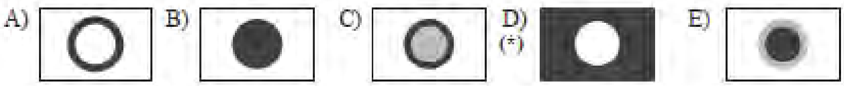
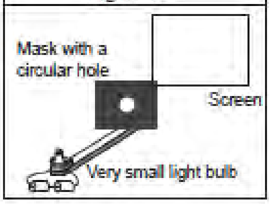


Figure 1



2. Two very small bulbs are held in front of a screen. A mask with a circular hole is placed between the bulbs and a screen as shown at Figure 2.
Which of the below seen on the screen when the bulbs are lighted?

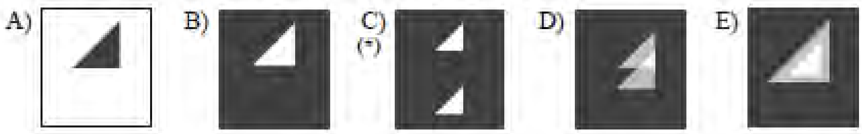
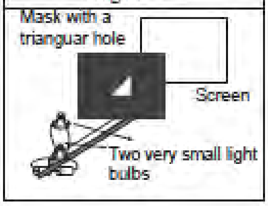


Figure 2



3. A bulb with long filament is held in front of a screen. A mask with a circular hole is placed between the bulb and a screen as shown at Figure 3.
Which of the below seen on the screen when the bulb is lighted?

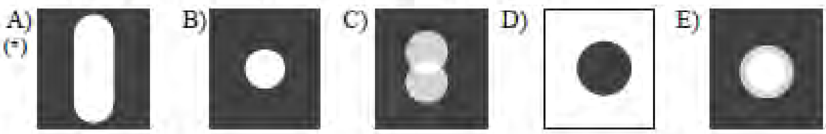
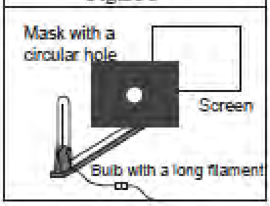


Figure 3



4. A very small bulb and a bulb with long filament are held in front of a screen. A mask is placed between the bulbs and a screen as shown at Figure 4.
Which of the below seen on the screen when the bulbs are lighted?

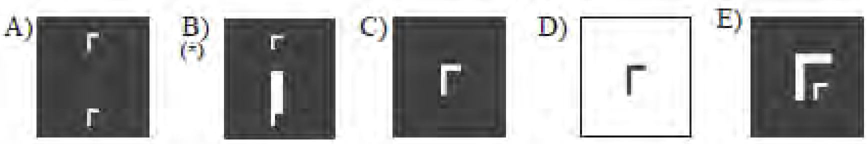
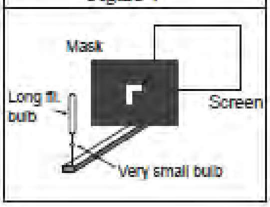


Figure 4

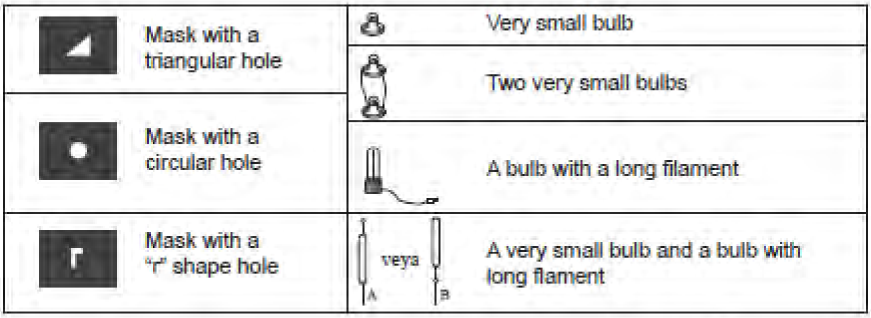


Appendix C

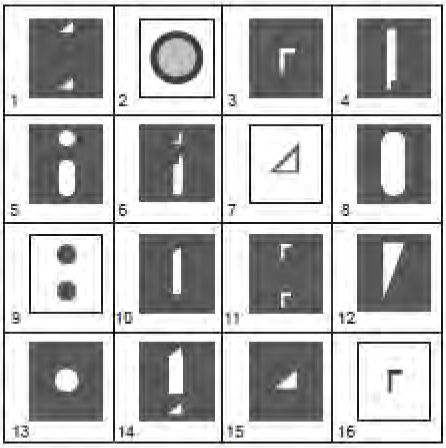
Name :
Gender : ☐G ☐B

IN ALL PARTS OF THIS PRETEST, ASSUME THAT THE ROOM IS VERY DARK BEFORE ANY BULBS ARE TURNED ON

Figure 1 (Masks) Figure 2 (Bulb System)



A mask (in Figure 1) with hole is placed between screen and a bulb system (Figure 2).



1. Which of the numbered images in the table are created using only one very small bulb?

ANSWER:

3, 13,15

2. Which of the numbered images in the table are created using two very small bulbs?

ANSWER:

1, 11

3. Which of the numbered images in the table are created using a bulb with a long filament?

ANSWER:

4, 8, 10

4. Which of the numbered images in the table are created using a bulb with a long filament and a very smal bulb?

ANSWER:

5, 14

Appendix D

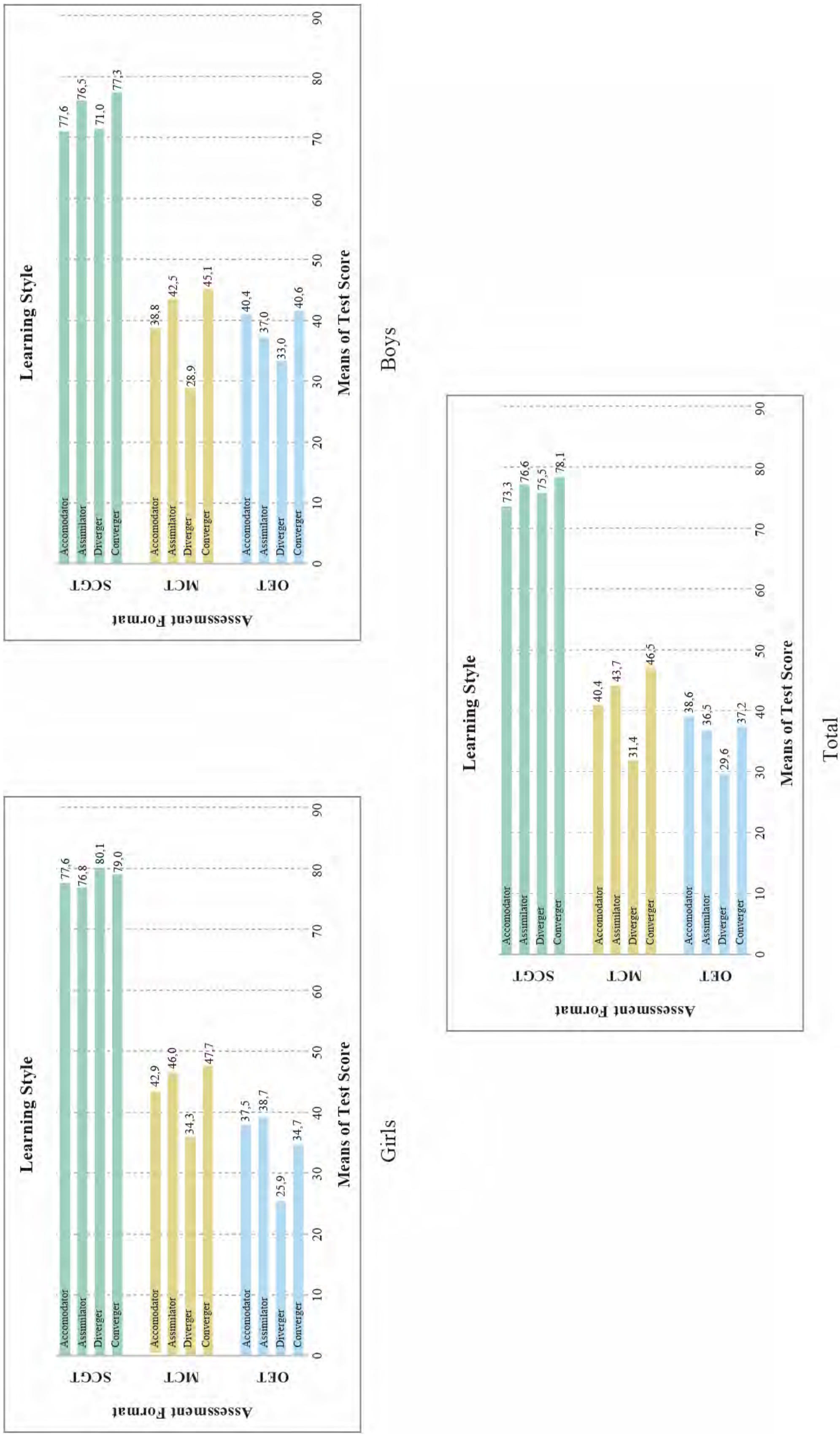


Figure 5. The mean scores of the students in different assessment formats according to the learning style and gender variables