Determining the Qualifications of the Secondary School Students in the Field of Quadrilaterals: A Scale Development Study

Serhan Ulusan [i]
Aydın Adnan Menderes University

Duygu Dinamit [ii]
Turkey Ministry of National Education

Gökhan Aksu [iii]
Aydın Adnan Menderes University

Abstract

Testing-assessment is needed to determine whether the education process is successful or at what level. For this reason, it is of great importance to determine the achievement level of the students at the secondary level. The purpose of this study is to develop an achievement test consisting of multiple choice questions about 'quadrilateral' subject of secondary school students by using alternative assessment techniques. The study group of this research, which is a descriptive survey type, consists of 372 secondary school students who are in the seventh grade in Aydın and Muğla during the 2018-2019 academic year. In the study, ITEMAN, SPSS and JMETRIK programs were used based on Classical Test Theory and Material Testing Theory. The formed item pool was subjected to the appearance validity by five trainers, who are experts in that field. According to the experts' opinions, it was decided that for the achievement test, 28 items were to be included in the test primarily. It was decided as a result of analysis of the data obtained from the students that an item with very low item-specificity index, four items that do not provide model data compliance in Rasch analysis and two items identified to have item bias may be excluded from the test were excluded from the test. After these processes, the final testing tool consisted of 21 items and the average difficulty index of the test was .64; the discrimination index was calculated as .40. The Cronbach Alpha internal consistency coefficient of the final state was determined as .81. According to these results, it can be said that the achievement test has sufficient validity and reliability.

Keywords: Quadrilateral, Hierarchy, Success, Test development, Testing , Rasch

----
[i] **Serhan Ulusan,** Assist. Prof. Dr., Mathematics and Science Education, Aydın Adnan Menderes University, Faculty of Education, ORCID: 0000-0003-0116-2696

[ii] **Duygu Dinamit,** Kumla Secondary School, Turkey Ministry of National Education, ORCID: 0000-0003-3760-6391

[iii] **Gökhan Aksu,** Instructor Dr., Vocational High School, Adnan Menderes Üniversitesi, ORCID: 0000-0003-2563-6112

Correspondence: gokhanaksu1923@gmail.com

INTRODUCITON

Geometry, a branch of mathematics, is a field of study that enables individuals to establish reasoning, problem solving, critical thinking and cause-effect relationship and to develop a high level of thinking skills as well as including many shapes and objects of knowledge, which also facilitates the understanding of the earth, direction and figure of the living world. Geometry has been included in the curriculum and educational programs since primary education because of the fact that it contributes to the students' critical thinking and problem solving skills, assists in teaching other subjects of mathematics, is an important part of mathematics used in daily life, is used in science and art, as well as helps students to better understand the world in which they live (Baykul, 2002).

Geometry teaching, as many researchers (Baykul, 1999; Duatepe, 2000; Fujita and Jones, 2007) emphasized, has been important not only to let the learners to comprehend the knowledge and relationships related to point, line, plane, planar shapes, space and spatial shapes, but also in terms of the development of spatial thinking and visual skills. Besides, it is stated that understanding the classification and properties of geometric shapes in geometry teaching contributes to the solution of problems related to real life and other fields of mathematics (testing, algebra and rational numbers) (NCTM, 2000). In this contex, in the mathematics curriculum of primary schools in Turkey, while recognition, naming, constructing, drawing, comparing and grouping activities of geometric objects and shapes are prioritized according to certain characteristics (MEB, 2013a); in the middle school mathematics curriculum, these figures are characterized by a small number of characteristic features and classifications (rectangles are rectangular parallellines etc.) (MEB, 2013b). Thus, while students are expected to recognize and understand the geometric shapes and properties in elementary mathematics curriculum, in secondary education mathematics curriculum, it is expected that they will form the relationships between these figures and classify the figures according to certain characteristics. However, students experience problems in understanding and classifying geometric concepts.

The memorization of the features of the forms, the inadequate sample presentation (eg, giving the typical image only) cause them to create limited structures related to geometric concepts and therefore not to understand the concept. On the other hand, the hierarchical classification of quadrilaterals is seen as a field of study to promote the development of geometric thinking (Fujita and Jones, 2007). For example, the parallelogram is defined as a quadrilateral with opposite sides of the parallel. Since the opposite sides of the rhombus, square and rectangle are parallel to each other, these rectangles are parallel edges. Therefore, if a property is correct for the parallelogram, it is also true for rhombus, square, and rectangle. However, the results of many studies have shown that students experience some problems in the hierarchical order of the quadrilaterals (Monaghan, 2000; Toluk, Olkun and Durmus, 2002; Olkun and Aydogdu, 2003; Aktas, 2005; Erez and Yerushalmy, 2006; Pickreign, 2007; , 2007; Fujita and Jones, 2007; Akuysal, 2007; Ergün, 2010; Aktaş and Aktaş, 2011; Türnüklü, Alaylı and Akkaş, 2013). In a study conducted by Olkun and Aydoğdu (2003), it was determined that students see geometric shapes only separately and independently from each other. Similarly, Okazaki and Fujita (2007) found that many of the students had difficulties in perceiving the square as a special state of rectangular and rhombus, but were more successful in perceiving the rhombus as parallelogram. Again, it is stated that most students experience difficulties in this issue with thoughts such as the square is not a parallelogram because the parallelogram appears oblique (Erez and Yerushalmy, 2006; Okazaki and Fujita, 2007). Fujita and Jones (2007) state that students' opinions about whether a rhombus is a special parallelogram, for example, are simply not enough to control their images, and that the features of the figure should be mutually correlated. In fact, in order to establish these relations, students should decide not by taking the shape of the given quadrilateral into account, but by using the properties of this quadrilateral (edge, angle, etc.). This process is considered to be a useful activity in the development of geometric thought (Fujita, 2008; Fujita, 2012).

Defining concepts in geometry is important for teaching. Although it is not approved by the researchers (De Villers, 1998; Tall and Vinner, 1981) that the definitions of the concepts are directly approved, according to Türnüklü, Alaylı and Akkaş (2013), it is also accepted that the definitions of concepts have important roles in the formation of concept image and problem solving situations.

According to Tall and Vinner (1981), students are able to filter out these definitions in their minds either the definitions of concepts are taught through direct narration, or the students are enabled to structure these definitions. These personally structured definitions may differ from formal definitions and these personal concept definitions may cause individuals to create their own concept images. According to Türnüklü, Alaylı and Akkaş kavrama (2013), the visual image contained by each geometric concept can be more prominent than the concept. In this context, typical (prototype) samples are the key factor. Each concept can have multiple prototypes. These prototypes are examples of some of the features included in the long feature list of the concept. These prototype shapes always have an effect on concept image (Fischbein, 1993; Hershkowitz, 1990). Fujita (2012), as a result of many researches, the definition of geometric shape and the family relationship in this figure has revealed that the properties can often lead to a contradiction. This contradiction leads to false perceptions and generalizations through conceptual perception from prototype form (Fujita, 2012; Fujita and Jones, 2006; Hershkowitz, 1990). For example, the definition of the parallelogram and the shape of the prototype are contradictory to the rectangle from the same family. It leads to false generalizations. In the case in this example, the perception that the parallelogram cannot be perpendicular develops (Türnüklü, Alaylı and Akkaş, 2013).

As it can be understood from the literature, the knowledge that students learn in geometry is important for students to acquire reasoning, problem solving and critical thinking skills while they are creating solutions for many problems they will encounter in daily life. In this study, it was aimed to develop a quadratic hierarchy test.

METHOD

This research can be accepted as a basic research because it is to develop a scale to reveal the knowledge of the 7th grade students on quadrilaterals and their ability to make inferences based on this information. Before the decision to develop an achievement test for this study, existing tests were examined. However, in the existing tests, even though there are questions that address the hierarchy of quadrilateral under the title of Quadrangles and Quadrilateral, it was determined that there were only certain special quadrilaterals' relationship such as square with rectangular or square with rhombus. In this study, since it is aimed to determine to what extent the students can see the relationships between all quadrilaterals, the test includes the items in which all quadrilaterals are related to each other.

Process

The test development process consists of preparation, implementation and reporting stages, and various steps are followed at this stages when the test development studies in educational researches are examined (Caliskan & Kaptan, 2009). In fact, in this research, the steps of writing test items, piloting, validity, reliability and item analysis, which were also used by Burns et al. (1985) and Karslı and Ayas (2013) for test development studies, were followed and applied respectively. In this study, the following process was followed while developing the test.

1. *Determination of the purpose of the test:* In the literature search, no test development studies have been observed on the hierarchy of quadrilaterals. Therefore, in order to determine the knowledge of the students about the definitions of the quadrilaterals, to determine the special cases of the quadrilaterals and to classify the quadrilateral hierarchical, developing an achievement test was aimed.

2. *Determining the subject:* While the studies were examined, although there are studies conducted with the teacher candidates or students about the quadrilaterals, since there was no test development study for the hierarchy of quadrilaterals, the subject of the test was determined as rectangles.

3. *Determination of the properties that can be measured by the test:* A table of statements containing the related learning outcomes and sub-learning outcomes were prepared and the items appropriate to the steps of Bloom Taxonomy were prepared.

4. *Writing the items of the test:* In order to determine how many items according to the determined learning outcomes and sub-learning outcomes will be written, the opinions of the mathematics teachers who are working in the primary school and field experts were taken.

5. As a result of the expert opinions, the 33-item test consisting of 9 items from the recall step, 17 from the comprehension step and 7 from the application, was formed.

6. *Spelling and obtaining expert opinion:* For the purpose of assessing the correctness and scientific accuracy of the prepared articles, the test was submitted to the opinion of the five trainers. In line with expert opinions, necessary corrections were carried out for deficiencies, errors and weakening of the scope validity, and 5 items have been removed by the experts and the total number of items has been decreased to 28.

7. *Implementation of the test:* The test was administered to 372 students from a private school and four state schools in Aydın and Muğla, determined by random sampling method.

8. *Item analysis:* The distinction and difficulty levels of items constituting the test were calculated using the following formulas through the scores of the upper group of students who gave the most accurate answers to the test and through the scores of the lower group formed from the students who gave the least correct answers.

$$D = \frac{D_{Ü} - D_{A}}{N_{Ü} \text{ ya da } N_{A}} \qquad P = \frac{D_{Ü} - D_{A}}{N_{Ü} + N_{A}}$$

D: Item discrimination index

P: Item difficulty index

$D_{Ü}$: Number of students correctly answering the item in the upper group

D_{A}: Number of students correctly answering the item in lower group

$N_{Ü}$: Number of students in the upper group

N_{A}: Number of students in the lower group

Study Group

Within the scope of the aim of the researcher, it was aimed to reach 10 times more than the total number of items in the item pool within the provinces where the researchers are in the 2017-2018 academic year. For this reason, 372 students who participated in the study voluntarily from the randomly selected schools in the mentioned provinces formed the study group of the research. The schools and distribution of students in the study group are given in Table 1.

Table 1. Distribution of research sample

Schools / Grade Level	Atatürk Secondary School	Efeler Secondary School	Dr. Fevzi Mürüvet Uğuroğlu Secondary School	Menteşe Secondary School	Private School	TOTAL
Grade 7 (f)	43	83	135	31	80	372
Grade 7 (%)	11,6	22,3	36,3	8,3	21,5	100

Data Collection Tools

In the process of developing data collection tools, firstly a literature search on the subject of research has been done and studies which include students' knowledge about quadrilaterals are examined. Among the existing studies on quadrilaterals, it was observed that there is no test development study on the hierarchy of rectangles. Therefore, developing an achievement test was intended in order to determine the knowledge of students of 7th grade on rectangle definitions, to determine the special cases of the quadrilaterals and to classify the quadrilateral hierarchical information.

During the process of preparing questions in the Quadrilateral Test, a related learning outcome in the area of Geometry and Measurement sub-learning area in the 7th Grade in the Curriculum of Mathematics Course in the Ministry of National Education was taken into consideration. It was thought that this learning outcome involves more than one behavior, and sub-gains were determined in order to measure a single behavior for some questions. For the main learning outcomes or sub-outcomes, it was tried to determine how many questions will be written for the recall, understanding, practice, analysis, evaluation and synthesis steps, which consists the cognitive step of Bloom Taxonomy.

Taking into account the learning outcomes and sub-learning outcomes, the number of questions to be written at the level of remembering, comprehension and application appropriate to 7th grade level was decided and 33 questions were written and a table was prepared. In the first instance of the test, there are 9 questions in the recall step, 17 in the comprehension step and 7 in the application step. In the test development studies in the literature, it is stated that the question root of the questions, the distractors in the question, the questions' coverage of the learning outcomes, the behavior measured in the question and the adaptation of the behavior to be measured in the question should be consulted to the field experts (Webb, 1997). The test was submitted to the opinion of five educators in order to evaluate the correctness and scientific accuracy of the prepared items. In line with the expert opinions, the number of questions was reduced to 28 by making necessary corrections for deficiencies, errors and conditions that weaken the content validity. Following the completion of the necessary arrangements, the test was applied to 372 students who had completed their learning outcomes in quadrilaterals.

Data Analysis

After the application of the test within the scope of reliability analysis, the data were analyzed according to two different measurement theories. Analysis based on the Classical Test Theory was evaluated using Iteman and SPSS program; reliability analysis in terms of internal consistency was evaluated according to the results obtained from Cronbach Alpha and Dot Dual Series Correlation Analysis. The analysis carried out within the context of Item Response Theory were carried out with the help of Rasch analysis which is a special case of a parameter logistics model with jMetric program. According to the criteria proposed by Linacre (2002), the items to be included in the test and the items to be removed were determined with the help of unweighted and standardized compliance statistics. Within the scope of validity analysis, expert opinions were asked and they were asked to express their opinions about whether the scale developed from experts according to the table of specifications has got content validity. In this study, the scattering graph obtained from the Iteman program shows to which outcome each item belongs on the figure. In the evaluation of the results obtained from the Rasch analysis, the criteria determined for the unweighted and standardized compliance statistics defined by Linacre (2002) will be used. Acceptable values for the criteria to be used in the study are shown in Table 2.

Table 2. Acceptable Ranges of Compliance Statistics

Criteria	Interval	Comment
Incompatibility Criteria (WMS and IAS)	> 2.0	The item deflects the property of measurement and reduces the quality of the measurement.
	1.5 – 2.0	The item is insufficient, but not too pessimistic for the measurement process.
	0.5 – 1.5	The item very suitable for measuring. *
	< .50	The item is not enough for the measurement process but not too bad.
Compliance Criteria (Std.WMS and Std.UMS)	≥ 3	The data do not correspond to the model.
	2.0 – 2.9	Data cannot be predicted significantly.
	-1.9 – 1.9	The data are reasonably predictable.*
	≤ -2.0	The data can be estimated at very low levels.

• Shows the ideal ranges.

Table 2 shows that the ideal range for WMS and UMS values is between 0.50 and 1.50. Similarly, the ideal range for standardized WMS and UMS values is between -1,90 and 1,90. The columns WMS and Std.WMS show the weighted and standardized squared mean mismatch statistics, respectively. UMS and Std.UMS columns show the weighted and standardized squared mean non-criterion statistics.

FINDINGS

Data from 372 students were first analyzed with Iteman program. The scatter plot of item discrimination indexes and item difficulty values for each of the 28 items in the test is shown in Figure 1.
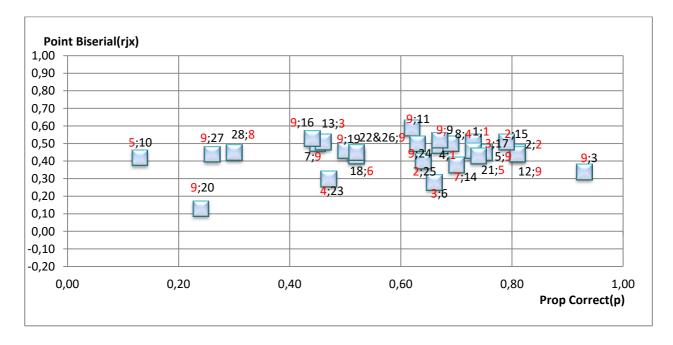

Figure 1. Scattering Graph of the Items in the Test

In Figure 1, the relationship between the points of the graphical items and the total score obtained from the whole of the test was formed according to the item discrimination index. While the item number is shown on itself, it was stated in the red numbering that the items belong to which learning outcomes or sub-learning outcomes. Whether these graphical items can be used shows which items will have the same or similar difficulty and discrimination while they measure the same outcome. This graph is used to determine the items in the test. When the graph is examined, it can be seen that the most difficult questions are item 10, item 20 and item 27 respectively; the easiest questions are item 2, item 12 and item 3. Since the test is intended to be composed of items with different difficulty and distinctiveness indices, no item was removed at this stage. The difficulty and discrimination values of the items in the 28-item test are shown in Table 3.

Table 3. Item Analysis Results

Question Number	Item Difficulty Index(p)	Item Discriminatin Index (r)	Question Number	Item Difficulty Index(p)	Item Discriminatin Index (r)
1	.74	.50	15	.79	.51
2	.82	.44	16	.47	.51
3	.94	.34	17	.73	.50
4	.69	.51	18	.52	.44
5	.74	.46	19	.54	.46
6	.67	.29	20	.26	.12
7	.47	.51	21	.75	.42
8	.72	.49	22	.50	.47
9	.71	.52	23	.48	.29
10	.15	.44	24	.64	.51
11	.63	.61	25	.62	.43
12	.81	.46	26	.51	.48
13	.48	.48	27	.28	.45
14	.70	.41	28	.34	.47

When Table 3 is examined, it is seen that the difficulty values of the items in the test vary between .15 and .94 and accordingly, there are questions at all levels in the test. When the item discrimination index is examined, it is determined that the discriminant value is below the critical value (.20) only for item 20. It can be seen that the average difficulty index of the item analysis results was 0.60; and the discrimination index is calculated as 0.45. The KR21 coefficient determined as the reliability value of the test was found to be 0.82. In addition, the lowest score obtained from the whole test was 4.00 and the highest score was 28.00; it was also found that the coefficient of skewness for the whole test was -0.03 and the coefficient of curtosis was -0.86. According to the skewness and kurtosis coefficients, the scores obtained from the test showed normal distribution (Baykul and Güzeller, 2014). Besides, it was determined that the mean score of the test was 16.72 and its standard deviation was 5.68. Item statistics for the whole test are shown in Table 4. According to the skewness and kurtosis coefficients, the scores obtained from the test showed normal distribution (Baykul and Güzeller, 2014). Besides, it was determined that the mean score of the test was 16.72 and its standard deviation was 5.68. Material statistics for the whole test are shown in Table 4.

Table 4. Item Statistics for the Complete Test

Item Statistics	Value
Number of Articles	28
Minimum	4,00
Maximum	28,00
Average	16,72
Median	17,00
Standard deviation	5,68
Quarter opening	9,00
Skewness	-0,03
Kurtosis	-0,86
KR21	0,802
Gutman	0,855
Beta value of Raju	0,851

When Table 4 is examined, it is determined that the reliability coefficient obtained according to three different methods of 28-item test varies between .80 and .85 and therefore the results obtained from the test are reliable (Cronbach, 2004). In order to determine the degree to which the results obtained according to the classical test theory are consistent with the item response theory, the item statistics obtained with the jMetric program are shown in Table 5.

Table 5. MTK Based Item Statistics to the Items in the Scale

Item Number	Item Difficulty	Std. Deviation	Distinction index
Item1	0,7446	0,4367	0,4342
Item 2	0,8172	0,3870	0,3836
Item 3	0,9355	0,2460	0,2986
Item 4	0,6882	0,4639	0,4419
Item 5	0,7392	0,4396	0,3949
Item 6	0,6747	0,4691	0,2121
Item 7	0,4731	0,4999	0,4418
Item 8	0,7151	0,4520	0,4295
Item 9	0,7124	0,4533	0,4625
Item 10	0,1532	0,3607	0,3868
Item 11	0,6344	0,4822	0,5474
Item 12	0,8145	0,3892	0,3996
Item 13	0,4812	0,5003	0,4114
Item 14	0,6989	0,4593	0,3366
Item 15	0,7903	0,4076	0,4547
Item 16	0,4731	0,4999	0,4418
Item 17	0,7312	0,4439	0,4400
Item 18	0,5215	0,5002	0,3699
Item 19	0,5430	0,4988	0,3911
Item 20	0,2554	0,4367	0,0446
Item 21	0,7527	0,4320	0,3585
Item 22	0,5000	0,5007	0,3946
Item 23	0,4758	0,5001	0,2021
Item 24	0,6425	0,4799	0,4482
Item 25	0,6210	0,4858	0,3533
Item 26	0,5054	0,5006	0,4085
Item 27	0,2849	0,4520	0,3793
Item 28	0,3414	0,4748	0,3965

When Table 5 is examined, it can be seen that item difficulty values varied between 0.15 and 0.94 while discriminant indexes were found to vary between 0.04 and 0.55. According to this result, it was decided to subtract from the measurement tool because the discriminant index of item 20 in the test was too low and it did not have item validity. In addition, the reliability coefficients of the 27 items included in the measurement tool with different reliability methods for each item are shown in Table 6.

Table 6. Reliability Values of 27 Items in the Test

Item number	Guttman's lambda-2	Alpha coefficient	Feldt-Gilmer coefficient	Feldt-Brennan coefficient	Raju's beta coefficient
Item1	0,8536	0,8498	0,8517	0,8513	0,8498
Item2	0,8551	0,8515	0,8532	0,8529	0,8515
Item3	0,8572	0,8541	0,8554	0,8550	0,8541
Item4	0,8535	0,8498	0,8516	0,8513	0,8498
Item5	0,8548	0,8512	0,8529	0,8526	0,8512
Item6	0,8599	0,8566	0,8581	0,8579	0,8566
Item7	0,8535	0,8498	0,8516	0,8513	0,8498
Item8	0,8537	0,8500	0,8518	0,8514	0,8500
Item9	0,8526	0,8490	0,8507	0,8504	0,8490
Item10	0,8555	0,8521	0,8538	0,8534	0,8521
Item11	0,8500	0,8464	0,8480	0,8477	0,8464
Item12	0,8549	0,8512	0,8530	0,8526	0,8512
Item13	0,8547	0,8510	0,8528	0,8525	0,8510
Item14	0,8563	0,8526	0,8544	0,8541	0,8526
Item15	0,8531	0,8494	0,8512	0,8509	0,8494
Item16	0,8534	0,8497	0,8514	0,8511	0,8497
Item17	0,8533	0,8497	0,8515	0,8511	0,8497
Item18	0,8556	0,8519	0,8537	0,8534	0,8519
Item19	0,8548	0,8511	0,8529	0,8526	0,8511

Item21	0,8552	0,8519	0,8537	0,8533	0,8519
Item22	0,8546	0,8510	0,8528	0,8525	0,8510
Item23	0,8610	0,8579	0,8594	0,8592	0,8579
Item24	0,8532	0,8495	0,8513	0,8509	0,8495
Item25	0,8555	0,8520	0,8539	0,8535	0,8520
Item26	0,8547	0,8510	0,8528	0,8525	0,8510
Item27	0,8555	0,8519	0,8537	0,8534	0,8519
Item28	0,8550	0,8516	0,8535	0,8531	0,8516

When Table 6 is examined, it is seen that 27 items in the measurement instrument have a reliability coefficient of over .80 which is considered as critical with 5 different reliability determination methods (Cronbach, 1951). According to this result, it was determined that the items in the measurement instrument met the assumption of reliability in terms of internal consistency. The reliability coefficients calculated for the whole scale and the 95% confidence interval and standard error values for this value are shown in Table 7.

Table 7. Results of the reliability analysis for the total of 27 items

Method	Coefficient	Confidence interval lower limit	Confidence interval upper limit	Standard error
Guttman's L2	0,8595	0,8380	0,8793	2,1185
Coefficient Alpha	0,8560	0,8340	0,8763	2,1445
Feldt-Gilmer	0,8577	0,8360	0,8777	2,1319
Feldt-Brennan	0,8574	0,8356	0,8775	2,1341
Raju's Beta	0,8560	0,8340	0,8763	2,1445

When the Table 7 is examined, it is determined that the reliability coefficients determined by different methods for the whole scale consisting of 27 items vary between .856 and .859. According to this result, the results obtained from the measuring instrument are considered to be reliable. With the thought that it will not be sufficient to look only at the item difficulty and discrimination indices in the determination of the items in the scale, compatibility indices with transformed discriminant indices related to items were calculated by Rasch analysis from 1 parameter logistic models. In the Rasch analysis, the item statistics obtained according to the parameters determined as maximum number of iterations 150, convergence criterion 0.005 and endpoints criterion 0.3 are shown in Table 8.

Table 8. Results of Rasch Analysis of Related to the Items in the Scale

Item Number	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
Item1	-0,73	0,13	0,90	-1,65	0,82	-1,21
Item2	-1,22	0,15	0,93	-0,91	0,97	-0,10
Item3	-2,56	0,22	0,94	-0,33	0,52	-1,29
Item4	-0,36	0,13	0,95	-0,83	0,86	-1,17
Item5	-0,71	0,13	0,97	-0,50	0,89	-0,70
Item6	-0,28	0,13	1,20	3,39	1,56	4,08
Item7	0,83	0,12	1,01	0,29	1,05	0,59
Item8	-0,56	0,13	0,92	-1,30	1,26	1,75
Item9	-0,55	0,13	0,88	-2,03	0,85	-1,06
Item10	3,14	0,18	0,87	-1,09	2,06	3,40
Item11	-0,09	0,12	0,81	-3,81	0,75	-2,48
Item12	-1,25	0,15	0,93	-0,84	0,83	-0,81
Item13	0,77	0,12	1,06	1,09	1,27	3,03
Item14	-0,51	0,13	1,03	0,58	0,96	-0,24
Item15	-1,14	0,15	0,84	-2,20	0,70	-1,69
Item16	0,72	0,12	1,00	0,05	1,07	0,87
Item17	-0,70	0,13	0,93	-1,12	0,80	-1,34
Item18	0,55	0,12	1,07	1,31	1,08	0,93
Item19	0,41	0,12	1,01	0,31	0,99	-0,10
Item21	-0,91	0,14	0,98	-0,31	1,12	0,73
Item22	0,60	0,12	1,06	1,26	1,13	1,49
Item23	0,77	0,12	1,31	5,72	1,37	4,05
Item24	-0,21	0,13	0,95	-0,95	0,89	-0,98

Item25	-0,05	0,13	1,07	1,27	1,04	0,40
Item26	0,55	0,12	1,02	0,46	1,08	0,96
Item27	1,93	0,14	1,02	0,24	1,20	1,55
Item28	1,59	0,13	0,98	-0,30	1,20	1,86

Unweighted Mean Square (UMS) and Weighted Mean Square (WMS) fit, which are shown in Table 8, are compliance statistics for fit statistics. From these values, WMS is accepted as the in fit criteria and UMS as the outfit criteria. When Table 8 is examined, it is seen that the scale consisting of 27 items is suitable for the measurement process according to WMS and UMS values. In other words, the items in the test were found to be in the ideal range in terms of non-compliance measures. In addition, according to standardized UMS and WMS values, item 10, item 11 and item 23 do not meet the compliance criteria, in other words, it was determined that the items do not provide the model data compliance. According to this result, it was decided that 4 items in the test should be removed. Accordingly, for the items in the Quadrature test consisting of a total of 23 items, the bias analysis which is another method of determining the validity of the items in the scale was started. DIF analysis was performed to determine whether the items are biased. In the comparison of the bias of the items in the scale, the type of school where the students studied was determinant. As a result of the determination of the students from private schools as the focus group and public school students of as reference group, the results obtained by using the Mantel-Haenszel method on the basis of common odds ratio are shown in Table 9.

Table 9. DIF Analysis Results

Item No	Chi-square	p-value	Valid N	Effect size	Class
Item1	4,40	0,04	175	2,67	B-
Item2	0,48	0,49	137	1,60	A
Item3	0,20	0,66	42	0,00	A
Item4	1,01	0,31	194	0,52	A
Item5	0,12	0,73	154	0,81	A
Item7	8,77	0,00	211	0,28	B+
Item8	0,27	0,61	194	1,35	A
Item9	0,67	0,41	201	1,60	A
Item12	0,53	0,47	125	0,59	A
Item13	0,13	0,72	220	0,87	A
Item14	4,93	0,03	192	0,27	B+
Item15	1,46	0,23	112	2,27	A
Item16	0,18	0,67	220	1,15	A
Item17	5,70	0,02	163	4,40	B-
Item18	0,01	0,92	211	0,96	A
Item19	3,00	0,08	201	2,24	A
Item21	53,37	0,00	192	40,63	C-
Item22	9,65	0,00	220	2,97	B-
Item24	0,04	0,85	201	1,10	A
Item25	10,45	0,00	201	3,31	B-
Item26	4,03	0,04	220	0,45	B+
Item27	5,79	0,02	220	0,44	B+
Item28	58,84	0,00	220	0,04	C+

When Table 9 is examined, it was determined that all of the items other than item21 and item28 are negligible and have got insignificant level of item function. However, the item characteristic curves (Item Characteristic Curve) were used to determine whether the two items identified as high (C + or C-) biased were working in favor of a particular group of students from public and private schools or against the other group. The material characteristic curves of the items determined to show significant DMF are shown in Figure 2.
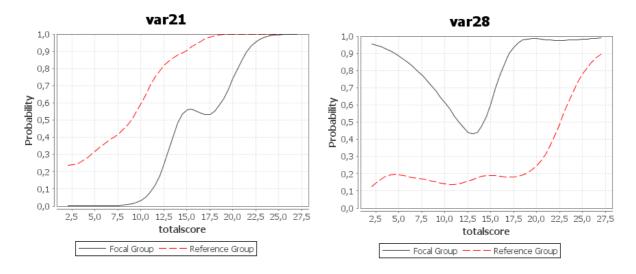
Figure 2. ICC Regarding Items Showing Significant Level of DMF

When Figure 2 was examined, it was determined that when the total scores taken are in the range of 2.50-20.00, item 21 is in favor of the public school students (C-) determined as the reference group. On the other hand, when the total scores taken from the test were within the range of 5,00-25,00, it was determined that item 28 works in favor of the private school students determined as the focus group (C +). According to this result, it was decided that both of the items in the test should be removed from the test (Koyuncu, Aksu and Kelecioğlu, 2018). After this process, the final version of the measurement tool was determined to include a total of 21 items. The test statistics obtained for the last version of the scale are shown in Table 10.

Table 10. Item Statistics for Final Version of the Test

Item Statistics	Value
Item Number	21
Minimum	2,00
Maximum	21,00
Average	13,43
Median	14,00
Standart Deviation	4,60
Quarter opening	7,00
Skewness	-0,26
Kurtosis	-0,89
KR21	0,809
Gutman L2	0,836
Raju's Beta Value	0,834

When Table 10 is examined, it was determined that the reliability coefficient obtained according to the three different methods of the test item consisting of 21 items varied between .81 and .84 and therefore the results obtained from the test were reliable (Cronbach, 2004). When the results were examined as a whole, it was decided that the results obtained from a total of 21 items were valid and reliable.

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

The purpose of the study is to develop an achievement test consisting of multiple choice questions which are valid and reliable in accordance with the learning outcomes of the quadrilaterals in order to determine the knowledge of the secondary school students about the definitions of the quadrilaterals, to determine the special cases of the quadrilaterals and to classify the quadrilaterals hierarchically. As it is known, multiple choice tests, which are used to measure student success, provide the opportunity to ask a large number of questions, and are among the most frequently used

measurement tools that are used today to poll all of the subjects learned in the courses in a short time (Kempa, 1986; Ogan Bekiroğlu, 2004). Multiple-choice tests provide with information about the errors students have and offers a wide range of questions about a subject or unit and the ability to measure and evaluate all knowledge. Besides, multiple choice tests are the most preferred measurement tool in measuring the success of students among the other measurement and evaluation tools in education. Achievement tests are passed through the standardized stages and tests which ensure the reliability and validity are provided (Narlı & Başer, 2008).

As a result of analysis, it is understood that there are a certain amount of both very difficult and very easy items in the test used in the study. The average value obtained shows that the test is not too difficult and has got medium difficulty.

The value of item discrimination, which determines the extent to which the items of the test are measured, is between +1 and -1. As this value approaches + 1, it is evaluated that the item measures the feature that the test aims to measure better, and that the closer to 0, the more the item is inadequate in measuring the property the test aims to measure. If the value of discrimination is minus, it is thought that the item measures another characteristic than the purpose (Kan, 2011). It was determined that the test used in this study consisted of 21 items and the average difficulty index of the test was .64; the discrimination index was .40. This result demonstrates that the discriminant of the test is sufficient. The Cronbach Alpha value, which was used to determine the reliability of the test prepared for quadrangles, was calculated as .82.

The validity and reliability analysis of the developed test showed that the test was a valid and reliable test. Since difficulty and discrimination indices are at desired levels showed that the test can be used as an achievement test. Based on the findings of this study, the following recommendations were made:

• This achievement test can be used by researchers who carry out relevant studies in order to determine the target and behavioral learning outcomes of the students.

• Although it is determined that the achievement test has sufficient validity and reliability, it can be said that the test needs supportive studies.

• It is believed that a test prepared in accordance with the acquisition of the curriculum will enable teachers to get information about his/her students' readiness and misconceptions and to teach their lessons in more planned and efficient way.

REFERENCES

Aktaş (Cansız), M. ve Aktaş, D.Y. (2011). *8. Sınıf öğrencilerinin dörtgenleri köşegen özelliklerinden yararlanarak tanıma sürecinin incelenmesi*, 10. Matematik Sempozyumunda sözlü olarak sunulmuştur. İstanbul, Işık Üniversitesi.

Aktaş, D. Y., ve Aktaş, M. C. (2012). 8. sınıf öğrencilerinin özel dörtgenleri tanıma ve aralarındaki hiyerarşik sınıflamayı anlama durumları. *İlköğretim Online*, *11*(3).

Aktaş, D.Y. (2005). *İşbirliğine dayalı grup çalışması ile öğrencilerin geometri anlama düzeylerinin incelenmesi*. Yayınlanmamış yüksek lisans tezi, Karadeniz Teknik Üniversitesi, Trabzon.

Akuysal, N. (2007). *İlköğretim 7. sınıf öğrencilerinin 7. sınıf ünitelerindeki geometrik kavramlardaki yanılgıları*. Yayınlanmamış yüksek lisans tezi, Selçuk Üniversitesi, Konya.

Baykul, Y. (1999). İlköğretimde matematik öğretimi. Ankara: Anı yayıncılık.

Baykul, Y. (2002). *İlköğretimde matematik öğretimi*. Ankara: Pegem A Yayıncılık.

Baykul, Y. ve Güzeller, C. O. (2014). *Sosyal bilimler için istatistik SPSS Uygulamalı*, Ankara: Pegem Akademi.

Bekiroğlu, F. O. (2004). *Ne kadar başarılı: klasik ve alternatif ölçme ve değerlendirme yöntemleri: fizikte uygulamalar*. Nobel Yayın Dağıtım.

Burns, J. C., Okey, J. R. & Wise, K. C. (1985). Development of an Integrated Process Skill Test: TIPS II. Journal of Research in Science Teaching, *22*(2), 169-177.

Bütüner, S. Ö., & Filiz, M. (2016). Matematik öğretmeni adaylarının dörtgenleri sınıflandırma becerilerinin incelenmesi. *Alan Eğitimi Araştırmaları Dergisi*, *2*(2), 43-56.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*,16, 297-334.

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.

Çalışkan, İ. Ö. ve Kaptan, F. (2009). Constructing Science Process Skills Test. *Çağdaş Eğitim Dergisi*, *34*(369), 27-34.

De Villiers, M. (1998). To teach definitions in geometry or teach to define? Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education, 2, 248-255.

Duatepe, A. (2000). An investigation of the relationship between van hiele geometric level of thinking and demographic maddeiable for pre-service elementary school teacher. Yayımlanmamış yüksek lisans tezi, ODTÜ, Ankara.

Erez, M. M., & Yerushalmy, M. (2006). "If you can turn a rectangle into a square, you can turn a square into a rectangle..." Young students experience the dragging tool. *International Journal of Computers for Mathematical Learning*, *11*(3), 271-299.

Ergün, S. (2010). İlköğretim *7. sınıf öğrencilerinin çokgenleri algılama, tanımlama ve sınıflama biçimleri.* Yayınlanmamış yüksek lisans tezi. Dokuz Eylül Üniversitesi, İzmir.

Fischbein, E. (1993). The theory of figural concepts. Educational Studies in Mathematics, 24 (2), 139-299.

Fujita, T. (2008). Learners' understanding of the hierarchical classification of quadrilaterals. *Proceedings of the British Society for Research into Learning Mathematics*, *28*(2), 31-36.

Fujita, T. (2012). Learners' level of understanding of the inclusion relations of quadrilaterals and prototype phenomenon. *The Journal of Mathematical Behavior*, *31*(1), 60-72.

Fujita, T. ve Jones, K. (2007). Learners' understanding of the definitions and hierarchical classification of quadrilaterals: towards a theoretical framing. *Research in Mathematics Education*, *9*(1), 3-20.

Fujita, T.,& Jones, K. (2006). Primary trainee teachers' understanding of basic geometrical figures in scotland. In J. Novotana, H. Moraova, K. Magdelena & N. Stehlikova (Eds.), Proceedings of The 30*th* Conference of the International Group for the Psychology of Mathematics Education,3, 14-21.

Gülbağcı, S. (2009). İlköğretim 7. sınıf dörtgenler konusunun öğretiminde dinamik geometri yazılımlarının etkisi. *Yayınlanmamış Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.*

Hershkowitz, R.(1990).Psychological aspects of learning geometry. In P. Nesher & J.Kilpatrick (Eds.),

Horzum, T. (2017). Matematik Öğretmeni Adaylarının Dörtgenler Hakkındaki Anlamalarının Kavram Haritası Aracılığıyla İncelenmesi. *Turkish Journal of Computer and Mathematics Education(TURCOMAT)*, 1-1.

Kan, A. (2011). *Ölçme Aracı Geliştirme*. S. Tekindal içinde, Eğitimde Ölçme ve Değerlendirme (s. 240-277). Ankara: Pegem Akademi.

Karakuş, F. ve Erşen, Z. B. (2016). Sınıf Öğretmeni Adaylarının Bazı Dörtgenlere Yönelik Tanımlama ve Sınıflamalarının İncelenmesi. *Karaelmas Eğitim Bilimleri Dergisi*, *4*(1).

Karslı, F. ve Ayas, A. (2013). Fen ve Teknoloji Dersi Öğretmen Adaylarının Bilimsel Süreç Becerilerinin Ölçülmesine İlişkin Bir Test Geliştirme Çalışması, *Türk Fen Eğitimi Dergisi*, *10*(1), 66-84.

Kayış, A. (2010). *Güvenilirlik Analizi*. Ş. Kalaycı içinde, SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri (s. 404-409). Ankara: Asil Yayın.

Kempa, R. (1986). Assessment in Science. Cambridge University Press, Cambridge, London.

Mathematics and cognition (pp. 70-95).Cambridge: Cambridge University Press.

Koyuncu, İ , Aksu, G , Kelecioğlu, H . (2018). Mantel-Haenszel, Lojistik Regresyon ve Olabilirlik Oranı Değişen Madde Fonksiyonu İnceleme Yöntemlerinin Farklı Yazılımlar Kullanılarak Karşılaştırılması. İlköğretim Online, 17 (2), 909-925. DOI: 10.17051/ilkonline.2018.419339

MEB. (2013a). İlkokul matematik dersi 1-4. Sınıflar öğretim programı. Ankara: MEB Talim ve Terbiye Kurulu Başkanlığı.

MEB. (2013b). Ortaokul matematik dersi 5-8. Sınıflar öğretim programı. Ankara: MEB Talim ve Terbiye Kurulu Başkanlığı.

Monaghan, F. (2000). What difference does it make? Children's views of the differences between some quadrilaterals**.** *Educational Studies in Mathematics*, *42* (2),179-196.

Narlı, S. ve Baser, N. E. (2008). " Küme, Bağıntı, Fonksiyon" Konularında Bir Başarı Testi Geliştirme ve Bu Test İle Üniversite Matematik Bölümü 1. Sınıf Öğrencilerinin Bu Konulardaki Hazırbulunuşluklarını Betimleme Üzerine Nicel Bir Araştırma. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (24).

NCTM (2000). *Principles and Standards for School Mathematics*. Reston, Va. NCTM.

Okazaki, M., & Fujita, T. (2007, July). Prototype phenomena and common cognitive paths in the understanding of the inclusion relations between quadrilaterals in Japan and Scotland. In *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 41-48).

Okumuş, S. (2011). Dinamik geometri ortamlarının 7. sınıf öğrencilerinin dörtgenleri tanımlama ve sınıflandırma becerilerine etkilerinin incelenmesi. *Unpublished master's thesis), Karadeniz Technical University, Trabzon*.

Olkun, S. ve Aydoğdu, T. 2003. Üçüncü Uluslararası Matematik ve Fen Araştırması (TIMSS) Nedir? Neyi Sorgular? Örnek Geometri Soruları ve Etkinlikleri. İlköğretim online, 2(1), 25 Eylül 2005 tarihinde http://ilkogretimonline.org.tr/vol2say1/v02s01d.htm adresinden alınmıştır.

Pickreign, J. (2007). RectangleandRhombi: How Well Do Pre-Service TeachersKnowThem? Issuesin theUndergraduateMathematicsPreparation Of School Teachers, *IUMPST*,1, 24 Ekim 2011 tarihinde http://www.k12prep.math.ttu.edu/journal/contentknowledge/pickreign01/article.pdf adresinden alınmıştır.

Tall, D. ve Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational studies in mathematics*, *12*(2), 151-169.

Toluk, Z., Olkun, S. ve Durmuş, S. (2002). Problem merkezli ve görsel modellerle destekli geometri öğretiminin sınıf öğretmenliği öğrencilerinin geometrik düşünme düzeylerinin gelişimine etkisi. *V. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi*, *2*, 913-920.

Türnüklü, E. (2014). Dörtgenlerde aile ilişkilerinin yapılandırılması: ilköğretim matematik öğretmen adaylarının ders planlarının analizi. *Eğitim ve Bilim*, *39*(173).

Türnüklü, E., Alaylı, F. G. ve Akkaş, E. N. (2013). İlköğretim matematik öğretmen adaylarının dörtgenlere ilişkin algıları ve imgelerinin incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, *13*(2), 1213-1232.

Webb, N.L. (1997). Determining Alignment of Expectations and Assessments in Mathematics and Science Education. NISE Brief 1(2). Madison, WI: University of Wisconsin Madison, National Institute for Science Education.