

DEVELOPING AN ACHIEVEMENT TEST FOR FRACTION TEACHING: VALIDITY AND RELIABILITY ANALYSIS

Belgin Bal-Incebacak^{1,*}, *Esen ERSOY*²

¹Education Faculty, Primary School Department, 55139 Samsun, Turkey

¹Education Faculty, Mathematics Department, 55139 Samsun, Turkey

Abstract. The aim of the study is to develop an achievement test that can be used to measure the achievement status of elementary school 4th grader students in terms of their fractions learning in mathematics courses. Examining the literature, it is visible that there are 8 development stages of the achievement test.. ITEMAN 3.5 package software was employed to analyze the validity and reliability of the test. Statistical analysis of the test reveals that KR-20 Internal Consistency and KR-21 Internal Consistency are 0.90. This value indicates that the consistency of the test is high. The average discrimination of the test is 0.80, which indicates that the test is highly distinctive among students. The average strength of the test has been calculated as 0.67. Therefore, it has been determined that the test should be both easy and distinctive when it comes to measuring the effectiveness of the teaching method. The result of the analysis suggest that when consistency and distinctiveness values are considered, the developed test is not only easy to use but also highly distinctive. Finally, it has been discovered that the academic achievement test for fraction teaching is valid and reliable.

1. Introduction

Education institutions are institutions that aim to provide students with goals and behaviors that the state considers appropriate. The purpose of education is to give an individual terminal behavior in schools, which is a well-known definition [25]. For this reason, teachers are trying to provide students to obtain information by planning learning experiences and activities. Teachers need to perform assessments to determine the level of knowledge they bring. Assessment is the most important step in the curriculum [12]. Questions such as "How much students could achieve the objectives?" and "How much knowledge, skills and attitudes are brought to students?" are most important questions of the assessment [24]. To perform these assessments, many methods are referred to. Measurement tools such as written examinations, open-ended questions, multiple choice tests, oral examinations, scale types, observation and interviewing are used by researchers and teachers to meet the mentioned demand [57] Achievement is an indication of how

* Corresponding author: belginbal33@gmail.com

much the participants benefit from a course or academic program. As for school success, it can be regarded as the average grade or score of the student from the academic courses in the school [46]. Common measurement tools used in contemporary education systems are multiple choice (standard achievement tests) tests [60]. By using assessment and evaluation methods, achievement is tried to be measured as reliable and valid as possible. The most common tool used for this measurement is multiple-choice tests from achievement tests. While developing an achievement test, there are various stages for a valid and reliable achievement test. The achievement test measures the competence of an individual including knowledge and skills in a field or subject [27]. By another definition, tools developed to measure how much an individual learns in an educational, instructional and learning environment are called achievement tests [57] With these tests, it is tried to determine the level of knowledge of students about a teaching subject.

When we think about mathematics teaching, it can be said that fractions are one of the most challenging topics. [4] stated that the teaching of mathematics is very important because of this difficulty of fractions. When [50] and [62] studies are examined, it is stated that students have low level of comprehension of fractions and mostly lacking knowledge. Fraction teaching is an important issue for mathematics. Among the reasons for being important is that it is related to many subjects in mathematics courses. It is involved in many math related subjects such as measurement, data, ratio and proportion, natural numbers, etc. Fractions are handled as part-whole, measurement, division, processing and ratio. Fractions are given the most part in part-whole relation and there is not much emphasis on its other concepts.

Many studies conducted show that students are unsuccessful in the subject of fractions. [13,53] stated that the students are lacking in operations with fractions. [32,43,51] stated that students have misconceptions regarding the fractions. [31] stated that there are problems with dividing fragments into equal parts and writing the defined fractions, [3] stated that there are problems in understanding basic concepts in every topic of fractions, and [8] stated that there are problems regarding problem solving about fractions. Even if concrete materials were used, they failed in fraction teaching. In short, it has been pointed out by many researchers that students have been unsuccessful in teaching the concept of fractions and many concepts related to fractions, and these subjects have been among the most difficult mathematics subjects to learn [23,33,38,40,49,52,54,58,61]. Şiap and Duru [56] indicate that the reason for this difficulty in fractions arises from the fact that students see the dividend and denominator of the fractions as two separate numbers. The difficulties in the concept of fractions are difficult for students because each process is unique and abstract. Fractions teaching is also very important to assess because it is intertwined with many subjects that make up a great part of mathematics. Correct assessment and evaluation is the most important step in the teaching process, since it will determine whether the desired subject is understood.

The aim of the assessment and evaluation in schools is to determine the achievement status of the students, how successful they are if successful, how unsuccessful they are if unsuccessful, which students have leveled up to next step or which students need to repeat the program [37] It is important for the procedures to be done in a reliable manner because they have an important place in the lives of the students. The data to be collected for this measurement work need an error-free, adequate, valid and reliable measuring tool. Since there are many difficulties about fractions, it is thought that an accurate measurement tool will be needed. For this reason, in the elementary school 4th grade Mathematics course, it was needed to develop an achievement test on fractions, and with this study, it is tried to shed light on new researches and researchers.

2. Method

The research was conducted by survey method. Survey research are studies to determine specific characteristics of a group [17]. According to Karasar [35] survey models are an approach aimed at describing the present situation in its both present and past form.

2.1. Study Group

In the selection of the participants in the study, purposeful sampling was used which allowed the researcher to select the people to answer the research questions [21]. The sample of the study consisted of 431 students who are studying in Samsun city center in 2015-2016.

Although the test developed in the research is directed to the fourth grade fractions unit, the inclusion of the fifth grade students in the sample of the research is to consider that the students in the 4th grade should not have studied this unit and that they will tend to leave the questions unattended when the test is implemented. In this research, the sample was formed of primary school 5th grade students for the purpose of minimizing this problem and ensuring all questions are answered.

The distribution of the sample by schools and gender is shown in Table 1.

Table 1: The distribution of the sample by schools and gender

School Type	Class	Gender and Numbers of Students			
		Female	Male	Total	Total
Normal State School (Secondary School)	A	11	12	23	170
	B	12	12	24	
	C	10	10	20	
	D	15	15	30	
	E	17	21	38	
	F	20	15	35	
YIBO (Boarding Elementary School Regional Middle School)	A	12	13	25	125
	B	14	16	30	
	C	20	20	40	
	D	16	14	30	
Imam-Hatip Secondary School	A	20	20	40	136
	B	22	24	46	
	C	9	10	19	
	D	16	15	31	
Total		214	217	431	431

When the table was examined, a total of 431 students were studied. Of these students, 214 are female and 217 are male. According to the school types, studies were carried out with 3 different schools: normal public school, regional boarding middle-school and vocational religious middle school. Our study group consists of 170 students from normal public school, 125 from regional boarding middle school and 136 from vocational religious middle school.

2.2. Data Analysis

All the steps of the Achievement Test for Fraction Teaching, which is planned to be developed in the scope of the study, are described in detail below. The achievement test was developed by following the process steps specified by [6].

2.2.1. The Area to be used for Test Scores

The purpose of the test to be applied in the test development process should be clearly stated. Every work to be done after the determination of the purpose is to realize this purpose in effective and efficient manner [44] For this purpose, we aim to develop the achievement test to determine the knowledge of students on the subject of fractions and to determine whether all achievements have been achieved at the end of the training.

2.2.2. Determining the behaviors representing the area or the statement

At this stage of the test development, learning areas, sub-learning areas and achievements related to these learning areas are determined within the scope of the test. In this research, the achievements of learning and sub-learning areas taken from the Curriculum of Elementary School Mathematics Lesson (1, 2, 3 and 4th Grades) dated 30/06/2005 and No. 2575 (2015) and presented in Table 2 were included within the scope of achievement test. This table of specifications was arranged in accordance with the "Cognitive Domain Taxonomy" prepared by Bloom et al [14]

Table 2. Table of specifications

			Time	Cognitive Domain Steps			Total Number of Questions Written
Learning Area	Sub-Learning Area	Achievements		Recall	Understanding	Application	
Numbers	Fractions	M4.1.27. Recognizes simple, compound, and integer fractions and presents with models.	12	x	x	x	8
		M4.1.28. Compares and orders unit fractions.			x	x	5
		M4.1.29. Determines a specified simple fraction of a quantity.			x		1
		M4.1.30. Compares up to three fractions with same or multiple denominators.			x	x	3
	Operations with Fractions	M4.1.31. Performs addition and subtraction operations with equal denominator fractions.	5			x	4
		M4.1.32. Solves problems requiring addition and subtraction operations with fractions.		x	x	x	5
	Decimal Notation	M4.1.33. Specifies when a whole is divided into 10 and 100 equal parts that the resulting units of fraction can be expressed in decimal notation.	12	x	x		3
		M4.1.34. Writes a fraction with a denominator of 10 and 100 by using decimal notation.		x	x	x	5
		M4.1.35. It specifies the whole part, fraction part and step names of the decimal notation.			x	x	3
		M4.1.36. Compares two numbers which the decimal notations are given and presents the relation between them by larger, smaller or equal symbol.			x	x	6
Total			29				43

The programs showed way to Bloom's Cognitive Domain Taxonomy, which was developed in 1956, when they were being developed [16, 28]. The six-stage original bloom taxonomy consists of categories of knowledge, comprehension, application, analysis, synthesis and evaluation, from simple to complex, in accordance with the cognitive domain [14, 15]. According to the table of specifications, Numbers were discussed in 10 achievements according to old and 13 achievements according to new program, consisting of 3 sub-learning fields as fractions, operations in fractions and decimal notation (Names fractions, of which the dividend and denominator are maximum two digit natural numbers, obtaining from the units of the fraction. Shows the fractions, of which the dividend and denominator are maximum two-digit number on the numerical axis. Compares fractions. Orders a maximum of four fractions with equal denominator, from largest to smallest or

smallest to largest. Orders a maximum of four fractions with equal dividends and different denominators, from largest to smallest or smallest to largest. Determines a specified simple fraction of a quantity. Specifies when a whole is divided into 10 and 100 equal parts that the resulting units of fraction are fractions. Writes decimal fractions using commas. It specifies the whole part, fraction part and step names of the decimal fractions. Compares two decimal fractions and shows the relation between them with a bigger, smaller and equal symbol. Performs addition operation with equal denominator fractions. Performs subtraction operations with equal denominator fractions. Solves and poses problems requiring addition and subtraction operations with fractions.) A total of 43 questions were written in accordance with these achievements.

2.2.3. Writing Test Items

At this stage, as the achievement test was developed for the 4th grades in primary school, questions suitable for the dimensions of remembering, understanding and application. Analysis, synthesis and evaluation are considered as metacognitive levels while knowledge, understanding, and practice are considered sub-levels[5]. When these questions are being created, care has been taken to ensure that they are appropriate for achievement and grade level. Attention has been paid to the rules of item writing and the types of items while the questions are being created. If the appropriate item type is not selected, the item will have low validity, since the behavior to be measured with the item cannot be measured sufficiently or not at all [6]. When writing the test items, attention was paid to the average difficulty of the test and the difficulty distribution of the test items. The positions of the questions have been changed in accordance with expert opinions by the difficulty and easiness and by achievement.

2.2.4. Reviewing The Test Items

The test items created need to be developed and reviewed in accordance with certain criteria. It is emphasized in the literature that it should be reviewed in terms of compliance with the following qualifications:

- For validity, determining whether the behavior to be measured is worth measuring,
- Determination of the accuracy in terms of science,
- Determination of whether it is comprehensible in terms of language and there are grammar and spelling errors,
- Determination of whether there are technical defects,
- Determination of whether it is suitable for development characteristics of students [11,39,59] should be done.

Table 3. Expert opinion table

Question No: 1			
Evaluation Criteria	Yes	No	Suggestion (if any)
Question text clear, understandable			
Question is appropriate for 4. grade			
Distractors are well prepared			
Question is appropriate for achievement			
Which cognitive process dimension should the question be placed according to Bloom's taxonomy.	Recall (Knowledge)	Understanding (Comprehension)	Application

Compliance with the criteria was determined by the questions in Table 3 prepared. This table was prepared for each question and opinions were taken from the experts in Table 4. Achievement test, consisting of 43 test items, were presented to a total of 17 experts, 12 academicians, 3 elementary mathematics teachers and 2 class teachers, in order to determine the compliance to the above mentioned qualifications.

Achievement test, consisting of 43 test items, were presented to a total of 17 experts, 12 academicians, 3 elementary mathematics teachers and 2 class teachers, in order to determine the compliance to the above mentioned qualifications.

Table 4. Expert opinion table

Kod	Sex	University/Shocol	Degree	Education
1	Male	Sinop University	Associate Professor	Mathematics Education
2	Male	Ondokuz Mayıs University	Associate Professor	Science Education
3	Male	İstanbul University	PhD, Assistant Professor	Mathematics Education
4	Female	Dokuz Eylül University	PhD, Assistant Professor	Mathematics Education
5	Female	İstanbul University	PhD, Assistant Professor	Mathematics Education
6	Female	Ondokuz Mayıs University	PhD, Assistant Professor	Mathematics Education
7	Female	Ondokuz Mayıs University	PhD, Assistant Professor	Mathematics Education
8	Female	Ondokuz Mayıs University	PhD, Assistant Professor	Mathematics Education
9	Female	İstanbul University	Research Assistant PhD	Mathematics Education
10	Female	Ondokuz Mayıs University	Research Assistant PhD	Science Education
11	Male	Mersin University	Research Assistant PhD	Assesment and evaluation
12	Female	Ondokuz Mayıs University	Research Assistant	Mathematics Education
13	Male	Ministry of Education	Master Degree Teacher	Primary School
14	Female	Ministry of Education	Teacher	Primary School
15	Female	Ministry of Education	Master Degree Teacher	Mathematics Education
16	Male	Ministry of Education	Master Degree Teacher	Mathematics Education
17	Male	Ministry of Education	Master Degree Teacher	Mathematics Education

It has been examined in terms of scope validity with the expert group formed by teachers and academic members. Such procedures increase the validity and reliability of the test [1,47].

2.2. 5.Preparing The Test Form

After the expert opinions are obtained, necessary arrangements are made and questions should be sorted in a certain order. Below are the steps to be considered when preparing the test form. These steps consist of three stages;

1. Distribution of the items in the test form
2. Writing of test instructions
3. Writing of items [6]

According to the opinions taken from the experts, the test was shaped by correcting the errors. Formally unsuitable questions, unreadable, incomplete and hard-to-comprehend questions are corrected. In addition, while the questions were being edited, [6] emphasized that the items should be grouped according to the subjects and the questions should be ordered from the easy to difficult, without disturbing the objective unity. Font size in the literature for the fourth grade students was determined as 12. According to this information the question texts are adjusted to be size 12 fonts. After these adjustments are made, information such as "Switch to the next page" or "Switch to the other page" should be placed. At the end of the test, there should be information such as "Test is over" and "Check your answers" should be placed [6].

2.2. 6.Putting The Test on A Trial Implementation

Before the moving on to application, it is necessary to determine the duration for the test. In this context, a random class was selected from a school with a low level of socio-economic level at the 5th grade level and a pilot application of the test was conducted on a total of 31 students without a time limit. As a result of the pilot application, it was determined that the best student in the class finished in 35 minutes, and a student who was lower in level finished within 50 minutes. Students at the average level finished in 45 minutes. In this context, the duration of the test was determined as 45 minutes.

The final form was given to the test, the application duration was determined as 45 minutes and trial application was performed. Accordingly, the test was applied to the 5th grade students so that the final test can be formed. As the number of students in the test group increased, the closeness of statistics obtained would get nearer to the actual statistics, so the test was applied to 400 students.

2.2.7. Selecting Materials by Analyzing Them According to The Trial Implementation

After the test application, item analysis was carried out so that the desired test can be formed, and suitable items for the final test were selected. IBM-SPSS, FINESSE, ITEM and hand calculations can be done to analyze the validity and reliability of a test. It is preferred to use ITEMAN 3.5 package program in order to obtain the most accurate and correct analyzes. The reason for preferring this program is to give an idea about the statistics of each question as well as their options. The distinctiveness and difficulty indices of the items are examined in the item selection [39] and also the distractor characteristics of the questions to be selected are reviewed. The ITEMAN program calculates p (item difficulty) and r (distinctiveness power) values for each question. Table 5 shows the item difficulty and distinctiveness index of each item in the item analysis for the developed achievement test.

Table 5. Item difficulty and distinctiveness index of each item in the item analysis

Seg No.	Number of Replies	Prop. Correct	Disc. Index (Point-to-Double Series Correlation)	item reliability	Item Varyans	Item Standard deviation	item correlation (Pearson)	item difficulty index evaluated	Item distinctiveness index evaluated
	Ij	pj	rjx	rj	sj ²	sj	rjy		
1	27	0,87	0,74	0,25	0,11	0,34	0,67	Easier *	Good item
2	30	0,97	0,80	0,14	0,03	0,18	0,57	Easier *	Good item
3	20	0,65	0,61	0,29	0,23	0,48	0,59	Easier *	Good item
4	25	0,81	0,55	0,22	0,16	0,40	0,51	Easier *	Good item
5	30	0,97	0,62	0,11	0,03	0,18	0,45	Easier *	Good item
6	12	0,39	0,54	0,26	0,24	0,49	0,53	Difficult ***	Good item
7	27	0,87	0,74	0,25	0,11	0,34	0,67	Easier *	Good item
8	19	0,61	0,66	0,32	0,24	0,49	0,65	Easier *	Good item
9	19	0,61	0,59	0,29	0,24	0,49	0,57	Easier *	Good item
10	20	0,65	0,62	0,30	0,23	0,48	0,60	Easier *	Good item
11	13	0,42	0,70	0,34	0,24	0,49	0,69	Medium difficulty **	Good item
12	16	0,52	0,29	0,15	0,25	0,50	0,29	Medium difficulty **	Analyze-correct
13	14	0,45	0,39	0,19	0,25	0,50	0,38	Medium difficulty **	Not distinctive
14	22	0,71	0,75	0,34	0,21	0,45	0,72	Easier *	Good item
15	5	0,16	-0,35	-0,13	0,14	0,37	-0,35	Difficult ***	Not distinctive
16	13	0,42	0,60	0,30	0,24	0,49	0,59	Medium difficulty **	Good item
17	5	0,16	-0,32	-0,12	0,14	0,37	-0,32	Difficult ***	Not distinctive
18	3	0,10	0,14	0,04	0,09	0,30	0,14	Difficult ***	Not distinctive

19	17	0,55	0,72	0,36	0,25	0,50	0,71	Medium difficulty **	Good item
20	6	0,19	-0,04	-0,02	0,16	0,40	-0,04	Difficult ***	Not distinctive
21	9	0,29	0,02	0,01	0,21	0,45	0,02	Difficult ***	Not distinctive
22	7	0,23	-0,40	-0,17	0,17	0,42	-0,40	Difficult ***	Not distinctive
23	4	0,13	-0,09	-0,03	0,11	0,34	-0,09	Difficult ***	Not distinctive
24	3	0,10	0,20	0,06	0,09	0,30	0,20	Difficult ***	Analyze-correct
25	5	0,16	0,35	0,13	0,14	0,37	0,35	Difficult ***	Distinctive
26	17	0,55	0,31	0,16	0,25	0,50	0,31	Medium difficulty **	Distinctive
27	22	0,71	0,44	0,20	0,21	0,45	0,42	Easier *	Good item
28	17	0,55	0,76	0,38	0,25	0,50	0,75	Medium difficulty **	Good item
29	18	0,58	0,46	0,23	0,24	0,49	0,45	Medium difficulty **	Good item
30	16	0,52	0,74	0,37	0,25	0,50	0,73	Medium difficulty **	Good item
31	17	0,55	0,76	0,38	0,25	0,50	0,75	Medium difficulty **	Good item
32	17	0,55	0,31	0,16	0,25	0,50	0,31	Medium difficulty **	Distinctive
33	22	0,71	0,44	0,20	0,21	0,45	0,42	Easier *	Good item
34	17	0,55	0,76	0,38	0,25	0,50	0,75	Medium difficulty **	Good item
35	18	0,58	0,46	0,23	0,24	0,49	0,45	Medium difficulty **	Good item
36	16	0,52	0,74	0,37	0,25	0,50	0,73	Medium difficulty **	Good item
37	17	0,55	0,76	0,38	0,25	0,50	0,75	Medium difficulty **	Good item
38	17	0,55	0,31	0,16	0,25	0,50	0,31	Medium difficulty **	Distinctive
39	22	0,71	0,44	0,20	0,21	0,45	0,42	Easier *	Good item
40	17	0,55	0,76	0,38	0,25	0,50	0,75	Medium difficulty **	Good item
41	18	0,58	0,46	0,23	0,24	0,49	0,45	Medium difficulty **	Good item
42	17	0,55	0,31	0,16	0,25	0,50	0,31	Medium difficulty **	Distinctive
43	22	0,71	0,44	0,20	0,21	0,45	0,42	Easier *	Good item

The distinctiveness index value can be found through the internal consistency Cronbach Alpha item-total table calculated by SPSS software. The item difficulty index takes values between 0 and 1. The difficulty index approaching to 0 indicates that the item is getting difficult, approaching to 1 indicates that the item is getting easier, being 0.50 indicates that the item has medium difficulty [18,19,21,45]. Item distinctiveness index is an important decision criterion when items are included in the test. It allows to distinguish between the informed and the uninformed student. The item difficulty index takes values between -1 and +1 [48]. It can be said that the item difficulty is related to the reliability, one of the qualifications required in measurement tool" [10,57]. The fact that the difficulty of any item in an achievement test is moderate, is a sign that that item is a good item.

The criteria to be used by Ebel (1995) in determining which of the test items should be tested is shown in Table 6 [57]

Table 6. Item selection criteria based on item distinctiveness indices

Item difficulty index (p)	Item distinctiveness index (r)	
0.90 >	-	- Preferred if there is an effective teaching
0.60-0.90	r>0.20	- A typical good item

0.60-0.90	$r < 0.20$	- An item needs to be worked on
$p < 0.60$	$r > 0.20$	- A difficult but distinctive item (This question is good if you have high standards)
$p < 0.60$	$r < 0.20$	A difficult but not distinctive item (This item cannot be used)

The questions were evaluated according to the above criteria. In this respect, the distribution of 43 questions according to the item difficulty index and item distinctiveness index given in table 6 is shown in Table 7:

Table 7. Item ranking according to item difficulty and distinctiveness index

Item difficulty index (p)	Item distinctiveness index (r)	Comment
0.90 >	-	--
0.60-0.90	$r > 0.20$	1,2,3,4,5,7,8,9,10,14,27,33,39,43
0.60-0.90	$r < 0.20$	--
$p < 0.60$	$r > 0.20$	6,11,12,13,16,19,22,24,25,26,28,29,30,31,32,34,35,36,37,38,40,41,42
$p < 0.60$	$r < 0.20$	15,17,18,20,21,23

When the Table 7 was examined, it was seen that the test was a test with questions of high standards. [44] points out that the range of difficulty indices of test items for achievement tests should generally range from 0.20 to 0.80. The test is in this range, but there are a lot of questions that are difficult and distinctive. Accordingly, the items need to be made easier.

In the item score analysis, item difficulty and distinctiveness were examined for analysis of optional items. After analyzing the results of the item analysis which was coded in the ITEMAN program of the test consisting of 43 items, it was determined that the item correlations of 15 items were low and removed from the test. ITEMAN program has provided us with the following information for each question. In light of this information, final test was created with 28 questions in total

Table 8. Item analysis results and evaluation for test item

Item Statistics					Alternative Statistics					
Seq. No.	Scale-item	Prop. Correct	Disc. index	Point Biser.	Alt.	Prop. Total	Endorsing low	Endorsing High	Point Biser.	Key
1	1-1	.81	.56	.80	A	.03	.09	.00	-.28	
					B	.04	.15	.00	-.32	
					C	.03	.03	.03	-.06	
					D	.81	.41	.97	.80	*

p: item difficulty index (p) D: Item distinctiveness index (r)

Table 8 shows the results of the item analysis of the 1st test item conducted by ITEMAN program. It is a question that upper group was successful and some of the subgroup were successful. The question in the table, as a result of the analysis, is an easy answer ($p = 0.81$) with a good distinctiveness power ($D = 0.56$). Since 1st item is a typical good item, it can be included in the test as it is.

2.2.8. Prognosis of the selected items that generate the statistics of the final test

When the literature is examined, the test-retest reliability expressing consistency and continuity of the scores obtained from a test emphasizes that, it can be determined by the

relationship coefficients between the before and after measurements [7,18,19,20,35,45]. The reliability coefficient is between 0 and 1. As the correlation coefficient approaches 1, it can be said that the scores obtained from the applications are close to each other and closer to 0, respectively. The calculation of the coefficient values as 0.70 and over is interpreted as the measurement tool making stable measurement [9,10]

A test's reliability is found when a test is run twice on the same group at a certain time interval. The time elapsed is a factor that affects the consistency of the test. For this reason, the time must not be too long or too short [20] Reported that ranges of 2-6 weeks are appropriate for this reliability.

The main test statistics such as average, standard deviation, distinctiveness and difficulty indices and reliability coefficient of the achievement test formed after the item analysis and selection of the items to be included in the test are given in Table 9:

Table 9. Final test statistics.

not be indented; subsequent paragraphs should be indented by 5 mm.

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in Table 2 should be used.

NUMBER OF ITEMS IN THE TEST	28	
NUMBER OF ANSWERERS TAKING THE TEST	115	
TEST AVERAGE	20,06	
TEST VARIANCE	44,42	
STANDARD DEVIATION OF TEST	6,66	
HIGHEST SCORE	30,00	
LOWEST SCORE	2,00	
MEDIAN	21,00	
RANGE	28,00	
SKEWNESS	-1,09	Skewed to Left
KURTOSIS	1,26	Acute
RELATIVE VARIABILITY COEFFICIENT	33,22	Kurtic and heterogeneous
KR-20 INTERNAL CONSISTENCY	0,90	High
KR-21 INTERNAL CONSISTENCY	0,90	High
SPLIT HALF CONSISTENCY	0,80	High
STND. ERROR OF AVERAGE	2,08	
AVERAGE DISTINCTIVENESS OF TEST	0.48	Very good
AVERAGE DIFFICULTY OF TEST	0,67	Easy
CRONBACH'S ALPHA VALUE	,901	

When the statistical results of the test are looked at, it is seen that KR-20 Internal Consistency and KR-21 Internal Consistency are 0.90. This shows that the consistency of the test is high. The average distinctiveness of the test is seen as 0.80. This also suggests that the test is distinctive between students at a very good level. Average difficulty of the test is seen as 0,67. This states that when it is desired to measure the effectiveness of the effective teaching method of the test, it should be easy but highly distinctive. When the values of the test are examined, it is seen that the test is easy, highly distinctive.

3. DISCUSSION AND CONCLUSIONS

Many achievement tests have been developed in the literature to measure the achievement of students. In the literature, there are achievement tests especially for the science [2,22,29,30,55] and geography lessons [60]. It was seen that there are very few achievement tests accessed as articles. Not many achievement test articles concerning mathematics were accessed. It is thought that the developed achievement test is a valid and reliable achievement test for mathematics teaching in elemental school.

Achievement tests generally consist of multiple choice questions, and [36] noted that multiple-choice tests are often used to measure student achievement. [42] stated that many questions were asked at the same time and that tests were used as a means of measuring the whole of the desired courses in a short period of time. For this purpose, the achievement test is a reliable measurement tool for making comparisons between classes and students. Developing an achievement test is actually obtaining a valid and reliable test [41]. The developed achievement test is a measurement tool that measures all achievements with sufficient questions, useful for elementary schools with small age groups. Fonthal (2004) reported that schools with more consistent curricula in reading and mathematics areas were more successful in evaluating students. In other words, he emphasized the importance of preparing a question to cover the subjects taught. [34] stated that developing a test that measures all achievements will be more effective in achievement measurement. For this purpose, questions including all the achievements have been prepared.

In short, 431 students who are studying in Samsun city center in 2015-2016 were worked with in the study. Questions including all the achievements given in the table of specifications have been prepared in accordance with the purpose. Expert opinions were obtained for the items prepared for testing and adjustments were made. Test form was prepared and a pilot study was conducted with 31 students to determine how many minutes the test would take. Here, the average of the first and last completing students is taken and an average duration is calculated. After that, 400 students were asked these questions and item analysis were made. Test questions consisting of 43 items were reduced to 28 questions after the item analysis. Then, test-retest application was made for the final results and reliability of the test. The test was applied to 115 students first and then two weeks later, applied again to the same 115 students and the consistency between the two tests was examined. When the statistical results of the test are examined, it is seen that KR-20 Internal Consistency and KR-21 Internal Consistency are 0.90. This result indicates that the test is highly reliable. The average of the test was calculated as 20,06, the variance as 44,42, the standard deviation as 6,66, the median as 21, and the range as 28. Moreover, average difficulty of the test was found as better than the values in the literature with 0,67, and an achievement test with high distinctiveness as 0,48 was obtained.

This study is a unique study showing that KÖYBT has valid and reliable results. Moreover, since the KÖYBT is prepared by following the test development steps, the method can serve as an example for other researches. This work is original in terms of inventory collection and construction and is parallel to the results of other scientific studies.

References

1. Abraham, M.R., Williamson, V.M., Westbrook, S.L. A., *Cross-Age Study of the Understanding Five Concepts*, Journal of Research in Science Teaching, **31**, 2, 147-165 (1994).
2. Akbulut, H. İ., Çepni, S. Bir Üniteye Yönelik Başarı Testi Nasıl Geliştirilir?: İlköğretim 7. Sınıf Kuvvet ve Hareket Ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, **2**(1), 18-44 (2013).
3. Aksu, M., Student performance in dealing with fractions. *The Journal of Educational Research*, **90**(6), 375-380 (1997).
4. Alacaci, C. Öğrencilerin kesirler konusundaki kavram yanılgıları. E. Bingölbali ve M. F. Özmantar (Ed.), *İlköğretimde karşılaşılan matematiksel zorluklar ve çözüm önerileri* (1. Baskı). Ankara: Pegem Akademi Yayınları (2009).

5. Arı, A. Bilişsel Alan Sınıflamasında Yenilenmiş Bloom, SOLO, Fink, Dettmer Taksonomileri ve Uluslararası Alanda Tanınma Durumları. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, **6** (2), 259-290 (2013).
6. Atılgan, H. Eğitimde Ölçme ve Değerlendirme. Hakan Atılgan (Ed.), *Test Geliştirme içinde* (s.315- 348). Ankara: Anı Yayıncılık (2013).
7. Balcı, A. Sosyal Bilimlerde Araştırma Yöntem, Teknik ve İlkeler. Ankara: Pegem A Yayıncılık (2015).
8. Başgün, M., Ersoy, Y. Sayılar ve Aritmetik I: Kesir ve Ondalık Sayıları Öğretilmesinde Bazı Güçlükler ve Yanılgılar, IV. *Fen Bilimleri Eğitimi Kongresi Bildiri Kitabı*: 604-608, MEB Yay., Ankara (2000).
9. Başol, G. *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem A Yayıncılık (2013).
10. Baştürk, S. *Eğitimde Ölçme ve Değerlendirme*. (Edt. Savaş Baştürk). Ölçme Araçlarının Taşınması Gereken Nitelikler (s. 23-53). Ankara: Nobel Yayıncılık (2014).
11. Baykul, Y. *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulanması*. Ankara: ÖSYM (2000).
12. Bayrak, B., Erden, A. M. Fen bilgisi öğretim programının değerlendirilmesi. *Kastamonu Eğitim Dergisi*, **15**(1), 137-154, (2007).
13. Behr, M. J., Lesh, R., Post, T., Silver, E. A. Rational number concepts. In R. Lesh, M. Landau (Eds.), *Acquisitions of mathematics concepts and processes* (pp. 92–126). New York: Academic Press (1983).
14. Bloom, B.S. *Taxonomy of Educational Objectives: The Classification of Educational Goals*, Handbook I: The Cognitive Domain. New York: David Mckay Company Inc (1956).
15. Bloom, B.S., Hastings, J.T., Madaus, G.F. *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw- Hill (1971).
16. Bümen, N.T. Program Geliştirmede Bir Dönüm Noktası: Yenilenmiş Bloom Taksonomisi. *Eğitim ve Bilim*, **31**(142), 3-14 (2006).
17. Büyüköztürk, Ş. *Bilimsel Araştırma Yöntemleri* Ankara: Pegem Akademi (2009).
18. Büyüköztürk, Ş. *Veri analizi El Kitabı*. Ankara: Pegem A Yayıncılık (2011).
19. Can, A. *SPSS ile Bilimsel Araştırma Sürecinde Nicel Veri Analizi*. Ankara: Pegem Yayıncılık (2014).
20. Christensen, L. B., Johnson, R. B., Turner, L. A. Research Methods Desing and Analysis (Çev Edt: Ahmet Alpay). Değişkenlerin Ölçümü ve Örneklem (s.149-180). Ankara: Anı Yayıncılık (2015).
21. Cohen, L., Manion, L., Morrison, K. *Research methods in education*. (6th edition). London: Routledge Falmer (2007).
22. Çalık, M., Ayas, A. Çözümlerde kavram başarı testi hazırlama ve uygulama. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, **2** (14), 1-16 (2003).
23. De Castro, B. Cognitive models: the missing link to learning fraction multiplication and division. *Asia Pacific Education Review*, **9**(2), 101-112 (2008).

24. Erden, M. *Eğitimde Program Değerlendirme*, Ankara: Pegem Yayıncılık (1993).
25. Ertürk, S. *Eğitimde Program Geliştirme*. Ankara: Meteksan (1997).
26. Fonthal, G. *Alignment of State Assessments and Higher Education Expectations: Definition and Utilization of an Alignment Index*. University of California, Irvine (2004).
27. Gay, L. R., Geoffrey E. Mills, Airasian, P. *Educational Research: Competencies for Analysis and Applications*. Boston: Pearson (2009).
28. Gezer, M., Şahin, İ. F., Öner Sünkür, M., Meral, E. 8. Sınıf Türkiye Cumhuriyeti İnkılâp Tarihi ve Atatürkçülük Dersi Öğretim Programı Kazanımlarının Revize Edilmiş Bloom Taksonomisine Göre Değerlendirilmesi. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, **3** (1), 433 – 455 (2014).
29. Gönen, S., Kocakaya, S., Kocakaya, F. Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, **8** (1), 40-57 (2011).
30. Güven, E. Çevre sorunları başarı testinin geliştirilmesi ve öğretmen adaylarının bilgi düzeylerinin belirlenmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, **3**(2), 114-127 (2013).
31. Haser, Ç., Ubuz, B., İlköğretim 5.sınıf Öğrencilerinin Kesirler Konusunda Kavramsal Anlama ve İşlem Yapma Performansı. *IV. Fen Bilimleri Eğitimi Kongresi*, s: 609-612 MEB Yay., Ankara (2001).
32. Işık, C., Kar, T. 7. sınıf öğrencilerinin kesirlerde toplama işlemine kurdukları problemlerin analizi. *İlköğretim Online*, **11**(4), 1021-1035. (2012).
33. İşeri, A., *Diagnosis on Students' Misconceptions on Decimal Numbers* (unpublished master teshiss), ODTÜ, Ankara (1997).
34. Kara, Y., Çepni, S. Investigation the Alignment between School Learning and Entrance Examinations through Item Analysis. *Journal of Baltic Science Education*, **10**(2),73-86 (2011).
35. Karasar, N. *Bilimsel araştırma yöntemleri*. Ankara: Nobel Yayıncılık (2009).
36. Kempa, R. *Assessment in Science*. Cambridge University Press, Cambridge, London (1986).
37. Küçükahmet, L. *Öğretimde Planlama Değerlendirme*. Ankara: Nobel Yayın Dağıtım (2004).
38. Malcolm, P. S. Understanding Rational Numbers. *Mathematics Teacher*. **80**, 518-521 (1987).
39. Metin, M. Kuramdan uygulamaya eğitimde bilimsel araştırma yöntemleri. (Edt). *Nicel Veri Toplama Araçları* (s. 161-2014). Ankara: Pegem A Yayıncılık. (2015).
40. Moss, J., Case, R. Developing children's understanding of the rational numbers: a new model and experimental curriculum. *Journal for Research in Mathematics Education*, **30**(2), 122 -147. (1999).
41. Narlı, S., Başer, N. Küme, Bağntı, Fonksiyon" Konularında Bir Başarı Testi Geliştirme ve Bu başarı testi İle Üniversite Matematik Bölümü 1. Sınıf Öğrencilerinin Bu Konulardaki Hazırbulunuşluklarını Betimleme Üzerine Nicel Bir Araştırma. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, **24**, 147-158 (2008).
42. Ogan Bekiroğlu, F. *Ne kadar Başarılı?*, *Klasik ve Alternatif Ölçme- Değerlendirme Yöntemleri ve Fizikte Uygulamalar* (1. baskı). Ankara: Nobel Yayın Dağıtım. (2004).
43. Olkun, S., Toluk-Uçar, Z. *İlköğretimde etkinlik temelli matematik öğretimi*. Ankara: Anı Yayıncılık (2012).

44. Özçelik, D. A. *Eğitimde Ölçme ve Değerlendirme*, Ankara: ÖSYM Yayınları (1992)
45. Özdamar, K. *Paket Programları İle İstatiksel Veri Analizi –I (MINITAB 16- IBM SPSS 21)*. Eskişehir: Nisan Kitapevi (2013).
46. Özgüven, İ. E. *Bireyi Tanıma Teknikleri*, Ankara: Pdrem Yayınları (1998).
47. Peterson, R.F., Treagust, D.F. Grade-12 Students' Misconceptions of Covalent Bonding and Structure, *Journal of Chemical Education*, **66** (6), 459-460. (1989)
48. Popham, W. J. *Classroom assessment: What teachers need to know*. Boston: Allyn Bacon (1995).
49. Post, T., Fractions and Other National Numbers. *Aritmetic Teacher*. **37**, 3, 28 (1989).
50. Sowder, J.T., Wearne, D. What do we know about eighth-grade student achievement? *Mathematics Teaching in Middle School*, **11**(6), 285-293. (2006).
51. Soylu, Y., Soylu, C. İlköğretim beşinci sınıf öğrencilerinin kesirler konusundaki öğrenme güçlükleri: Kesirlerde sıralama, toplama, çıkarma, çarpma ve kesirlerle ilgili problemler. *Erzincan Eğitim Fakültesi Dergisi*, **7**(2), 101-117 (2005).
52. Soylu, Y. Öğrencilerin kesirler konusundaki hata ve yanlış anlamaları ve sınıf öğretmen adaylarının tahmin edebilme becerileri. *Çağdaş Eğitim Dergisi*, **33**(356), 31-39 (2008).
53. Stafylidou, S., Vosniadou, S. The development of students' understanding of the numerical value of fractions. *Learning and Instruction*, **14**, 503-518. (2004).
54. Sweetland, R. Understanding Multiplication of Fractions, *Arithmetic Teacher*. **32**, 48-52. (1984).
55. Şen, H. C., Eryılmaz, A. Bir başarı testi geliştirme çalışması: Basit elektrik devreleri başarı testi geçerlik ve güvenirlik araştırması. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, **8** (1), 1-39. (2011).
56. Şiap, İ., Duru, A. Kesirlerde geometriksel modelleri kullanabilme becerisi. *Kastamonu Eğitim Dergisi*, **12** (1), 89-96 (2004).
57. Tekin, H. *Eğitimde Ölçme ve Değerlendirme*. Ankara: Yargı Yayınevi (2009).
58. Tiros, D. Enhancing prospective teachers' knowledge of children's conceptions: the case of division of fractions. *Journal for Research in Mathematics Education*, **31**(1), 5-25 (2000).
59. Turgut, M. F., Baykul, Y. *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi (2010).
60. Uzunöz, A., Buldan, İ. ortaöğretim coğrafya dersi doğal sistemler konu alanı atmosfer ve iklim ünitesi başarı testi geliştirme çalışması. *Kastamonu Eğitim Dergisi*, **20** (1), 291-312 (2012).
61. Ünlü, M., Ertekin, E. Why do pre-service teachers pose multiplication problems instead of division problems in fractions? *Procedia - Social and Behavioral Sciences*, **46**, 490-494 (2012).
62. Wearne, D., Kouba, V.L. Rational numbers. In E.A. Silver and P.A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Education Progress* (pp. 163-191). Reston, VA: NCTM (2000).