Education Mind

EDUCATION MIND
2024, VOL 3, NO 3, 389-405
DOI: 10.58583/EM.3.3.8

Research Article

Open Access

Developing an achievement test on length measurement

Gizem Nur Rençber [1], Galip Genç [1]

[1] Department of Elementary Education, Department of Classroom Education, Adnan Menderes University, Aydın, Turkey

ARTICLE HISTORY

Received: 13.08.2024
Accepted: 08.11.2024

KEYWORDS

Length measurement
Achievement test
Validity and reliability

CORRESPONDENCE

Gizem Nur Rençber,
rencbergizemnur09@gmail.com

ABSTRACT

In this study, it was aimed to explain the development process of the achievement test related to the subject of measuring length in the 2nd grade mathematics course. Within the scope of the study, in line with the 2018 Primary School Mathematics Curriculum, the list of achievements related to the subject of measuring length in the second-grade mathematics course was determined and a specification table was created according to the achievements. In line with this table, 25 multiple-choice trial items were created. The items were placed in the test in accordance with the level of difficulty and face validity of the test after receiving expert opinion. A period of 30 minutes was determined for the implementation of the test. The test was applied one-to-one with the think aloud technique to 5 students who had studied the subject of length measurement in the 3rd grade. According to the results obtained in the applications, necessary arrangements were made on the points that were not understood in the items. For the application of the test, 211 students studying in the central primary schools in Efeler district of Aydın province in the second semester of 2021-2022 academic year were determined by random sampling method. The results of the test were analyzed by TAP test analysis program and Tetrachoric Factor analysis. Twenty-two items with an item discrimination index of 0.40 and above were included in the test. However, as a result of the tetrachoric factor analysis, 2 more items were removed from the test due to the fact that the items' Measure of Sampling Adequacy (MSA): value and item loadings were below the expected value. Thus, a final test consisting of 20 multiple-choice items was created to be used in the study. As a result of the reliability analysis of the test, the KR 20 value was found to be 0.85. The answer time of the test was determined as 25 minutes. It was revealed that the test is a valid and reliable test that can determine the achievement of 2nd grade students on the subject of length measurement.

Introduction

Mathematical science and therefore mathematical activities are used in many different areas of human life, whether for daily life or for professional requirements. Mathematical situations are perceived negatively by many people as consisting of calculations that are difficult to do and that

only intelligent people can do mathematics (Akkaş, & Toluk Uçar, 2020). Although there are many different reasons why people perceive mathematics as a negative and impossible science in this way, it is possible to state that some of them are people's attitudes towards mathematics, their anxieties, indirect motivations and indirect learning they have acquired in school life or daily life. Failure to take the necessary precautions in time to prevent mathematics from being perceived as a challenging process at the beginning of the academic year causes people to perceive mathematics as a challenging process. This situation, which may start in the first years of education, may negatively affect the future years of education (Dally et al., 2007). In his 2008 study, Skinner stated that it is important to ensure that individuals have a say in mathematics from the primary school years, to find and eliminate their deficiencies, and to support their success. For this reason, it can be stated that taking measures that will enable individuals to develop positive perceptions towards mathematics from the beginning of their education processes will contribute to the healthy progress of their learning in the mathematics course that they will encounter throughout their education processes.

Measurement and evaluation studies date back to the time when humanity existed, and people knew their environment and themselves. Especially since the mid-18th century, the systematic development of measurement in various fields of science has accelerated and this has contributed to the progress of measurement and science (Çapraz, 2023). When education is defined as the process of achieving behavioral changes in students by exposing them to various experiences, it is possible to state that it is necessary to resort to measurement and evaluation in order to reveal the extent of behavioral changes that occur in students in many different areas of education. Measurement is the objective determination of the number of objects in empirical observations based on a criterion, while evaluation is the work of assigning value to the results obtained from measurement. Measurement and evaluation are an integral part of the teaching process, provided that there are valid and reliable measurements (Başol, 2019). Measurement is used to obtain reliable and valid information in order to reach educational results and make decisions (Baykul, 2021). Measurement in education aims to ensure both the development of students in the educational process and the functioning of teaching in the process and to make inferences about the process of making educational decisions. For this purpose, it is important to determine the purpose of measurement correctly and to choose the appropriate method (Fidan, 2013). In education, the evaluations carried out for the purposes of determining the degree of appropriateness of the methods used in teaching, directing students to areas where they have talents and can be successful, determining learning difficulties and attitudes towards the course, and revealing students' achievements are based on the results obtained from the measurement process (Baykul, 2021). When the literature is examined, it is seen that the measurement tools used to measure achievement are very diverse, and some of them are multiple-choice tests, true-false tests, fill-in-the-blank tests, matching, open-ended tests. However, it is important that the measurement tool to be used for determining success in education has proven validity and reliability in the field of measurement. Here we come across the concepts of validity and reliability, which are frequently used in the field of measurement. Reliability is the ability of the instrument used in measurement to measure in such a way as to give the same result every time it is used, free from errors (Aker et al.,2005). Although reliability in education is generally related to the measurement tool, it is considered in relation to the consistency of the items in the test used with the whole test (Baykul, 2021). The reliability of a test is that it gives similar results to the first application when it is reapplied under the same conditions (Başol, 2019). Validity is the fact that the measurement tool used serves the purpose of use determined while creating it (Aker et al., 2005). Validity is the demonstration of the durability of the comments revealed by the scores obtained from the measurement tool about whether the measurement tool measures the

variable it aims to measure (Özen et al., 2006). While validity is concerned with what will be measured, reliability is concerned with the purity of the measurement process from errors and the rigor of the application (Başol, 2019).

According to Presley et al. (1997), multiple-choice tests are used in Turkey, as in many other countries, to measure students' cognitive status and to place students in higher education institutions as cited in (Akbulut, & Çepni, 2013). Today, most of the tests that teachers use to measure students' achievement in education are multiple-choice tests with unproven validity and reliability. This situation is likely to cause misdirection regarding students' achievement (Balcı, 2019). Turgut & Baykul (2012) includes achievement tests used in education should be valid, reliable and useful in order to provide real results and to be used scientifically (Turgut, & Baykul, 2012). Therefore, it is possible to state that the measurement tools to be used in education should be valid, reliable and useful in order to be of high psychometric quality.

Tests used in education are divided into two as maximum performance and performance tests; achievement and ability tests are included in maximum performance tests; interest and personality tests are included in performance tests (Başol, 2019). Achievement tests among maximum performance tests are used to measure cognitive skills (Şahin et al., 2023). When the literature is examined, it is seen that achievement tests are frequently preferred measurement tools in the field of education (Balcı, 2019; Ersoy, & Bayraktar, 2018; Kurtça, 2016; Yılmaz, & Yılmaz, 2021). The first achievement test used for educational assessment was administered by Meyer Rice in 1895 in the United States (Mehres, & Lehman, 1987). Achievement tests, which have many different purposes, generally aim to determine the achievement level of the student at the end of the applied education program and to find the area or areas in need of support (Kurtça, 2016). The purpose of achievement tests used in the educational process in schools is to be a compass for both teachers and students in determining the level of students' achievement in the subject and in the teaching-learning process (Özçelik, 2010). The use of achievement tests is especially important in terms of determining how much children in primary school have learned the learning outcomes of the relevant course, their learning deficiencies or misconceptions. In the literature, there are achievement tests used in different courses and at different grade levels. As in many subjects, it is seen that achievement tests are used in mathematics courses in primary school period and there are achievement test development studies (Balcı, 2019; Ersoy, & Bayraktar, 2018; Yılmaz, & Yılmaz, 2021).

When the Primary School Mathematics Curriculum used in our country is examined, the acquisitions related to the "Measuring Length" sub-learning area within the measurement learning area are located in the 1st-4th grade (MONE, 2018). The "Measuring Length" sub-learning area is located after the "Measuring Time" sub-learning area in the "Measurement" learning area in the 2nd grade. The subject of "Measuring Length", especially as a continuation of the first grade of primary school, appears as the distinction between non-standard length measurement units and standard measurement units, and also consists of the skills of separating the use of centimeters and meters from standard measurement units, measuring and expressing lengths in centimeters and meters, and solving problems using length measurement units. For this reason, when the literature was reviewed, it was seen that there was no achievement test development study related to the "Measuring Length" area of the mathematics course in the 2nd grade of primary school. It is thought that this study will contribute to eliminating the existing gap in the field. In this study, the process of developing an achievement test on measuring lengths for primary school 2nd graders within the scope of the mathematics curriculum published in 2018 was explained in detail. With this purpose, the research questions are as follows:

Is the test prepared to determine 2nd grade students' achievement on "Measuring Length" valid?

Is the test prepared to determine 2nd grade students' achievement in "Measuring Length" reliable?

Purpose and importance

The measurement area, which is a part of our daily life, appears in many places. Sometimes we measure time, which is the planner of our daily work, sometimes we measure the area for measuring the areas used in the architectural field, and perhaps there is length measurement that we frequently encounter and use, such as time measurement in daily life. In schools, mathematics education, which starts in preschool with counting, is taught in a more systematic program with a spiral structure with the start of primary school. It is seen in the mathematics curriculum that the sub-learning area of length measurement, which starts from the 1st grade of primary school, includes standard length measurement units in the content of the acquisitions in the 2nd grade of primary school and progresses by expanding in a spiral structure at later grade levels (MONE, 2018).

As in all courses, it is important to measure and evaluate students' achievements in the process and in the end result in mathematics. One of the measurement tools that we encounter in measuring students' achievement is achievement tests. The purpose of achievement tests used in the educational process in schools is to be a compass for both teachers and students in determining the level of students in the subject and in the teaching-learning process (Özçelik, 2010). What is important here is that the achievement tests to be used are scientifically valid and reliable. Therefore, the aim of this study is to reveal whether the length measurement achievement test to be developed in the 2nd grade of primary school is valid and reliable.

When the literature was examined, it was seen that there was no achievement test developed related to the sub-learning area of "Length Measurement", one of the sub-learning areas in the "Measurement" learning area in primary school mathematics education. This study is considered important because it will contribute to academic literature in this respect. In addition, since a test form with scientific validity can be made available for classroom teachers who teach 2nd grade primary school classes in order to measure whether the learning outcomes are learned or not, it is considered worth conducting because it will contribute to classroom teachers in terms of the evaluation step, which is one of the stages of the teaching process. The study was also deemed worthwhile for 2nd grade primary school students to evaluate their own learning on the subject and to see their deficiencies.

Method

Research model

The research is a descriptive study since it is an achievement test development study prepared to determine the achievement of 2nd grade students on the subject of "Measuring Length". In this method, the situation, subject, person or events related to the study are defined and described carefully and completely at the maximum level (Büyüköztürk et al., 2020). In this study, it is a descriptive study in which the existing situations of the 2nd grade students are discussed in order to measure their success in the sub-learning area of length measurement. Since the aim of the study was to develop a valid and reliable measurement tool to measure 2nd grade students' achievement in the sub-learning domain of length measurement, Baykul (2021)'s steps on how the test development process should be were followed.

Sample of the study

In order to develop the achievement test, 211 students studying in the 3rd grade in the central primary school in the Efeler district of Aydın province, who have seen the subject of "Measuring Length" in the 2nd grade of primary school in the 2021-2022 academic year. Of these students, 105 were male and 106 were female. The achievement test was determined for 2nd grade students, but since the 2nd graders had not seen the subject at the time of the application and there was a possibility that they might leave the questions blank, the 3rd grade students who had already seen the subject were selected. The simple random sampling method was used because these students were independent of each other and the selection of one for the sample did not depend on the selection of the other.

Data collection tools

Since this study is a valid and reliable achievement test development study, the stages of achievement test development are given under the title of data collection tools.

Test development process:

1. ***Identifying the purpose of the test***

The aim of this study is to develop a valid and reliable achievement test for the subject of "Measuring Length" in the "Measurement" learning domain of the 2nd grade mathematics course "Measurement" based on the primary school mathematics curriculum (2018).

2. ***Determining the characteristics to be measured in the test***

The main feature to be measured in the "Measuring Length" achievement test to be developed in line with the aim of this research is considered as measuring the learning levels and therefore the achievements of 2nd grade primary school students on the subject of "Measuring Length". In addition, the question items in the test to be created are effective in determining the learning deficiencies of the students since they cover all of the acquisitions of the "Measuring Length" subject and which item is related to which acquisition is clearly stated with the prepared specification table.

3. ***Preparation of test items***

During the development phase of the test, a literature review was conducted and the outcomes related to the "Measuring Length" sub-learning area in the elementary school 2nd grade Mathematics Curriculum (2018) were listed. A specification table was prepared to determine the level at which the majority of the determined acquisitions should take place according to Bloom's taxonomy. In the light of the table of specifications, 31 trial items to be included in the test were written by the researchers. The table of specifications of the test items consisting of 31 items is given below.

Table 1 Length measurement achievement test specification table 1

Learning Area (Length)	Cognitive Domain (Bloom)							
	Knowledge	Comprehension	Application	Analyze	Synthesis	Evaluation	Total	Percentage
1. Measures a length using different non-standard units of length measurement and makes repeated measurements with parts of the non-standard unit divided by two and four.	(2nd and 3rd Question)		(Question 30)	(Question 4 and 5)		(Question 1 and 6)	7	%22,55
2.Recognizes standard length measurement units and explains their usage places.	(Question 7)	(Question 8, 9, 10 and 11)					5	%16,12
3.Measures lengths in meters or centimeters using standard tools.		(Question 16.)	(Question 12 and 14)	(Question 13 and 15)			5	%16,12
4.Estimates lengths in units of meters or centimeters and checks the estimate against the measurement result.						(Question 27, 28 and 29)	3	%9,67
5.Construct length models with standard or non-standard units of measurement.			(17th, 18th and 19th Question)				4	%12,9
6. Solves problems using units of length measurement.				(20th, 21st, 22nd, 23rd, 24th, 25th, 26th and 31st Question)			8	%25,80
Total	4	5	7	12	0	4	31	%100
Percentage	%12,9	%16,12	%22,55	%38,7	%0	%12,9	%100	

4. Redaction of articles

In order to ensure the redactions of the items in the test form consisting of 31 items created by the researchers based on the 2018 primary school mathematics curriculum and the literature, the trial items were sent to 1 Turkish language expert, 2 Mathematics experts, 2 Measurement and

Evaluation experts and feedback on validity was received. The test form consisting of 25 items was created by making the necessary arrangements in the trial items with the feedback from the field experts (for reasons such as not fitting the outcome, the question stem not being suitable for the student level, etc.).

Table 2 Length measurement achievement test specification table 2

Learning Area (Length)	Cognitive Domain (Bloom)							
	Knowledge	Comprehension	Application	Analyze	Synthesis	Evaluation	Total	Percentage
1. Measures a length using different non-standard units of length measurement and makes repeated measurements with parts of the non-standard unit divided by two and four.	(2nd and 3rd Question)		Question 21)	(Question 4)		(Question 1)	5	%20
2.Recognizes standard length measurement units and explains their usage places.	(Question 5)	(6th, 7th and 8th Question)					4	%16
3.Measures lengths in meters or centimeters using standard tools.			(Question 9 and 12)	(Question 11 and 13)				%16
4.Estimates lengths in units of meters or centimeters and checks the estimate against the measurement result.						(Question 20 and 24)	2	%8
5.Construct length models with standard or non-standard units of measurement.			(Question 14, 24 and 23)				3	%12
6. Solves problems using units of length measurement.				(10th, 15 th, 16th, 17th, 18th, 19th and 22nd Question)			7	%28
Total	3	3	6	10	0	3	25	%100
Percentage	%12	%12	%24	%40	%0	%12	%100	

Based on the literature and the 2018 primary school mathematics curriculum, acquisitions for the 2nd grade "Measuring Length" subject were determined and a specification table was created as follows for the 25-item test form for which level questions should be weighted according to Bloom's taxonomy.

In line with the specification table, 25 items suitable for the learning outcomes were developed by the researchers.

5. Selection and implementation of trial group

In this study, as a data collection tool, an achievement test consisting of 25 multiple-choice questions on the topic of Measuring Length in Grade 2 was developed by the researchers. This test was applied to 211 students for validity and reliability study.

Data analysis

Within the scope of the validity and reliability study of the achievement test, Test Analysis Program (TAP) and Tetrachoric Factor Analysis were used for data analysis based on Classical Test Theory.

Findings and comments

At this stage of the study, the item difficulty and item discrimination index and reliability results in the Tap program were explained. The item analyses of the achievement test in the TAP program are given in the table below:

Table 3 Item difficulty and discrimination index of the length measurement achievement test 1

Questions	Substance difficulty	Item discrimination index	A	B	C	KR-20
1	0,53	0,09	16	112*	77	
2	0,73	0,43	153*	23	31	
3	0,75	0,62	30	19	158*	
4	0,80	0,36	14	168*	21	
5	0,64	0,69	34	33	136*	
6	0,81	0,37	30	6	171*	
7	0,68	0,67	24	34	143*	
8	0,71	0,44	45	12	149*	
9	0,68	0,33	14	14	143*	
10	0,75	0,58	15	14	158*	
11	0,15	0,13	31*	152	23	
12	0,64	0,52	26	19	134*	
13	0,93	0,13	8	197*	3	
14	0,55	0,39	115*	30	24	0,82
15	0,75	0,56	17	20	158*	
16	0,80	0,51	168*	18	9	
17	0,61	0,56	29	27	129*	
18	0,79	0,56	167*	8	19	
19	0,64	0,64	25	26	134*	
20	0,72	0,58	27	23	152*	
21	0,37	0,49	38	36	79*	
22	0,62	0,61	27	31	130*	
23	0,75	0,62	11	31	159*	
24	0,65	0,45	20	15	138*	
25	0,38	0,14	49	62	80*	

As a result of the analyses made in the Tap analysis program, Question 1, Question 11 and Question 13, which had low item discrimination, were analyzed one by one. As a result of this analysis, since the item discrimination of the last question was on the borderline, it was left in the test by looking at the specification table and question distributions. However, the 22-question test was also analyzed with the factor analysis program and the problematic items were reviewed again.

Factor analysis is applied to reveal the factor structure or to confirm the previously estimated factor structure instead of giving a single coefficient regarding the validity of the measurement tool (Çokluk et al., 2012). Tabachnick & Fidell (2001) includes factor analysis refers to the process of revealing new variables called factorization or common factor or obtaining functional definitions of concepts by using factor loading values of items (Tabachnick, & Fidell, 2001). Factor analysis can be performed in two types: exploratory and confirmatory. Exploratory Factor Analysis (EFA) aims to discover conceptually meaningful and fewer new variables by bringing together a large number of interrelated variables (Çokluk et al., 2012). Various types of EFA are encountered in the literature depending on the difference in the way they are calculated. In this study, exploratory factor analysis was conducted based on the tetrachoric correlation matrix.

"FACTOR 12.01" program was used to calculate the tetrachoric correlation matrix. Prior to factor analysis, Kaiser-Meyer-Olkin (KMO) and Bartlett's Sphericity Test results were examined to test whether the data were suitable for factor analysis.

Within the scope of the research, a tetrachoric correlation matrix was obtained from the data in the form of 1-0 in order to determine the factor structures of the Primary School 2nd Grade Length Measurement Achievement Test Development: Validity and Reliability Study, a tetrachoric correlation matrix was obtained from the 1-0 data and subjected to EFA. First of all, KMO and Barlett Sphericity Test results were examined to determine whether the data were suitable for factorization. If the KMO value is between 0.50-0.60, it is interpreted as poor, between 0.60-0.70 as poor, between 0.70-0.80 as moderate, between 0.80-0.90 as good and above 0.90 as excellent (Leech et al., 2005; Tavşancıl, 2005). The KMO value for the data was found to be approximately 0.83 and Barlett's Test of Sphericity was found to be significant at the .05 level (χ 2=2323.0; sd=231; p=.00). These findings indicate that the relevant assumptions of EFA were met, that is, the data were suitable for factorization. After it was determined that the data were suitable for EFA, EFA was performed on the tetrachoric correlation matrix. One of the most important stages of EFA is to decide on the number of important factors. In this study, the number of factors was determined by examining the eigenvalues, interpreting the eigenvalue graph and interpreting the explained variance ratio. Information about the statistics obtained from the factor analysis based on the tetrachoric correlation matrix and orthogonal rotation is given below.

It is noteworthy that the AGFI value obtained for model fit is 0.94. AGFI is a GFI value corrected for the sample size. When the sample size is particularly large, AGFI is a more representative fit index. The AGFI value is between 0-1. The closer this value is to 1, the better the model fit. It can be said that the obtained AGFI value is within the acceptable range (>.94) (Aksu et al., 2016; Şimşek, 2007).

Although the AGFI value is within the acceptable range, when the factor loadings of the items are examined, the Measure of Sampling Adequacy (MSA) of the 8th and 22nd items: MSA values below .50 indicate that the item measures the same domain as the remaining items in the pool and thus should be removed (Ferrando, & Lorenzo-Seva, 2021). It was also concluded that item 8 should be removed because its factor loading was 0.082 (item 9 in Table 3) and item 22 was 0.197 (item 25 in Table 3). After the factor analysis, Tap program was used for the analysis of the retest items and the analysis was performed. The results of these analyses are given in the table below:

Table 4 Item difficulty and discrimination index of the length measurement achievement test 2

Questions	Substance difficulty	Item discrimination index	A	B	C	KR-20
1	0,73	0,44	153*	23	31	
2	0,75	0,61	30	19	158*	
3	0,80	0,34	14	168*	21	
4	0,64	0,66	34	33	136*	
5	0,81	0,37	30	6	171*	
6	0,68	0,65	24	34	143*	
7	0,71	0,47	45	12	149*	
8	0,75	0,58	15	14	158*	
9	0,64	0,56	26	19	134*	
10	0,55	0,41	115*	30	24	
11	0,75	0,55	17	20	158*	0,85
12	0,80	0,49	168*	18	9	
13	0,61	0,55	29	27	129*	
14	0,79	0,52	167*	8	19	
15	0,64	0,68	25	26	134*	
16	0,72	0,59	27	23	152*	
17	0,37	0,51	38	36	79*	
18	0,62	0,68	27	31	130*	
19	0,75	0,61	11	31	159*	
20	0,65	0,49	20	15	138*	

Other statistical data of the Measuring Length achievement test (20-question final version) are given below:

Table 5 Test statistics

Features	Value
Number of Articles	20
Number of Individuals	211
Minimum	3,00
Maximum	20,00
Mean	13,73
Median	15,00
Standard deviation	4,61
Mean Item Difficulty of the Test	0,68
Mean Test Item discrimination	0,53
Upper group minimum value	18
Subgroup maximum value	11
KR- 20	0,85

Results and discussion

In this study, it was aimed to develop a valid and reliable test to determine the learning levels (achievements) of 2nd grade elementary school students in the sub-learning area of "Length Measurement" within the learning area of "Measurement" in the mathematics course. In order to achieve this goal, the test development stages of Baykul (2021) were followed (Baykul, 2021). In the process of developing the test, after determining the purpose of the test, the scope including the acquisitions of the "Measurement of Length" sub-learning area within the "Measurement" learning area in the 2nd grade mathematics curriculum (2018) was determined and a specification

table was created in line with this scope (Tablo 1 Length Measurement Achievement Test Specification Table 1). Then, an item pool was created in accordance with the specification table. The prepared item pool was presented to the relevant field experts (mathematics, Turkish, measurement and evaluation) together with the specification table and content validity studies were conducted. After the expert opinions, necessary examinations were made on the test items, and it was decided to use a test consisting of 25 items by removing 6 items from the test. The specification table of the 25-item test was reorganized and the number of items used to measure each outcome and which items were used were detailed in the specification table.(Tablo 2 Length Measurement Achievement Test Specification Table 2). Within the scope of the study, the implementation of the trial test was carried out by the researcher under the supervision of the classroom teachers of the relevant grades with 211 third grade primary school students who had similar characteristics to the main group. The trial test was applied to the third-grade students of the primary school since the relevant acquisitions of the second grade students had not yet been processed as of the application period. After the realization of the trial application, the answers were first scored in accordance with the classical test theorem. After the scoring of the answers, item analysis was performed by calculating the item difficulty and item discrimination index values of the items. According to the item analyses, it was seen that items 2,3,4,6,8,10,10,13,15,16,18,20 and 23 were easy, items 1,5,7,9,12,14,17,19,22 and 24 were of medium difficulty, and items 11,21 and 25 were of difficult difficulty. In addition, item analysis showed that items 1, 10 and 11 had very poor discrimination, items 4, 6, 9 and 14 had good discrimination and items 2,3,5,7,8,12,13,15,16,17,18,19,20,21,22,23,24 and 25 had very good discrimination. Since the discrimination index values of items 1, 10 and 11 in the test were lower than 0.19, these items were removed from the test (Tekin, 2017). After the related items were removed from the test, factor analysis and then item analyses were performed again for each item. As a result of the factor analysis, the Measure of Sampling Adequacy (MSA) of the 8th and 22nd items: MSA values below .50 are outside the acceptance limits for measuring the same domain as the remaining items in the pool (Lorenzo-Seva, & Ferrando, 2021). In addition, it was concluded that item 8 should be removed from the test since its factor loading was 0.082 and item 22 had a factor loading of 0.197. As a result of the analysis, it was seen that items 1, 2, 3, 5, 7, 8, 11, 12, 14, 16 and 19 were at the easy difficulty level, items 4, 6, 9, 10, 13, 15, 18 and 20 were at the medium difficulty level and item 17 was at the difficult difficulty level. When each difficulty index of the test items was summed and the arithmetic mean was taken, it was found that the average difficulty index value of the test was 0.68 (medium difficulty). It is desirable that the average difficulty index value of the test is at medium difficulty (Gönen et al., 2011). The item analyses showed that items 3 and 5 had good discrimination and items 1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 had very good discrimination. When the arithmetic mean of the sum of the discrimination index values of each item in the test was taken, it was determined that the average discrimination index value of the test was 0.53 (very good discrimination). In the study, KR-20 method was used to determine the reliability of the test. The KR-20 reliability coefficient value of the test was calculated as 0.85. The KR-20 reliability coefficient value obtained shows that the reliability of the test is sufficient. Tavşancıl (2006) includes the reliability of a test depends on the reliability coefficient value being between 0.60 and 0.80 (Tavşancıl, 2006). After the validity and reliability procedures of the test items were completed, the final form of the test consisting of 20 items was created (EK 1 Length Measurement Achievement Test). The KR-20 reliability coefficient value obtained in the study was found to be 0.78, which was obtained in the study of Ersoy and Bayraktar (2018) in the study of developing an achievement test related to the decimal representation sub-learning area from the 4th grade mathematics course subjects in primary school, and the KR-20 reliability coefficient was found to be 0.78, which revealed that the test was a reliable test, and Balcı (2019) in the study of developing a

valid and reliable achievement test for determining the mathematics course achievement of 3rd and 4th grade students in primary school. and 4th grade students' achievement in mathematics course, and the reliability coefficient of the primary school 3rd and 4th grade tests was 0.78, which showed that the developed tests were valid and reliable tests. grade tests (KR-20) were found to be 0.834 and 0.814, respectively; the reliability coefficient (KR-20) of the primary school 4th grade tests were found to be 0.841 and 0.828, respectively, and Yılmaz &Yılmaz (2021) found that the developed tests were valid and reliable tests. grade mathematics course sub-learning area of natural numbers sub-learning area and the results of the analyses (item discrimination 0.402, item difficulty 0.593, and reliability coefficient 0.797), which revealed that the developed test was a valid and reliable test, were found to support each other.

Acknowledgements

Extended summary of this article was presented as an oral presentation at the International Congress of Integrated Social Research and Interdisciplinary Studies (ISRIS), held from May 30 to June 1, 2024, in Batumi, Georgia.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Akbulut, H. İ., & Çepni, S. (2013). How to develop an achievement test for a unit?: A study for grade 7 force and motion unit. *Amasya University Journal of Faculty of Education*, 2(1), 18-44. https://dergipark.org.tr/tr/download/article-file/19600

Aker, S., Dündar, C., & Pekşen, Y. (2005). Two Vitaly Consept of Measurement instruments: Validity and Reliability. *Ondokuz Mayıs University Medical Journal*, 22(1), 50-60.

Akkaş, E. N. & Toluk Uçar, Z. (2020). Public Opinions on Mathematics.*West Anatolian Journal of Educational Sciences,* 11 (2), 473-491.

Aksu, G., Eser, M.T., & Güzeller, C. O. (2016). *Açımlayıcı ve doğrulayıcı faktör analizi ile yapılan eşitlik modeli uygulamaları [Structural equation model applications with exploratory and confirmatory factor analysis]*. Detay.

Balcı, O. (2019). Developing a mathematics achievement test for primary school 3rd and 4th grade students. (Published master's thesis). Ordu University.

Başol, G. (2019). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. (6. Baskı) [6th Edition]. Pegem Academy Publications. https://ws1.turcademy.com/ww/webviewer.php?doc=76425

Baykul, Y. (2021). Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması [Measurement in education and psychology: Classical test theory and practice]. (4th ed.). Pegem Akademi. https://ws1.turcademy.com/ww/webviewer.php?doc=79856

Büyüköztürk, Ş., Çakmak Kılıç, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2020). *Eğitimde bilimsel araştırma yöntemleri [Scientific research methods in education]*. Pegem Akademi [Pegem Academy].

Çapraz, M. (2023*)*. Developing a mathematics achievement test for 7th grade students**.** (Published Master's Thesis). Erciyes University.

Çokluk, Ö. S., Şekercioğlu, G. ve Büyüköztürk, S. (2012). *Sosyal bilimler için çok değişkenli istatistik: Spss ve Lisrel uygulamaları* [*Multivariate for social sciences statistics: Spss and Lisrel applications]*. Pegem Akademi Yayıncılık [Pegem Academy Publishing].

Dally, E. J., Martens, B. K., Barnett, D., Witt, J. C. & Olson, S.C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement instruction and intensity. *School Psychology Review*, 36, 562-581. https://doi.org/10.1080/02796015.2007.12087918

Ersoy, E., & Bayraktar, G. (2018). Devolopment of Achivemenet Test Related to Sub-Learning of Decimal Projection in Math Class of 4th Grade in Primary School. J*ournal of Buca Faculty of Education*, 46, 240-266. https://dergipark.org.tr/en/download/article-file/610412

Ferrando, P. J., & Lorenzo-Seva, U. (2021). The appropriateness of sum scores as estimates of factor scores in the multiple factor analysis of ordered-categorical responses. *Educational and Psychological Measurement, 81*(2), 205–28. https://doi.org/10.1177/0013164420938108

Fidan, E. (2013). Development of achievement test in the number domain of mathematics course for elementary school students. (Published master's thesis). Ankara University.

Gönen, S., Kocakaya, S., & Kocakaya, F. (2011). A study on developing an achievement test which has reliability and validity on dynamics. *Yüzüncü Yıl University Journal of Faculty of Education*, 13(1), 40-57. Retrieved from https://dergipark.org.tr/tr/pub/yyuefd/issue/13707/165951

Kurtça, V., E. (2016). *The development of mathematics achievement test battery in order to scan first stage of primary school mathematics learning difficulties.* (Published master's thesis). Eskişehir Anadolu University.

Leech, N. L., Barrett, K. C., Morgan, G. A., & Clay, J. N., & Quick, D. (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd ed.). Lawrence Erlbaum Associates Publishers.

MEB (2018). *Matematik dersi öğretim programı (ilkokul ve ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. sınıflar) [Mathematics course curriculum (primary and secondary school 1, 2, 3, 4, 5, 6, 7 and 8. Grades)]*. Ankara: MEB.

Mehrens, W. A. & Lehmann, I. J. (1987). Using standardized tests in education. New York: Longman/Addison Wesley Longman.

Özçelik, D. A. (2010). *Test hazırlama klavuzu (4. Baskı) [Test preparation manual* (4th edition)]. Pegem Akademi.

Özen, Y., Gülaçtı, F., & Kandemir, M. (2006). The problem of validity and reliability in educational research. *Journal of Erzincan Faculty of Education*, 8(1), 69-89.

Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review*, 37, 309-314. https://doi.org/10.1080/02796015.2008.12087878

Şahin, M., Başkurt, İ. & Deringöl, Y. (2023). Development of mathematics problem solving achievement test for primary school 3rd grade students. *Bayburt Education Faculty Journal,* 18 (39), 811-838. https://doi.org/10.35675/befdergi.1211836

Tavşancıl, E. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi* [*Measurement of attitudes and data analysis with SPSS]*. Nobel.

Tekin. H. (2017). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. (25. Baskı) [25th Edition]. Yargı Yayınevi [Yargı Publishing House].

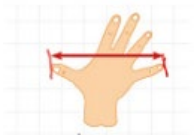
Turgut, M. F., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. Pegem Akademi.

Yılmaz, İ. & Yılmaz, E. (2021). Achievement test development study related to the sub-learning area of natural numbers in fourth grade math. *Kesit Academy Journal*, 7(26), 295-310. https://dergipark.org.tr/en/download/article-file/1751232.

Appendix 1

Length Measurement Achievement Test

1. Which of the following length measurement tools is **a non-standard length measurement tool?**

A) 

B) 

C) 

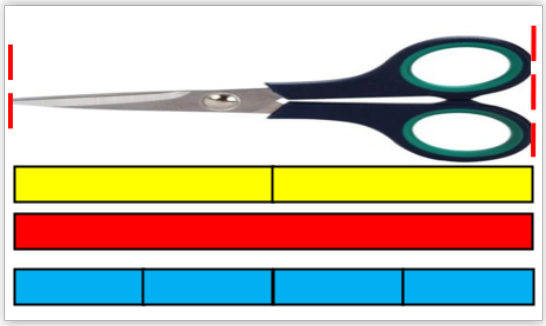
2. Which of the following length measurement tools **is not a non-standard length measurement tool?**

A) 

B) 

C) 

3.

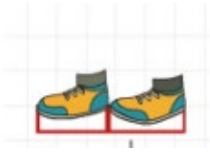


Which of the following is **correct** about the measurement of the scissors given above with colored strips?

A) The scissors are 2 red stripes long.
B) The scissors are 4 blue stripes long.
C) The scissors are 1 yellow strip long.

4. The results of measurements made with non-standard units of length measurement may vary from person to person. Which of the following units of length measurement does not vary **from person to person?**

A) 

B) 

C) 

5. Which of the following **is more appropriate** to measure its length in meters (m)?

A) Notebook
B) Earrings
C) Height of the refrigerator

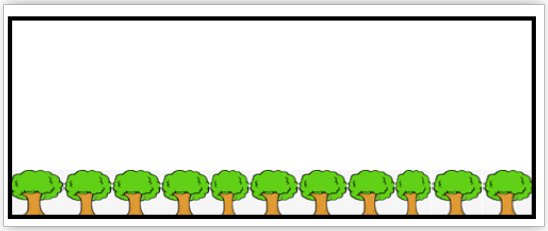
6. Which of the following **is best measured** in centimeters (cm)?

A) The width of a wardrobe
B) Length of garden hose
C) Eraser

7. Measuring the length of which of the following with the centimeter (cm) length unit **is not the first choice?**

A) Paper clip
B) Pen
C) Garden

8.



Uncle Ahmet planted trees on the long side of the garden of his rectangular house. How many meters (m) is the long side of the garden when the distance between two trees is 1 meter (m)?

A) 5 meters (m)
B) 15 meters (m)
C) 10 meters (m)

9.

Nesne	Uzunluk
Kitap	40 cm
Kalemtıraş	5 cm
Okul Çantasının yüksekliği	50 cm

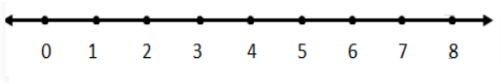
According to the table above, which of the following is the longest object?

A) Book
B) Height of the school bag
C) Sharpener

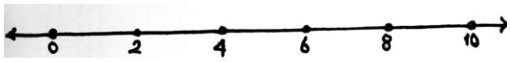
10. Each of the pencil sharpeners below is 1 cm long.

Ali put pencil sharpeners side by side to measure the length of the short side of his notebook. He created a number line model to measure the length of the short side of his notebook with sharpeners of equal length. According to this, which of the following number line models **is appropriate** for the number line model created by Ali?

A)



B)



C)

11.

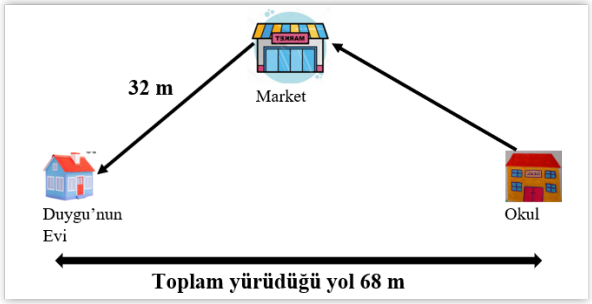
Sevgi adds sticks of 5 centimeters each end to end to form the square on the right. According to this, **how many centimeters is the sum of** the lengths of all sides of the square?

A) 10 cm
B) 40 cm
C) 20 cm

12. Ahmet Bey, a carpenter, used 18 cm of the 50 cm wooden bar he had. How many cm of wooden sticks does the carpenter have left?

A) 32 cm
B) 40 cm
C) 28 cm

13.



The distance between Duygu's house and the market is 32 meters. One day on her way home from school, Duygu stopped by the market to buy bread. When Duygu stops at the supermarket after school and walks home from there, she walks 68 meters. According to this, how many meters is the distance between Duygu's school and the market?

A) 32 m
B) 34 m
C) 36 m

14.

What is the length of the pencil sharpeners when 19 of them are added end to end?

A) 38 cm
B) 30 cm
C) 20 cm

15. The length of Ahmet's pencil is 10 cm shorter than half the length of his bag. According to this, how many cm is the length of Ahmet's pencil?

Length of the bag 50 cm

A) 25cm
B) 20 cm
C) 15 cm

16.

The teacher asked Ayşe, Fatma and Oğuz to first estimate the length of the notebook and then measure it using their rulers. The estimation and measurement results of Ayşe, Fatma and Oğuz are given above. According to this, who made **the closest** estimate of the length of the notebook?

A) Fatma
B) Oguz
C) Ayşe

17. A seller cuts three meters of fabric into quarter-meter pieces and sells each piece for 10 TL. How much money did the seller earn?

A) 60
B) 90
C) 120

18. The poplar tree in the school garden is 73 cm. If this tree grows 5 cm every year, how many cm will it be after 4 years?

A) 78 cm
B) 82 cm
C) 93 cm

19. Kemal added 6 sticks end to end, each of which is 10 cm long. How many cm is this new rod that Kemal created?

A) 36 cm
B) 16 cm
C) 60 cm

20. Ali and his brother Hasan want to find out how many meters long the long side of the dining table is in their house. The difference between Ali's estimate and measurement of the length of the long side of the dining table is 3 meters. His brother Hasan's estimate and measurement of the length of the long side of the dining table differ by 5 meters. Since the length of the long side of the dining table is 2 meters, which of the following **is definitely wrong?**

A) Ali has the closest estimate of the length of the long side of the dining table.
B) The sum of the differences between Ali and his brother Hasan's estimations and measurements of the length of the long side of the dining table is 8 meters.
C) Ali made the farthest estimate of the length of the long side of the dining table.