ACTA DIDACTICA NAPOCENSIA
Volume 15, Number 1, 2022 - DOI: 10.24193/adn.15.1.2

6TH GRADE SCIENCE ABILITY TEST VALIDITY AND RELIABILITY RESEARCH

İclal ALKAN, Nevzat BAYRİ

Abstract: The aim of this research is to develop a test to determine the Science ability of 6th grade students. For this purpose, taking into consideration the 6th grade Science course outcomes; ability test was developed and validity and reliability studies were conducted. In the study, the steps of determining the purpose of the test, determining the properties to be measured, writing the items, reviewing the items, preparing the trial form, applying the trial form, scoring the test, item analysis and selection, and the final form of the test were followed, respectively. The study group of the research consists of 295 students studying in two different schools. According to the findings, item difficulty indexes of the questions in the developed ability test ranged between 0.26 and 0.80, while item discrimination indexes varied between 0.32 and 0.68. Item-total correlations of questions take values between 0.27 and 0.51. While the overall difficulty coefficient of the questions in the test was 0.53, the discrimination index was calculated as 0.47. The KR-20 reliability coefficient calculated for the reliability of the test was 0.84, and the split-half value was 0.71. It can be said that the ability test developed in line with all these results is valid and reliable.

Key words: Science education, ability test, test development

1. Introduction

Education is defined as the process of revealing the desired changes in an individual's behavior (Ertürk, 1984). It is necessary to determine whether it is successful in the process or the level of behavioral change expected from the individual (Erdoğdu & Kurt, 2012). This requirement can be met by measurement and evaluation. For this reason, measurement and evaluation (Ayaydın, 2010), two different concepts, is an indispensable element of the education process (Yeşilyurt, 2012). According to Turgut and Baykul (2010), measurement is defined as "observing the properties of objects and expressing the results with numbers or symbols". Evaluation is the comparison of the value obtained as a result of the measurement with a criterion and reaching a decision (Bahar, Nartgün, Durmuş, & Bıçak, 2012). It is very important to make the measurement process flawlessly in determining whether the goals in the curriculum have been reached or not. In order to make important decisions in education, it was emphasized that teachers (practitioners) should correctly choose the data collection tool they will measure (McDonald, 2002).

Relevant valid and reliable measurement tools are needed in order to determine the realization levels of the goals and behaviors determined in science education and to measure student achievement (Gönen, Kocakaya, & Kocakaya, 2011). Open-ended questions, true-false, short-answer, multiple-choice tests, questionnaires, and two-stage tests related to the measurement of students' achievement are used in all areas of education (Şimşek, 2007). Among the mentioned measurement tools, multiple choice tests are frequently preferred because they provide the opportunity to measure both simple and complex concepts with ease of use (Küçükahmet, 2002; Ogan Bekiroğlu, 2004; Özçelik, 1998). Multiple-choice tests, which are frequently used to reveal students' lack of knowledge and misleading, are easy to apply and score together with well-prepared distractors (Demirci & Efe, 2007; Kan, 2014).

Received: March 2021.

Cite as: Alkan, İ; & Bayri, N. (2022). 6th grade science ability test validity and reliability research. *Acta Didactica Napocensia,* 15(1), 21-29, https://doi.org/10.24193/adn.15.1.2

The relationship between the mental development of individuals and their ages is measured through skill tests (Tezbaşaran, 1994). With these measurement tools, the individual's learning power, general abilities, special abilities or ability to cope with cognitive problems are measured (Atılgan, Kan, & Aydın, 2017). Ability tests; it is examined in three groups as general ability, special ability and different ability tests (Tezbaşaran, 1994; Yıldırım, 1999). The test that is used to measure the mental abilities of individuals is general ability tests. These abilities are believed to be innate and change little under the influence of the environment. It is used to predict what an individual can do in the future and learning power (Yıldırım, 1999). The tests that measure the mind power of an individual in a limited and narrow area are special ability tests (Atılgan et al., 2017). Tests that measure the individual's special abilities in painting, music or art branches are included in this group. Tests measuring different mental abilities such as mechanical comprehension ability, abstract comprehension ability, and language ability are called different ability tests (Özgüven, 1998). Achievement tests, achievement tests, follow-up tests and readiness tests are also collected under this group (Atılgan et al., 2017).

When the literature is examined within the scope of science education research, there are many achievement tests that include multiple choice questions on specific science topics. Şener and Taş (2017) “Let's Solve Our Body's Riddle”, Demir, Kızılay and Bektaş (2016) “Solutions”, Ayvacı and Durmuş (2016) “Heat and Temperature”, Kenan and Özmen (2014) “Particulate Structure of Matter”, Şen and Eryılmaz (2011) developed success tests on the topics of "Simple Electric Circuits", Çakır and Aldemir (2011) "Genetics". In this context, the absence of a measurement tool that can measure the general science ability of the individual makes this study necessary. The fact that the study does not address a specific science topic, but consists of question items that include biology, physics and chemistry, is the most important feature of this measurement tool that distinguishes it from others. Through this study, an easy-to-use and understandable measurement tool will be developed, with validity and reliability studies that will help students gain information about their general science abilities, and facilitation will be provided to researchers and practitioners.

2. Method

In this part of the research, information about the developed measurement tool, the development process and the study group are given.

2. 1. Measuring tool

In the study, a test was developed to determine the Science skills of middle school 6th grade students, and validity and reliability studies were conducted. While developing the Science ability test, the following steps specified by Turgut and Baykul (2010) were followed:

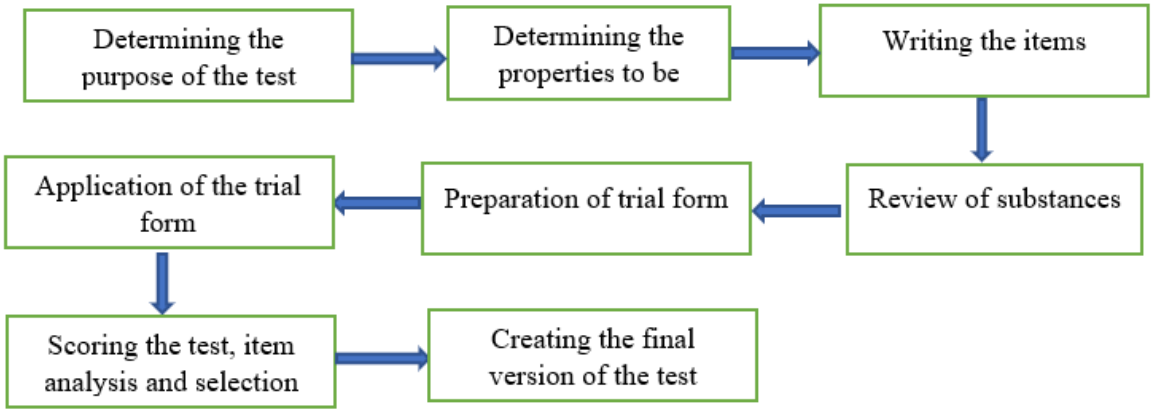

Figure 1.*Test Development Process*

a. In order to determine the purpose of the test and the features to be measured, secondary school science curricula (5th and 6th grades) were examined and a literature review was made taking into account their achievements.

b. In order to determine the scope of the test, a table of specifications was prepared and an aptitude test consisting of 50 items was created.

c. The opinions of 2 faculty members, 1 doctor and 2 Science teachers in the field of Science Education were consulted in order to evaluate the content, appearance validity and comprehensibility of the items and their compatibility with the target group.

d. The pilot studies of the last edited test were conducted with 295 (162 girls, 133 boys) middle school 6th grade students selected through appropriate sampling.

e. The validity and reliability studies of the scores obtained from the test were performed using the Test Analaysis Program (TAP, Version 19.1.4) software.

f. After the pilot application, the final version of the revised test has been reached.

When the related literature is examined, it is seen that there are quite a lot of achievement tests for a single science subject; However, it was observed that there is no test to measure general science ability. It is important that the achievements of the test to be prepared cover all three of the physics, chemistry and biology fields of science. As a result of the interviews with the experts, the achievements planned to be included in the test are as follows:

- "Explains the structures of support and movement system with examples.*"
- "Describes the functions of the structures and organs that make up the digestive system using models.*"
- "He/She concludes that nutrients must undergo physical (mechanical) and chemical digestion in order to pass into the blood."
- "Explains the functions of organs that help digestion."
- "Describes the functions of the structures and organs that make up the circulatory system using a model.*"
- "Examines the large and small blood circulation on a diagram and explains their functions."
- "Defines the structure and functions of blood.*"
- "Refers to the exchange of blood between blood groups.*"
- "Evaluates the importance of blood donation for society."
- "Describes the functions of the structures and organs that make up the respiratory system using models.*"
- "Summarizes their tasks by showing the structures and organs that make up the excretory system on the model.*"
- "Shown the direction, direction and magnitude of the force affecting an object by drawing."
- "Observed that trying more than one force acting on an object.*"
- "Compares balanced and unbalanced forces by observing the motion states of objects.*"
- "Defined speed and expresses its unit.*"
- "Shown that the relation between road, time and speed on the graph."
- "Expressed that the materials are granular, hollow and mobile.*"
- "He/She compares the space between the particles of matter and the mobility of the particles by experimenting with the change of state.*"
- "Density is defined.*"

- "Densities of various substances are calculated as a result of the designed experiments."
- " By experimenting, he/she compares the density of insoluble liquids in each other.*"
- "Comparing the densities of the solid and liquid states of water and discussing the importance of this situation for living beings."
- "Classifies materials in terms of heat conduction.*"
- "Determines the selection criteria for the thermal insulation materials used in buildings.*"
- "Develops alternative thermal insulation materials."
- "Discusses the importance of thermal insulation in buildings in terms of family and country economy and efficient use of resources."
- "Classifies fuels as solid, liquid and gaseous fuels and gives examples of commonly used fuels.*"
- "Discusses the effects of using different types of fuels for heating on human and environment."
- "Investigates and reports the measures to be taken regarding stove and natural gas poisoning."
- "Compares the planets in the solar system with each other.*"
- "Creates a model by sorting the planets in the solar system according to their proximity to the Sun"

(The 18 target achievements included in 34 questions in the Science Ability Test are indicated by *.)

Considering the grade level to which the test will be applied, the ability to answer the questions, the scoring period and the objective scoring, the questions were decided to be multiple choice. While developing the science aptitude test, the distribution of the questions according to the renewed Bloom taxonomy steps was also examined along with the validity and reliability studies. Accordingly, the distribution of 34 questions including 18 gains in taxonomy is as in Table 2:

Table 2. *Science Ability Test Taxonomy Table*

Information size	**The cognitive process dimension**					
	1. Remembering	**2. Understanding**	**3. Applying**	**4. Analyzing**	**5. Evaluating**	**6. Creating**
A. Factual	3, 11, 15, 25	13, 14, 19, 31, 32				
B. Conceptual	17, 24	5, 6, 8, 9, 12, 16, 18, 20, 21, 22, 28, 29, 30, 34				
C. Procedural	4, 26	1, 2, 10, 23, 27	7, 33			
D. Metacognitive						

According to Table 2, it is observed that aptitude test questions show distribution in the stages of recall, comprehension and application. It is very difficult to measure the analysis, evaluation and creation levels in the cognitive process dimension in taxonomy and the metacognitive knowledge level in the knowledge dimension. However, it is possible to measure these levels with projects or homework. Measuring these levels in terms of multiple choice question types, time and quality is not considered appropriate (Karaman, 2005). For this reason, the questions in the aptitude test were limited to the "recall, understanding and application" in the cognitive process dimension, and the "factual knowledge, conceptual knowledge and operational knowledge" steps in the knowledge dimension, which are among the revised Bloom taxonomy steps.

2. 2. Working Group

The pilot application of the science aptitude test was carried out with 295 (Female = 162, Male = 133) secondary school 6th grade students attending two public schools in Malatya province.

2. 3. Data Analysis

The analysis of the data obtained after the pilot application was carried out using the TAP program. Difficulty index, discrimination index and item total correlation coefficients were calculated for the construct validity of the test, while the KR-20 internal consistency coefficient was calculated for reliability.

3. Results

In this part of the research, the analysis results of the data obtained after the pilot application are included. First, the difficulty indexes, discrimination indexes and item total correlation coefficients of the questions in the aptitude test were calculated and it was decided which questions would be excluded from the test (Table 3).

Table 3. *Item Analysis Results of the Questions in the Science Ability Test*

Question	Difficulty Index	Discrimination Index	Item Total Correlation	Question	Difficulty Index	Discrimination Index	Item Total Correlation
01	0,52	0,44	0.34	**26**	0,26	0,35	0.32
02	0,52	0,60	0.47	**27**	0,43	0,34	0.32
03	0,48	0,32	0.27	**28**	0,39	0,38	0.35
04	0,83	0,27	0.34	**29**	0,32	0,30	0.32
05	0,54	0,60	0.50	**30**	0,41	0,39	0.37
06	0,59	0,29	0.29	**31**	0,44	0,23	0.21
07	0,84	0,19	0.27	**32**	0,63	0,69	0.53
08	0,85	0,08	0.17	**33**	0,56	0,67	0.52
09	0,76	0,37	0.36	**34**	0,64	0,59	0.50
10	0,75	0,38	0.38	**35**	0,43	0,18	0.17
11	0,53	0,54	0.45	**36**	0,52	0,54	0.42
12	0,82	0,32	0.34	**37**	0,27	0,22	0.29
13	0,36	0,15	0.17	**38**	0,42	0,57	0.47
14	0,63	0,42	0.35	**39**	0,74	0,34	0.34
15	0,78	0,39	0.42	**40**	0,65	0,37	0.34
16	0,50	0,57	0.43	**41**	0,33	0,16	0.18
17	0,27	0,24	0.29	**42**	0,33	0,45	0.35
18	0,71	0,53	0.45	**43**	0,80	0,35	0.41
19	0,57	0,38	0.31	**44**	0,28	0,30	0.27
20	0,80	0,36	0.38	**45**	0,46	0,49	0.40
21	0,54	0,74	0.56	**46**	0,45	0,51	0.44
22	0,42	0,04	0.02	**47**	0,46	0,48	0.37
23	0,40	0,34	0.28	**48**	0,53	0,50	0.40
24	0,80	0,35	0.37	**49**	0,44	0,60	0.46
25	0,27	-0,04	-0.06	**50**	0,62	0,43	0.38

When Table 3 was examined, it was decided to exclude 4, 6, 7, 8, 12, 13, 17, 21, 22, 25, 31, 35, 37, 39, 41, 43. questions with difficulty index above 0.80 and below 0.20 and item discrimination index below 0.30 from the Science ability test (Tekin, 2011; Turgut & Baykul, 2010). After the questions were removed, the item analysis was repeated. Item difficulty indexes, item discrimination indexes and item total correlations of the questions in the final test, which includes 34 questions, are presented in Table 4.

Table 4. *Item Analysis and Descriptive Analysis Results of the Questions in the Final Test*

Question	Difficulty Index	Discrimination Index	Item Total Correlation
01	0,52	0,48	0.35
02	0,52	0,59	0.51
03	0,48	0,36	0.27
05	0,54	0,60	0.50
09	0,76	0,37	0.36
10	0,75	0,40	0.37
11	0,53	0,60	0.45
14	0,63	0,42	0.35
15	0,78	0,38	0.41
16	0,50	0,60	0.45
18	0,71	0,51	0.44
19	0,57	0,35	0.29
20	0,80	0,34	0.32
23	0,40	0,38	0.32
24	0,80	0,38	0.35
26	0,26	0,33	0.33
27	0,43	0,36	0.34
28	0,39	0,41	0.38
29	0,32	0,36	0.35
30	0,41	0,44	0.41
32	0,63	0,68	0.54
33	0,56	0,64	0.51
34	0,64	0,57	0.49
36	0,52	0,58	0.46
38	0,42	0,57	0.50
40	0,65	0,38	0.34
42	0,33	0,47	0.38
44	0,28	0,32	0.29
45	0,46	0,56	0.43
46	0,45	0,55	0.47
47	0,46	0,45	0.40
48	0,53	0,50	0.41
49	0,44	0,60	0.51
50	0,62	0,50	0.40

Item difficulty indexes of 34 questions in the final test take values between 0.26 and 0.80, while item discrimination indices take values between 0.32 and 0.68. Item-total correlation coefficients of the questions vary between 0.27 and 0.51. The results of the TAP analysis of the Science aptitude test are shown in Table 5.

Table 5. *TAP Analysis Results of the Science Ability Test (Final Test)*

Statistics	Value
N	295
Total Article	34
Average	18.06
Standard Deviation	6.55

Minimum	2
Maximum	34
Skewness	0,363
Kurtosis	-0,571
Average Difficulty	0.53
Average Discrimination Index	0.47
KR-20	0.84
Split-half	0.71

When Table 5 is evaluated, it is observed that the average difficulty value calculated for the Science Ability Test (FBYT), which includes 34 questions, is 0.53, and the discrimination value is 0.47. While the KR-20 reliability coefficient was calculated as 0.84, the split-half coefficient was determined as 0.71.

4. Discussion and Conclusion

It was observed that the validity and reliability levels of the data collected with the science ability test developed in this study were within acceptable ranges. Within the scope of item analysis, item difficulty index and discrimination index for each question in the test were calculated. The average difficulty and distinctiveness of the test were also calculated. By calculating the average difficulty of the data, the development of a test that is too easy or too difficult has been avoided. While the test gets harder with the difficulty coefficient approaching 0; it becomes easier when it approaches 1 (Saraç, 2018). The difficulty coefficients of the questions in the science ability test take values between 0.26 and 0.80. The average difficulty coefficient of the test was calculated as 0.53. In the literature, it is stated that the difficulty of a test being close to 0.50 means a medium level of difficulty and this is a desirable situation (Downing & Haladyna, 2006; Gömleksiz & Erkan, 2010; Tekin, 2011). As a result, it can be said that the questions in the Science aptitude test and the general difficulty coefficient of the test provide the moderate difficulty value stated in the literature and contribute to the construct validity (Downing & Haladyna, 2006; Hingorjo & Jaleel, 2012; Tekin, 2011). It can be said that the developed science ability test provides a good discrimination between low and high achievement students with a medium level of general difficulty (Büyüköztürk, 2012).

By calculating the discrimination indexes of the science aptitude test, it can be distinguished whether the students have the measured feature or not. Downing and Halaydna (2006) emphasized that items with a discrimination index between 0.20-0.29 should be corrected, and items with a value of 0.19 or less should be excluded from the test. In addition, they stated that it is an excellent value for the discrimination index to approach 1. The discrimination indices of the items in the Science Ability Test developed in the research took values between 0.32 and 0.68. It is stated that the discrimination indices of the items in the tests conducted in the classroom environment should be higher than 0.20 (Wells & Wollack, 2003). The average discrimination index of the test was also calculated as 0.47. The discrimination index also contributes to the construct validity of the test, just like the difficulty coefficient. According to Gronlund (1977), questions with a difficulty level of 50 percent and high discrimination should be used in a test.

As a result of the reliability analysis of the Science Ability Test, the KR-20 reliability coefficient was calculated as 0.84 and the split half coefficient as 0.71. Gronlund (1977) stated that each of the methods for calculating the reliability coefficient provides a different type of information. Therefore, the use of KR-20 and split-half coefficients calculated for the aptitude test reliability analysis is considered important in terms of providing two different information. It can be said that the reliability coefficient of the Science Ability Test developed according to the findings obtained is sufficient. The reliability coefficient of the tests applied in the classroom is expected to be 0.50 or 0.60 (Rudner & Schafer, 2002). The reliability coefficient is lower than 0.40, the scores from the test are not reliable, the coefficient is between 0.40 and 0.60, the reliability of the test scores is low, it is between 0.60 and 0.90, the scores obtained from the test are quite reliable and the reliability coefficient is 0.90 and above shows that the scores are highly reliable (Can, 2014).

According to the findings obtained from the research, it can be said that the developed Science Ability Test is valid and reliable. The fact that the questions in the test are multiple-choice, includes 18 acquisitions, and is prepared for the level of 6th grade secondary school students can be expressed as the limitations of the study. Based on this, more comprehensive tests can be prepared for different grade levels in future studies. In addition to multiple choice questions, different question types can be used to measure higher-order thinking skills.

References

Atılgan, H., Kan, A., & Aydın, B. (2017). *Eğitimde ölçme ve değerlendirme*, (Edt. Hakan Atılgan). Ankara: Anı Yayıncılık.

Ayaydın, A. (2010). Desen eğitiminde ölçme ve değerlendirme üzerine bir araştırma. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, 11*(2), 159-172.

Ayvacı, H. Ş., & Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: Isı ve sıcaklık başarı testi geçerlik ve güvenirlik araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, *35*(1), 87-103.

Bahar, M., Nartgün, Z., Durmuş, S. ve Bıçak, B. (2012). *Geleneksel-tamamlayıcı ölçme değerlendirme teknikleri* (5. Baskı). Ankara: Pegem Akademi.

Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı.* Ankara. Pegem A Yayıncılık.

Can, A. (2014). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi (2. Baskı).* Ankara: Pegem A Yayıncılık.

Çakır, M., & Aldemir, B. (2011). İki aşamalı genetik kavramlar tanı testi geliştirme ve geçerlik çalışması/ Developing and validating a two tier mendel genetics diagnostic test. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(16), 335-353.

Demir, N., Kızılay, E., & Bektaş, O. (2016). 7. sınıf çözeltiler konusunda başarı testi geliştirme: geçerlik ve güvenirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *10*(1).

Demirci, N. ve Efe, S. (2007). İlköğretim öğrencilerinin ses konusundaki kavram yanılgılarının belirlenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi.1*(1). 23-56.

Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ: Erlbaum.

Erdoğdu, M. Y. ve Kurt, F. (2012). Öğretmenlerin ölçme ve değerlendirme yeterlik algılarının bazı değişkenler açısından incelenmesi. *Eğitim Bilimleri Dergisi, 1*(2), 23-36.

Ertürk, S. (1984). *Eğitimde program geliştirme*. Yelkentepe Yayınları, Ankara.

Gömleksiz, M. & Erkan, S. (2010). *Eğitimde ölçme ve değerlendirme (2. Baskı).* Ankara: Nobel Yayın Dağıtım.

Gönen, S., Kocakaya, S. ve Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, VIII* (I), 40-57.

Gronlund, N. E. (1977). *Constructing achievement test (2nded.).* New Jersey: Prentice Hall, Inc.

Hingorjo, R. M., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty, discrimination and distractor efficiency, *JPMA-Journal of the Pakistan Medical Association, 62*(2), 142–147.

Kan, A. (2014). *Ölçme aracı geliştirme, eğitimde ölçme ve değerlendirme* (ed. Satılmış Tekindal), Ankara, Pegem A.

Karaman, İ. (2005). Erzurum ilinde bulunan liselerdeki fizik sınav sorularının Bloom taksonomisinin basamaklarına göre analizi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *25*(1).

Kenan, O., & Özmen, H. (2014). Determination of students' alternative conceptions on the particulate nature of matter subject by using. Paper presented at *International Conference on Education in Mathematics, Science and Technology*. Konya, Turkey, 16-18 May 2014.

Küçükahmet, L. (2002). *Öğretimde planlama ve değerlendirme (13. Baskı)*. Ankara: Nobel Yayın Dağıtım

McDonald, P. (2002). Sustaining fertility through public policy: The range of options. *Population*, *57*(3), 417-446.

Ogan Bekiroğlu, F. (2004). *Ne kadar Başarılı? Klasik ve Alternatif Ölçme- Değerlendirme Yöntemleri ve Fizikte Uygulamalar* (1. baskı). Ankara: Nobel Yayın Dağıtım.

Özçelik, D. A. (1998). *Ölçme ve değerlendirme*. Ankara: ÖSYM Yayınları

Özgüven, İ. E. (1994). *Psikolojik testler.* Ankara: Yeni Doğuş Matbaası.

Rudner, L. M., & Shafer, W. D. (2002). *What teachers need to know about assessment.* Washington, DC: National Education Association.

Saraç, H. (2018). Fen bilimleri dersi "Maddenin değişimi" ünitesi ile ilgili başarı testi geliştirme: Geçerlik ve güvenirlik çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *18*(1), 416-445.

Şen, A. G. D. H. C., & Eryılmaz, A. (2011). Bir başarı testi geliştirme çalışması: Basit elektrik devreleri başarı testi geçerlik ve güvenirlik araştırması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *8*(1), 1-39.

Şener, N., & Taş, E. (2017). Developing achievement test: A research for assessment of 5th grade biology subject. *Journal of Education and Learning, 6*(2), 254.

Tekin, H. (2011). *Eğitimde ölçme ve değerlendirme* (5b.). Ankara: Yargı Kitap ve Yayınevi.

Tezbaşaran, A. A. (1994). ÖSYS testlerinde yoklanmak istenen bilişsel davranışlar, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 10,* 79-84.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Ankara: Pegem Akademi.

Wells, C. S., & Wollack, J. A. (2003). *An instructor's guide to understanding test reliability. Testing & Evaluation Services.* University of Wisconsin. Retrieved from http:// https://testing.wisc.edu/Reliability.pdf

Yeşilyurt, E. (2012). Fen ve teknoloji dersinde kullanılan ölçme-değerlendirme yöntemleri ve karşılaşılan güçlükler. *Turkish Studies, 7*(2), 1183-1205.

Yıldırım, C. (1999). *Eğitimde ölçme ve değerlendirme.* Ankara: ÖSYM Yayınları.

Authors

İclal ALKAN, İnönü University, Department of Science Education, Malatya (Türkiye). E-mail: iclal.alkan@inonu.edu.tr

Nevzat BAYRİ, İnönü University, Department of Science Education, Malatya (Türkiye). E-mail: nevzat.bayri@inonu.edu.tr

Acknowledgement

The findings of the research were presented as an oral paper at the 2nd International Congress on Seeking New Perspectives in Education (UYEAK 2019), held on 26-28 October 2019 in İstanbul, Türkiye.

This research was produced from the doctoral thesis supported by the Scientific Research Projects Unit of İnönü University (Project Code: SDK-2019-1734)