Kastamonu Education Journal, 2025, Vol. 33, No:2, 355-371
doi: 10.24106/kefdergi.1683483

| Research Article / Araştırma Makalesi |

Development of a Science Skill-Based Questions Achievement Test on Sound and Its Properties

Ses ve Özellikleri Konusunda Bilimsel Beceri Temelli Başarı Testi Geliştirilmesi

Hayriye Nisa ŞAHİN[1] Esen ERSOY [2], Mustafa ERGUN[3]

Keywords
1.Skill-Based
2.Achievement test
3.Sound
4.Science education

Anahtar Kelimeler
1.Beceri temelli
2.Başarı testi
3.Ses
4.Fen eğitimi

Received/Başvuru Tarihi
13.11.2024

Accepted / Kabul Tarihi
14.04.2025

Abstract

Purpose: This study aimed to develop a skills-based achievement test for the learning outcomes of the subject of "Sound and Its Properties" in the newly implemented science curriculum.

Design: This research used a survey design as a skill-based test development study. The skill-based achievement test developed per the purpose of the study was applied to 375 6th-grade students, and the results were analyzed using the TestAn package program.

Findings: An achievement test consisting of 27 multiple-choice descriptive questions was developed within the scope of this study by examining PISA (Program for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), various achievement tests, resources, and previous studies in order to develop a test consisting of skill-based questions for the new science curriculum. It was determined that 23 of the 28 items in total had an item discrimination index of 0.40 and above; 4 items were between 0.30-0.39. One item, Question 27, was between 0.20-0.29 and removed from the test. The item analysis showed difficulty index values ranging from 0.28 to 0.72. The reliability coefficient of the skill-based achievement test was 0.80.

Highlights: This achievement test is particularly valuable for science teachers preparing students for the LGS (Turkish High School Entrance Exam), as it enables assessing their students' current proficiency levels.

Öz

Çalışmanın amacı: Güncellenen ortaokul fen bilimleri dersi öğretim programı yönelik olarak ses ve özellikleri konusuna uygun beceri temelli sorulardan oluşan akademik başarı testi geliştirmektedir.

Materyal ve Yöntem: Araştırmada beceri temelli test geliştirme çalışması olarak tarama deseni kullanılmıştır. Çalışmanın amacına uygun olarak geliştirilen beceri temelli başarı testi 375 6. sınıf öğrencisine uygulanmış ve sonuçlar TestAn paket programı kullanılarak analiz edilmiştir.

Bulgular: Çalışmada yeni fen bilimleri öğretim programına yönelik beceri temelli sorulardan oluşan bir başarı testi geliştirmek amacıyla PISA, TIMSS, başarı testleri, kaynaklar ve önceki çalışmalar incelenerek 27 adet çoktan seçmeli betimsel başarı testi geliştirilmiştir. Toplam 28 maddenin 23'ünün madde ayırt edicilik indeksinin 0,40 ve üzerinde, 4 maddenin 0,30-0,39 arasında, 1 maddenin ise 0,20-0,29 arasında olduğu belirlenmiştir. 0,20-0,29 arasında olan 27. soru testten çıkarılmıştır. Madde analizi, 0,28 ile 0,72 arasında değişen güçlük indeksi değerleri göstermiştir. Beceri temelli başarı testinin güvenirlik katsayısı 0,80'dir.

Önemli Vurgular: Bu başarı testi özellikle LGS'ye öğrenci hazırlayan fen bilimleri öğretmenleri tarafından öğrencilerinin mevcut durumlarını değerlendirmek için kullanılabilir.

[1] Ondokuz Mayıs University Department of Mathematics and Science Education, Ondokuz Mayıs University, Samsun, Türkiye; https://orcid.org/0009-0005-5523-9427
[2] Ondokuz Mayıs University Department of Mathematics and Science Education, Ondokuz Mayıs University, Samsun, Türkiye; https://orcid.org/0000-0002-7594-8838
[3] **Corresponded Author,** Ondokuz Mayıs University Department of Mathematics and Science Education, Ondokuz Mayıs University, Samsun, Türkiye; https://orcid.org/0000-0003-4471-6601

Citation/Alıntı: Şahin, H. N., Ersoy, E., & Ergun, M. (2025). Development of a Science Skill-Based Questions Achievement Test on Sound and Its Properties, *Kastamonu Education Journal, 33*(2), 355-371. doi: 10.24106/kefdergi.1683483

INTRODUCTION

Education enables the individual to become competent and responsible members of society, equipping them to contribute effectively to the environment and the world around them. A key factor in the effective realization of this process is the individual's capacity to learn. Learning is defined as a dynamic process in which an individual gains knowledge, skills, attitudes, or values through experience, interaction, and education and ensures these achievements become permanent by reinforcing them over time (Khorram et al., 2024). This process also involves the development of the individual in cognitive, affective, and psychomotor domains and is usually shaped by environmental factors, teaching methods, and personal efforts. Learning brings about various changes in students. These changes include transformations in knowledge, skills, attitudes, values, motivation, reasoning skills, and behaviours. Evaluating these changes provides feedback to the student, referred to as the learner, and to the teacher. Drawing from the provided explanations, it becomes evident that rigorous evaluation is essential in education; learning is measured using a diverse array of specialized assessment tools, which are crucial for accurately capturing and understanding every facet of the learning process.

Education uses various measurement tools, which should be designed in accordance with their intended purpose. In other words, the validity and reliability of the measurement tool applied must be established (Gönen et al., 2011). Measurement, which is indispensable in education, provides information about the degree, level, and learning process of the student. It also indicates how changes should be made in the learning process and in which direction the method should evolve. For these reasons, the measurement process uses various measurement methods and techniques. One of these is multiple-choice achievement tests (Çetin & Çakan, 2010). In the Turkish education system, multiple-choice tests are generally preferred when measuring and evaluating students (Akbulut & Çepni, 2013). While there are many measurement and evaluation techniques such as gap filling, true-false, word matching, concept mapping, and short answer questions, the reason for choosing multiple-choice tests is that they are easy, understandable, and useful in terms of both time and application while measuring students' achievements (Karadağ, 2014). Additionally, multiple-choice tests are the most preferred measurement tool for assessing both simple and complex concepts and covering questions about the whole acquisition (Demir, 2012). In this context, beyond the general contribution of multiple-choice tests in measuring knowledge, the use of achievement tests that comprehensively reveal the qualitative gains students acquire during the instructional process supports a holistic assessment approach in education.

Achievement tests are generally a model designed to assess the degree of learning an individual has acquired in a specific content area as a result of planned instruction or training (Nitko & Brookhart, 2011). A review of the literature reveals numerous achievement tests designed for various topics in science courses, including the following: Force and Motion (İdin & Aydoğdu, 2016; Özcan & Orhan, 2024), the Mystery of the Earth's Crust (Sontay & Karamustafaoğlu, 2017), Simple Machines (Özkan & Muştu, 2018), Change of Matter (Saraç, 2018), Matter and Heat (Soylu, Karamustafaoğlu & Karamustafaoğlu, 2020; Avcı, 2020), Cell and Divisions (Karslı, Karamustafaoğlu & Kurt, 2019), the Solar System and Beyond (Uçar & Aktamış, 2019; Armağan & Demir, 2019; Altan, Yadigaroğlu & Eryılmaz Muştu, 2024), Systems (Keçeci, Yıldırım & Zengin, 2019; Akar, Güneş & Yıldırım, 2023; Özkılıç, Bektaş & Karaca, 2023; Meço & Görgülü Arı, 2024), Heat and Temperature (Ayvacı & Durmuş, 2016), Matter and Its Change (Nacaroğlu, Bektaş & Kızkapan, 2020), Force and Energy (Ermiş & Karaman, 2022), Properties of Matter (Dağ & Karamustafaoğlu, 2023), Propagation of Light (Sevim, Uysal & Demirci, 2021), and Human and Environment (Çiçek Şentürk & Selvi, 2021). Although the units included in the science curriculum have been studied separately by different researchers, and achievement tests have been developed accordingly, these tests and the measurement tools prepared are not equivalent to national-level examinations. This is because national-level exams reflect the variations in science questions over time (Sabri, 2019; Pradana et al., 2020; Kolomuc & Karagölge, 2021; Sidiq et al., 2021). Therefore, skill-based questions help students generate solutions and ideas more easily for problems they encounter or may encounter in various aspects of their lives while also providing them with the opportunity to apply their knowledge in everyday life (Rennie & Parker, 1996).

An analysis of national-level exams in the Turkish education system shows that the exam system, along with the types of questions, undergoes frequent changes. The changes in the question system necessitate modifications in both preparatory resources and teachers' approaches to questions and problem-solving skills. To fully understand this situation, it is essential to examine the historical development of the examination system. One of these changes is the high school transition exam. Between 1998 and 2007, the Secondary Education Institutions Selection and Placement Examination (OKS) was evaluated with a single central exam. Between 2008 and 2012, students were assessed with the Placement Examination (SBS), where students took the exam every year in the 6th, 7th, and 8th grades. From 2013 to 2017, the Transition from Basic Education to Secondary Education (TEOG) assessed students with exams in two separate periods. Since 2018, the High School Entrance System (LGS) exam has been used for evaluation.

The purpose of the LGS exam is to measure students' knowledge and skills. The exam consists of skill-based questions similar to the TIMSS (Trends in International Mathematics and Science Study) and PISA (Program for International Student Assessment) exams conducted worldwide (Şan & İlhan, 2022). These new-generation exam questions allow students to apply their knowledge and skills in real-life situations (Karabulut, Tosunbayraktar, & Kariper, 2022). Through skill-based questions, students develop strategic, analytical, and critical thinking skills and abilities in observation, information gathering, comparison, questioning, and discussion.

A review of the literature reveals the opinions of teachers in different branches about the skill-based questions in Turkish, mathematics, and science tests in LGS (Ersoy & Bayraktar, 2018; Erden, 2020; Ünsal & Kaba, 2022), indicating that such questions contribute to science education and should be included in science education (Ceylan & Orhan, 2023; Yuzuak & Recepkethüda 2022). Achievement tests consist of skill-based questions in different disciplines (Erhan, Doğan, & Haser, 2024; Bozkuş & Özgeldi, 2024). Studies examined the theoretical and conceptual framework of science course skill-based questions (Şan & İlhan, 2022). Karabulut, Tosunbayraktar, and Kariper (2022) examined middle school students' views on skill-based science questions and concluded that skill-based questions are difficult and complex. Ercan and Çalışkan (2023) developed a research inquiry and skill-based science test related to the electrical circuit unit, while Turan and Timur (2023) developed a skill-based achievement test for the solar system. Likewise, Büyükkara (2011) and Sözen & Bolat (2014) conducted achievement test studies related to the sound subject and its outcomes. Additionally, Aksoy and Özcan (2020) conducted a test development study to measure 6th-grade students' achievements related to the "Sound and Its Properties" unit.

Within the scope of the science course, the achievement tests on the subject of sound prepared before 2018 included light and sound units together. However, in the science curriculum updated in 2018, the unit "Sound and Its Properties" is included as a separate topic. Sound and its properties are among the concepts that students have difficulty with in science lessons. The curriculum of the science course, which has been gradually started to be used in classes as of the 2024-2025 academic year, reveals that the acquisition of sound and its properties was removed from the sixth grade and transferred to the 8th grade (MoNE, 2024). Considering the present research, it can be stated that the updated science curriculum was prepared in a way that the importance of skill-based questions and the development of the student's skills of observation, information gathering, comparison, questioning, and discussion in the learning process, and making connections in daily life. The new science curriculum can be considered to integrate constructivist, inquiry-based, and STEM-oriented contemporary educational theories while adopting a holistic approach to skill and value development (MoNE, 2024). The program aims to enable students to engage in active learning through skill-based activities, relate scientific knowledge to daily life for meaningful understanding, and consequently enhance their level of scientific literacy (Ak & Köse, 2024). Teaching sound and its properties in science education is crucial for fostering a comprehensive understanding of wave phenomena, which are fundamental to both physics and everyday experiences. Integrating sound into the curriculum can enhance student engagement and motivation, as it connects abstract scientific concepts with tangible, real-world applications, such as music and acoustics (Ramsey, 2015; Iliaki, Velentzas, Michailidi, & Stavrou, 2018). Moreover, addressing common misconceptions and difficulties related to sound through innovative pedagogical strategies, such as active learning and contextualized teaching, can significantly improve students' conceptual understanding and retention of scientific principles (Panjaitan, 2016; Rico, González, Azula, & Aranzábal, 2021; Aygün & Hacıoğlu 2022). The literature shows that students face difficulties in teaching the concept of sound and have misconceptions (Syah, 2025; Azahra et al., 2024; Batlolona & Jamaludin, 2024; Guerra-Reyes et al., 2024; Sözen & Bolat, 2014).

Upon reviewing the literature, it becomes evident that a skill-based achievement test for the unit of "Sound and Its Properties" has yet to be developed, and no existing measurement tool is available for recent acquisitions. Therefore, we decided to conduct a study on developing skill-based questions on sound and its properties. By comparing the 2018 science curriculum with the current curriculum on the topic of "Sound and Its Properties", common achievements were identified, and a test was developed to assess these achievements.

Purpose of the Research

This study aimed to develop a skills-based achievement test for the learning outcomes of the unit of "Sound and Its Properties" in the newly implemented science curriculum.

METHOD

Research model

In this study, a survey design was employed as part of the skill-based test development process. A survey design is a process of systematic data collection to identify specific characteristics, behaviours, opinions, or situations in a population. Surveys, tests, or observations are often used in this design. This method aims to reach generalizations by collecting data from large sample groups and is, therefore, a suitable model for statistical analysis (Gömleksiz & Erkan, 2010). In this research, the development of the achievement test comprising skill-based questions involved crafting multiple-choice items by taking into account the objectives of the "Sound and Its Properties" unit in the science curriculum, aligned with the test's purpose. A pilot study was conducted before finalizing the achievement test, and necessary arrangements were made.

Sample

Fraenkel and Wallen (2013) mentioned two types of populations in research. The first is the accessible population, while the second is the target population. In this study, we preferred the accessible population. Ethics committee permissions were obtained before conducting the study. The study group consists of 375 6th-grade students studying the unit of "Sound and Its Properties" in public schools in a metropolitan city and its districts in the Black Sea Region in the 2023-2024 academic year. The study group was determined through convenience sampling. Since the learning outcome was moved to a different grade level in the updated curriculum, sixth-grade students were chosen for the study, as they had already studied the relevant learning outcomes in their coursework.

Research Instrument and Procedure

In order for a measurement tool to be developed and used in education and psychology, validity, reliability, and usefulness must be ensured. The Standards for Educational and Psychological Testing (SEPT, 1999) was examined, which deals with these standards and steps in detail. According to the data obtained from this review, test developers must follow some steps to develop a measurement tool with the desired criteria and characteristics. The stages to be followed in the test development process can be listed as follows: Test development purpose and general test plan, determining the scope, creating the specification table, writing the articles, reviewing the articles, combining test items into one form, conducting a trial application, selecting or correcting items by item analysis, keeping the tested items in a question bank to be applied later.

Before starting this study, the purpose and planning of the test to be developed were specified through a review of the literature. Next, the scope of the test to be developed was determined, specifically identifying which topics or information should be included in the test. A specification table was created according to the framework determined. Questions were prepared and sent to expert opinions based on the specification table. Necessary adjustments were made in light of the feedback provided by the experts. The test was then collected in a form and piloted as a preliminary application, and the validity and reliability of the items were checked. Following the feedback received from the pilot application, adjustments were made to the developed test. Finally, a skill-based achievement test was created and administered to 375 students to measure their success with the new-generation questions on sound and its properties.

For the achievement test, the achievements related to the subject of "Sound and Its Properties" in the 2024 science curriculum were determined (Table 1).

Table 1: Matching the Skill-Based Achievement Test Questions with the Outcomes of the "Sound and Its Properties" Unit

Item Number	Related Learning Outcome Number
26	8.4.1.1.a. Describes the qualities that make up sound.
6,11,19	8.4.1.1.b. Records the data collected about the formation of sound
4,5,7,8,20,27	8.4.1.1.c. Evaluates that sound is formed as a result of vibration and spreads in waves.
3,12,23,24	8.4.2.1.a. Defines the problem of the effect of distance from the sound source and sound intensity on hearing
13,14,	8.4.2.1.b. Determines the cause-and-effect relationships between the distance from the sound source and the effect of sound intensity on hearing.
9,10,21,25	8.4.2.2.a. Describes the transmission, reflection and absorption of sound
16	8.4.2.2.b. Collects and records data on the transmission, reflection and absorption of sound
1,2,15,17,18,22,28	8.4.2.2.c. Evaluates acoustic applications in relation to the properties of sound

The achievements were examined to ensure the content validity of the achievement test consisting of skill-based questions. The prepared questions were designed by directly aligning them with the learning outcomes. Although the 2024 curriculum consists of various skills, the competencies of the skills employed in the problem-solving process are not evaluated within the scope of this study.

In order to ensure the content validity of the achievement test consisting of skill-based questions, some studies were carried out by examining the gains. A specification table was prepared consisting of the common learning outcomes determined and the questions from each learning outcome (Table 2). 28 questions were prepared. In developing the questions, Bloom's Taxonomy was carefully considered. Bloom's Taxonomy is a well-established framework that categorizes learning objectives into hierarchical levels, systematically guiding students' cognitive processes to achieve these objectives (Bloom, 1956; Anderson & Krathwohl, 2001). This structure forms the foundation for Bloom's Taxonomy's preference in preparing skill-based questions. Bloom's Taxonomy facilitates the acquisition of fundamental concepts at the knowledge and comprehension levels, associating these concepts with experimental processes at the application and analysis levels and developing critical thinking and problem-solving skills at the synthesis and evaluation levels. This hierarchy enables teachers to clearly define the cognitive processes students must engage in when designing questions related to the topic of sound. Consequently, it allows for assessing not only students' ability to recall information but also their capacity to apply, analyze, synthesize, and evaluate knowledge. This approach directly contributes to developing students' scientific process skills, such as observation, data collection, experimentation, and inference. Furthermore, Bloom's Taxonomy provides a methodological basis for developing skill-based questions by ensuring that learning objectives are explicitly and measurably defined in the assessment and evaluation process. The instructional programs published by the Ministry of National Education also emphasize the systematic development of students' cognitive processes, reinforcing the adoption of Bloom's Taxonomy in structuring teaching and assessment processes (MoNE, 2024).

In this study, Bloom's Taxonomy is adopted as an academic and practice-based approach to developing skill-based questions on sound. This preference is due to its advantages in clarifying learning objectives, systematically classifying cognitive processes,

and developing objective and comprehensive assessment tools. Table 2 gives the specification table of the test consisting of 28 trial items created according to Bloom's Taxonomy.

Table 2. Specification Table of Test Items

Learning Outcome	Remember	Understand	Apply	Analyze	Evaluate	Create
8.4.1.1.a.		26				
8.4.1.1.b.		6	19		11	
8.4.1.1.c.	7		20	4,5	27	8
8.4.2.1.a.	3, 12, 24				21	
8.4.2.1.b.					13	
8.4.2.2.a.	9, 23, 25	10				
8.4.2.2.b.	16					
8.4.2.2.c	17	1,2, 18, 22, 28				15

The prepared questions were then submitted to expert opinions. The experts consisted of two associate professors, one assistant professor, a PhD candidate research assistant, as well as two science teachers with a master's degree in science education (Appendix 1). In addition, two Turkish teachers were also consulted for the conformity of the questions to the rules. Based on the feedback of the experts, necessary arrangements were made, and some questions were removed from the achievement test. Additionally, two sample questions approved by the experts are provided in Appendix 2.

After adjustments, the final version of the achievement test was prepared, and the pilot was carried out by applying the test to 35 students. Students were given one class hour (40 minutes) to solve the prepared test. The students could read and answer the questions in the skill-based achievement test within the given time. Based on the students' feedback and the data gathered from the test application, issues such as unclear expressions and spelling errors were identified. Necessary revisions were made to the wording of the items, and the confusion was resolved. After these arrangements, the final version of the skills-based achievement test consisting of 28 items was made ready for the main application.

Procedure

The "Skill-Based Achievement Test on Sound and Its Properties" was used as a data collection tool by obtaining ethics committee permission. The achievement test was applied to 375 6th-grade students in public middle schools in the districts of a metropolitan city in the Black Sea Region.

Data Analysis

The data obtained as a result of the pilot application were coded in the SPSS 20 package program as "1" for correct and "0" for incorrect answers. Then, to ensure the reliability of the test, the Cronbach's Alpha reliability coefficient was calculated. The Cronbach's Alpha reliability coefficient takes values between 0 and 1, and the closer the value is to 1, the more reliable it is, which means that it is adequate in terms of reliability (Ercan & Kan, 2004). Based on the data obtained from the pilot application, the Cronbach's Alpha value was calculated as 0.766, which was sufficient to prove the reliability of the achievement test.

The ranges of item discrimination (Table 3) and item difficulty (Table 4) of a reliable test are given in Tables 3 and 4. TestAn package program was used to calculate item discrimination and item difficulty. Based on the item difficulty index, a ratio approaching +1 indicates that the item is relatively easier, whereas a ratio nearing 0 signifies increased difficulty. When preparing the item difficulty indices of the questions, they are expected to take values close to 0.5 (Turgut, 1995). According to Atılgan (2017), as the item difficulty index approaches 0, it becomes difficult, while it becomes easy towards 1 (Table 4).

Table 3. Item Selection Criteria According to Item Discrimination Indices

Item Discrimination Index	**Item Selection Decision**
0.19 and less	Item to be excluded from the test
between 0.20 and 0.29	Revisable item
between 0.30 and 0.39	Good item can be included in the test
0.40 and higher	Very good item can be included in the test

Table 3 shows the item discrimination indices to be taken as reference for the test questions. According to the discrimination range, it is determined which items should be included in the test, which ones should be removed, and which ones should be revised. A very good item means that the item discrimination index is 0.40 and above. Good items are between 0.30-0.39, and if the value is between 0.20-0.29, it means that the item should be revised. If the value is less than 0.19, the item should be removed from the test (Turgut, 1995).

Table 4. Item Evaluation According to Item Difficulty Indices

Item Difficulty Index	Evaluation of the Item
0.00-0.29	Difficult
0.30-0.49	Medium Difficulty
0.50-0.69	Easy
0.70-1.00	Very Easy

In light of the explanations above, the data were analyzed with the TestAn test analysis program. During the analysis, item discrimination and item difficulty indices of each item were examined separately. Analyses and comments for each item are given in the findings section. In addition, the Kuder-Richardson 20 (KR-20) formula was calculated to find the reliability coefficient of the test, which was found to be 0.77 (r). The reliability coefficient value is within the accepted range according to the values in the literature.

Data Analysis and Interpretation

The "Skill-Based Achievement Test on Sound and Its Properties" prepared was applied to 375 students. The answers given by the students to the multiple-choice questions were entered into the TestAn package program. All answers given by the students to 28 questions were analyzed in this program. The appropriateness of the distractors, the number of correct and incorrect answers, the validity and reliability of each item, and the item difficulty and discrimination indices of the 28 questions were calculated. In line with these data, the items that should be removed from the test and the questions that should be revised were identified.

Validity and Reliability

The data obtained in the pilot application were coded in the SPSS 20 package program as "1" for correct and "0" for incorrect. The Cronbach's Alpha value was found to be 0.766, and the test was determined reliable due to the internal consistency coefficient (Figure 1). The achievement test prepared in the first stage was applied to 375 6th-grade students. TestAn Test Analysis program was used for the analysis of the application result and the Cronbach's Alpha value of the achievement test was found to be 0.80.

Reliability Statistics

Cronbach's Alpha	N of Items
,766	28

Figure 1. Cronbach's Alpha Value of the Test

FINDINGS

Findings Related to the Questions in the Achievement Test

After the pilot study, adjustments were implemented in the achievement test, resulting in the development of a skill-based test comprising 28 questions. As a result of the application, the data of each item were used in the TestAn program, and item discrimination and item difficulty indices were calculated. The findings of item difficulty indices and item discrimination indices for each item are given below.

Table 5 presents the output of the TestAn package program for the first item of the achievement test. For all the items in the tables, p represents the item difficulty index, D denotes the item discrimination index, and * denotes the correct answer.

Table 5. Item Analysis Table for Item 1

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	53	44	24	254	p=0.62
Percentage of responses	14	12	6	68	D=0.46

p: Item difficulty index D: Item discrimination index, *the correct answer.

Table 5 indicates that item 1 demonstrated a very good discrimination power (D=0.46) and an appropriate difficulty level (p=0.62). Consequently, this item was retained in the test.

Table 6. Item Analysis Table for Item 2

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	66	33	264	12	p=0.65
Percentage of responses	18	19	70	38	D=0.50

Table 6 indicates that item 2 was easy (p=0.65) and exhibited very good discrimination power (D=0.50. Therefore, it was included in the test.

Table 7. Item Analysis Table for Item 3

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	21	41	30	283	p=0.69
Percentage of responses	6	11	8	75	D=0.44

Table 7 shows that item 3 was easy (p=0.69) with very good discrimination power (D=0.44, leading to its inclusion in the test.

Table 8. Item Analysis Table for Item 4

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	156	79	34	106	p=0.41
Percentage of responses	42	21	9	28	D=0.38

Table 8 indicates that item 4 had moderate difficulty (p=0.41) and good discrimination power (D=0.38). Therefore, it was included in the test.

Table 9. Item Analysis Table for Item 5

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	33	72	233	37	p=0.55
Percentage of responses	9	19	62	10	D=0.51

Table 9 presents that item 5 had medium difficulty (p=0.55) and very good discrimination power (D=0.51), leading to its inclusion in the test.

Table 10. Item Analysis Table for Item 6

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	273	47	17	38	p=0.67
Percentage of responses	73	12.5	4.5	10	D=0.48

Table 10 shows that item 6 was easy (p=0.67) with very good discrimination power (D=0.48). Hence, it was included in the test.

Table 11. Item Analysis Table for Item 7

Choice	A	B*	C	D	Analysis of the TestAn program
Frequency of Response	45	242	40	48	p=0.60
Percentage of responses	12	64	11	13	D=0.68

Table 11 indicates that item 7 was easy (p=0.60) and had very good discrimination power (D=0.68). Therefore, it was included in the test.

Table 12. Item Analysis Table for Item 8

Choice	A	B*	C	D	Analysis of the TestAn program
Frequency of Response	35	28	272	40	p=0.69
Percentage of responses	9	7	73	11	D=0.55

Table 12 indicates that item 8 was easy (p=0.69) with very good discrimination power (D=0.55). Consequently, item 8 was included in the test.

Table 13. Item Analysis Table for Item 9

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	308	15	30	22	p=0.73
Percentage of responses	82	4	8	6	D=0.41

Table 13 shows that item 9 was easy (p=0.73) with very good discrimination power (D=0.41. Therefore, it was included in the test.

Table 14. Item Analysis Table for Item 10

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	276	26	37	36	p=0.67
Percentage of responses	73	7	10	10	D=0.52

Table 14 indicates that item 10 had a difficulty index of p=0.67 and a very good discrimination power of D=0.52. It was hence included in the test.

Table 15. Item Analysis Table for Item 11

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	57	25	140	153	p=0.37
Percentage of responses	15	7	37	41	D=0.31

Based on Table 15, item 11, with an item difficulty index of p=0.37 and discrimination power of D=0.31, was of medium difficulty and had high discrimination, leading to its inclusion in the test.

Table 16. Item Analysis Table for Item 12

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	294	15	28	38	p=0.67
Percentage of responses	78	4	8	10	D=0.59

Table 16 above indicates that item 12 was easy (p=0.67) and had very good discrimination power (D=0.59). Therefore, it was included in the test.

Table 17. Item Analysis Table for Item 13

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	58	25	96	196	p=0.49
Percentage of responses	15	7	26	52	D=0.49

Based on the analysis in Table 17, item 13 was identified as having moderate difficulty (p=0.49) and very good discrimination power (D=0.49), leading to its inclusion in the test.

Table 18. Item Analysis Table for Item 14

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	146	138	43	48	p=0.37
Percentage of responses	39	37	11	13	D=0.35

Based on the analysis in Table 18, item 14 was found to be of medium difficulty (p=0.37) with good discrimination power (D=0.35), leading to its inclusion in the test.

Table 19. Item Analysis Table for Item 15

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	120	63	141	51	p=0.44
Percentage of responses	32	17	37	14	D=0.42

Table 19 above shows that item 15 was easy according to the item difficulty index (p=0.44) and had very good discrimination power (D=0.42). Therefore, it was included in the test.

Table 20. Item Analysis Table for Item 16

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	277	27	46	25	p=0.67
Percentage of responses	74	7	12	7	D=0.63

Table 20 shows that item 16 was easy (p=0.67) with very good discrimination power (D=0.63), leading to its inclusion in the test.

Table 21. Item Analysis Table for Item 17

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	62	34	68	211	p=0.52
Percentage of responses	17	9	18	56	D=0.59

Table 21 indicates that item 17, with an item difficulty index of p=0.52 and discrimination power of D=0.59, was easy and highly discriminative, leading to its inclusion in the test.

Table 22. Item Analysis Table for Item 18

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	64	70	50	191	p=0.49
Percentage of responses	17	19	13	51	D=0.59

Table 22 shows that item 18 had moderate difficulty (p=0.49) and excellent discrimination power (D=0.59), leading to its inclusion in the test.

Table 23. Item Analysis Table for Item 19

Choice	A	B*	C	D	Analysis of the TestAn program
Frequency of Response	38	252	29	56	p=0.61
Percentage of responses	10	67	8	15	D=0.49

Table 23 indicates that item 19, with a difficulty index of p=0.61 and discrimination power of D=0.49, was easy and highly discriminative, leading to its inclusion in the test.

Table 24. Item Analysis Table for Item 20

Choice	A*	B	C	D	Analysis of the TestAn program
Frequency of Response	254	50	41	30	p=0.67
Percentage of responses	68	13	11	8	D=0.46

Table 24 indicates that item 20 was easy (p=0.67) with very good discrimination power (D=0.46), leading to its inclusion in the test.

Table 25. Item Analysis Table for Item 21

Choice	A	B*	C	D	Analysis of the TestAn program
Frequency of Response	64	183	89	39	p=0.52
Percentage of responses	17	49	24	10	D=0.40

Table 25 reveals that item 21, with a difficulty index of p=0.52 and discrimination power of D=0.40, was both easy and highly discriminative, leading to its inclusion in the test.

Table 26. Item Analysis Table for Item 22

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	78	55	186	56	p=0.49
Percentage of responses	21	14	50	15	D=0.65

Table 26 indicates that item 22 had moderate difficulty (p=0.49) and very good discrimination power (D=0.65), making it a suitable item for inclusion in the test.

Table 27. Item Analysis Table for Item 23

Choice	A	B*	C	D	Analysis of the TestAn program
Frequency of Response	42	227	57	49	p=0.58
Percentage of responses	11	61	15	13	D=0.65

Table 27 indicates that item 23, with a difficulty index of p=0.58 and discrimination power of D=0.65, was both easy and highly discriminative, thus included in the test.

Table 28. Item Analysis Table for Item 24

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	67	90	177	41	p=0.47
Percentage of responses	18	24	47	11	D=0.47

Table 28 shows that item 24, with moderate difficulty (p=0.47) and very good discrimination power (D=0.47), was suitable to be included in the test.

Table 29. Item Analysis Table Item 25

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	63	77	31	204	p=0.54
Percentage of responses	17	21	8	54	D=0.72

Table 29 shows that item 25, with a difficulty index of p=0.54 and discrimination power of D=0.72, was easy and highly discriminative, leading to its inclusion in the test.

Table 30. Item Analysis Table for Item 26

Choice	A	B	C	D*	Analysis of the TestAn program
Frequency of Response	201	93	41	40	p=0.52
Percentage of responses	53	25	11	11	D=0.40

Table 30 reveals that item 26 was easy (p=0.52) with good discrimination power (D=0.46) and thus included in the test.

Table 31. Item Analysis Table for Item 27

Choice	A	B	C*	D	Analysis of the TestAn program
Frequency of Response	77	80	174	44	p=0.46
Percentage of responses	21	21	46	12	D=0.28

According to Table 31, item 27 exhibited medium difficulty (p=0.46) and a discrimination index of 0.28, indicating the need for revision. Consequently, item 27 was excluded from the final test.

Table 32. Item Analysis Table for Item 28

Choice	A	B	C*	D*	Analysis of the TestAn program
Frequency of Response	38	45	228	63	p=0.61
Percentage of responses	10	12	61	17	D=0.36

Table 32 shows that item 28 was easy (p=0.61) with good discrimination power (D=0.36) and thus included in the test.

Table 33 summarizes the difficulty and discrimination indices of the 28 test items.

Table 33. Item Difficulty and Item Discrimination of Test Items

Item No	p	D	Item No	P	D
1	.62	.46	15	.44	.42
2	.65	.50	16	.67	.63
3	.69	.44	17	.52	.59
4	.41	.38	18	.49	.59
5	.55	.51	19	.61	.49
6	.67	.48	20	.67	.46
7	.60	.68	21	.52	.40
8	.69	.55	22	.49	.65
9	.73	.41	23	.58	.65
10	.67	.51	24	.47	.47
11	.37	.31	25	.54	.72
12	.67	.59	26	.52	.40
13	.49	.49	27	.46	.28*
14	.37	.35	28	.61	.36

Accordingly, the item marked with * was removed from the final test due to its discrimination index falling between 0.20-0.29.

Table 34 shows the evaluation of the 28 questions in the achievement test.

Table 34. Evaluation of Items According to the Item Discrimination Index

Item Discrimination Index	Evaluation of Item
0.40 and bigger	1.2.3,5,6,7,8,9,10,12,13,15,16,17,18,19,20,21,22,23,24,25,26
0.30- 0.39	4,11,14,28
0.20- 029	27
0.19 and less	-

Table 34 reveals that no item had an item discrimination index of 0.19 and below; item 27 was in the range of 0.20-0.29 per the item discrimination index; therefore, the item needs to be revised or removed. Hence, item 27 was removed. The remaining questions were appropriate in terms of item discrimination.

DISCUSSION AND CONCLUSION

In the study, a total of 27 multiple-choice descriptive achievement test questions were developed through an examination of PISA, TIMSS, various achievement tests, relevant resources, and previous studies. The aim was to construct an achievement test comprising skill-based questions for the 8th grade "Sound and Its Properties" subject as delineated in the renewed science curriculum (MoNE, 2024). This section includes a discussion of the findings of the developed test in relation to the existing literature.

A review of the literature on the development of an achievement test with validity and reliability revealed studies that followed similar steps (Avcı, 2019; Sontay & Karamustafaoğlu, 2020; Sarıçetin, 2021; Aksoy & Özcan, 2020; Kurt, Aydın, & Bekereci 2023). The similarities between these studies and the current one include the creation of a question pool, seeking expert opinions, and revising or removing items as needed to ensure validity and reliability. The differences include that the current study was conducted for the renewed 2024 science curriculum in the subject area of sound and its properties. That being said, the most significant difference is that our study consists of skill-based questions.

An examination of the literature on skill-based questions indicates that the study by Turan and Timur (2023) on the Solar System unit involved the development of a question pool, the revision of selected items through validity and reliability assessments, the elimination of certain questions, and the solicitation of expert opinions—all aimed at ensuring the questions were skill-based. In contrast, the study was conducted within the framework of the renewed 2024 science curriculum and specifically addressed the domain of sound and its properties.

Within the scope of the theoretical and conceptual framework for skill-based questions in science, it is important to answer whether students have attitudes, prejudices, and expectations toward these questions (Şan & İlhan, 2022). In this respect, our study responds to the current expectations and needs. Furthermore, the present study may serve as a model by emphasizing the significance of language use and delineating essential elements in the construction of skill-based questions. These elements include the development of such questions, the definition of the scope, the establishment of theoretical and conceptual frameworks, the design of questions that meet students' expectations, and the formulation of questions in a clear and concise manner rather than in an overly lengthy or complex form (Şan & İlhan, 2022).. In this study, questions were developed by establishing the theoretical and conceptual framework of the unit of "Sound and Its Properties", which is part of the 8th-grade outcomes of the revised science curriculum.

As part of the study, a skill-based achievement test focusing on the subject of "Sound and Its Properties" was created. In line with this purpose, the achievement test consists of questions whose validity and reliability have been ensured. While developing this achievement test consisting of skill-based questions, the followed the steps suggested by Crocker and Algina (2006) as test development steps. In the implementation steps, the purpose and framework of the subject were first determined. In line with the determined purpose, specification tables were prepared for the content validity of the subject. Then, an item pool consisting of 28 questions was created for the questions of the test. While preparing the questions in the item pool, international TIMSS and PISA exams, as well as national LGS questions, were used, besides studies and resources on this subject previously available in the literature. The literature includes studies where the item pool was prepared using a similar approach (Özcan & Orhan, 2024; Saylan Kırmızıgül & Kaya, 2019). Expert opinions were taken for each question per the specification table prepared. With the feedback received from the experts, the questions became ready. The pilot application, which included the pre-application of the 28-question skill-based achievement test created following the expert opinions, was carried out. The test was applied to 35 students studying in the 6th grade for the pilot application. As a result of the application, issues such as spelling mistakes, incorrect word usage, Turkish sentence disorders, inaccuracies, and insufficient time were corrected. Finally, the validity and reliability of the test were examined by analyzing the items in the pilot application, and the Cronbach's Alpha value was found to be 0.766, indicating that the test was valid and reliable.

The skill-based achievement test developed was applied to 375 6th grade students, and the results were analyzed using the TestAn package program. Each item was analyzed separately in detail. As a result of the analyses, the item discrimination and item

difficulty indices for the items were calculated. Based on this analysis, questions that did not meet the required item index and difficulty standards were removed, and the final set of test questions was selected.

Item analysis is very useful in calculating the success criteria of students as well as developing a measurement tool (Bichi, 2015). The item analysis of the achievement test consisting of skill-based questions on the subject of "Sound and Its Properties" revealed that 23 of the 28 items had an item discrimination index of 0.40 and above, four items were between 0.30-0.39, while one item was between 0.20-0.29. Question 27, which was between 0.20-0.29, was removed from the test. According to extant literature, an item discrimination index of 0.4 or higher is deemed acceptable (Keçeci, Yıldırım, & Zengin, 2019; Ayvacı & Durmuş, 2016). Consequently, the questions in the achievement test are structured in a way that effectively distinguishes between students.

The item analysis results showed that the item difficulty index values of the items in the test were between 0.28 and 0.72. The average item difficulty index value of the achievement test consisting of skill-based questions developed according to the TestAn results was found to be 0.6. An item difficulty index of 0.6 indicates that it is a medium difficulty achievement test.

The reliability coefficient of the "Skill-based Achievement Test on Sound and Its Properties" was found to be 0.80. The literature on achievement test development includes studies with similar results (Özkılıç, Bektas, & Karaca, 2023; Dağ & Karamustafaoğlu, 2023; Boz, Bostan Sarıoğlan, & Özcan, 2022). On the other hand, an item must be greater than 0.70 to be reliable. Therefore, as a result of the analysis, the KR-20 value was determined as 0.8. Since this value is greater than 0.7, the developed test was deemed reliable.

A specification table and item pool were prepared for the content validity of the study. Expert opinions were taken. In addition to the opinions of experts in the field of science education, the opinions of teachers working in the field were also taken. Furthermore, the opinions of Turkish teachers were also taken in order to avoid sentence structure and expression errors. The validity of the test was ensured with the expert opinions. This study aligns with extant research in the literature, wherein expert opinions were solicited during the development of achievement tests in science education. Such an approach underscores the importance of expert input in enhancing the validity and reliability of these assessments. (Meço & Görgülü Arı, 2024; Çiçek Şentürk & Selvi, 2021).

RECOMMENDATIONS

In conclusion, a valid and reliable measurement tool was developed through the achievement test development study, specifically designed for skill-based questions at the middle school level. To ensure continuous improvement, it is recommended to regularly evaluate and update the tool, incorporating feedback to refine its content and format. Additionally, support materials and guides with sample questions should be provided to help students prepare for these skill-based questions. Performance analysis of test results is crucial to identify students' strengths and weaknesses, guiding instructional strategies and curriculum enhancements. Establishing feedback mechanisms with teachers and students will help in continuously refining the tool and the testing process. Research should be conducted to assess the tool's effectiveness and its impact on students' academic performance, ensuring inclusivity across different student groups. These efforts will contribute to a more accurate and comprehensive evaluation of students' skills.

It is important to note that this research was conducted exclusively with students from a specific region and within a limited time frame. Additionally, uncontrollable variables such as individual differences among participants, varying levels of motivation, and external factors could influence the results. Furthermore, the cultural and social backgrounds of the students may also impact the outcomes. In light of the findings obtained from this study, test development studies can be carried out with different samples by taking into account the limitations mentioned.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, author-ship, and/or publication of this article.

Statements of publication ethics

We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully.

Researchers' contribution rate

The first and third authors contributed to writing the scale items, formulating the problem statement of the article, and collecting the data. The first and second authors performed the data analysis, while all of the authors collaborated on writing the discussion and recommendations section. The study was conducted and reported with equal collaboration of the researchers.

Ethics Committee Approval Information

Ethical committee approval for this study was obtained from the Ethics Committee of Ondokuz Mayıs University (Number of Decisions:2024-569; Date: 31.05.2024.

REFERENCES

Ak, B. S. ve Köse, M. (2024). 2024 Fen bilimleri dersi öğretim programı hakkında öğretmen görüşlerinin incelenmesi. *Akademik Platform Eğitim ve Değişim Dergisi*, 7(2), 132-169. https://doi.org/10.55150/apjec.1582677

Akar, B., Güneş, İ., & Yıldırım, İ. (2023). Yenilenmiş Bloom taksonomisine göre vücudumuzdaki sistemler başarı testi geliştirme çalışması. *Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi, 7*(12), 161-183. https://doi.org/10.56677/mkuefder.1369107

Akbulut, H. İ., & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir?: İlköğretim 7. sınıf kuvvet ve hareket ünitesine yönelik bir çalışma. *Amasya Üniversitesi Eğitim Fakültesi Dergisi, 2*(1), 18-44.

Aksoy, Ş., & Özcan, H. (2020). Altıncı sınıf öğrencilerinin ses ve özellikleri ünitesi ile ilgili başarılarını ölçmeye yönelik bir test geliştirme çalışması. *Eğitimde Kuram ve Uygulama, 16*(2), 193-214. https://doi.org/10.17244/eku.787792

Aktamış, H., & Uçar, R. (2019). Astronomi'ye yönelik tutum ölçeği ve 7. sınıf "Güneş Sistemi ve Ötesi" ünitesine yönelik başarı testi geliştirme çalışması. *Batı Anadolu Eğitim Bilimleri Dergisi*, 10(1), 57-78.

Altan, T., Yadigaroğlu, M., & Eryılmaz Muştu, Ö. (2024). Bir başarı testi geliştirme çalışması: 6.sınıf Güneş sistemi ve tutulmalar ünitesi. *Fen Bilimleri Öğretimi Dergisi, 12*(1), 143-192. https://doi.org/10.56423/fbod.1401126

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision ofBloom's taxonomy of educational objectives: complete edition.* Addison Wesley Longman.

Armağan, F. Ö. & Demir, N. (2019). Astronomi başarı testi geliştirme: Geçerlik ve güvenirlik çalışması. *Maarif Mektepleri Uluslararası Eğitim Bilimleri Dergisi, 3*(1), 52-70. https://doi.org/10.46762/mamulebd.582883

Atılgan, H. (2017). *Eğitimde ölçme ve değerlendirme* (10. baskı). Anı yayıncılık

Avcı, F. (2020). Yenilenmiş Bloom taksonomisine göre madde ve ısı başarı testi: geçerlik güvenirlik çalışması. *Iğdır Üniversitesi Sosyal Bilimler Dergisi, 21*, 263-292.

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

Aygün, M., & Hacıoğlu, Y. (2022). Teaching the Sound Concept: A Review of Science and Physics Education Postgraduate Theses in Turkey. *Athens Journal of Education, 9*(2), 257-275.

Azahra, H., Pasaribu, M., Haeruddin, H., Zaky, M., Supriyadi, S., & Worku, H. (2024). Identification of misconceptions on the material of sound waves using the four tier diagnostic test in Tolitoli regency. *EduFisika: Jurnal Pendidikan Fisika*, 9(2), 228-236. https://doi.org/10.59052/edufisika.v9i2.36839

Batlolona, J. R., & Jamaludin, J. (2024). Students' misconceptions on the concept of sound: a case study about Marinyo, Tanimbar Islands. *Journal of Education and Learning (EduLearn)*, 18(3), 681-689. https://doi.org/10.11591/edulearn.v18i3.21135

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green

Boz, C., Bostan Sarıoğlan, A., & Özcan, H. (2022). Kimyasal bağlar konusunda başarı testi geliştirme çalışması. *Manisa Celal Bayar Üniversitesi Eğitim Fakültesi Dergisi, 10*(2), 149-166. https://doi.org/10.52826/mcbuefd.1065432

Bozkuş, Ç., & Özgeldi, M. (2024). Analysis of middle school mathematics teachers' evaluations of skill-based questions within the framework of mathematical competencies. *Uluslararası Türk Eğitim Bilimleri Dergisi*, *12*(1), 535-583. https://doi.org/10.46778/goputeb.1425323

Büyükkara, S. (2011). İlköğretim 8. sınıf fen ve teknoloji dersi ses ünitesinin bilgisayar simülasyonları ve animasyonları ile öğretiminin öğrenci başarısı ve tutumu üzerine etkisi. [Yayınlanmamış Yüksek Lisans Tezi]. Selçuk Üniversitesi.

Çetin, M. O., & Çakan, M. (2010). Fen ve teknoloji dersi başarılarının farklı yaklaşımlarla ölçülmesi ve bu yaklaşımlara ilişkin öğrenci görüşleri. *Journal of Measurement and Evaluation in Education and Psychology, 1*(2), 93-99.

Ceylan, G., & Orhan, A. T. (2023). Beceri Temelli Fen Sorularına Yönelik Fen Bilimleri Öğretmenlerinin Görüşlerinin İncelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 19*(1), 84-111. https://doi.org/10.17860/mersinefd.1193968

Çiçek Şentürk, Ö., & Selvi, M. (2021). Fen Bilimleri dersi "İnsan ve Çevre" ünitesi akademik başarı testi geliştirme: güvenirlik ve geçerlik çalışması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 41*(2), 601-630. https://doi.org/10.17152/gefad.940400

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Thomson Learning.

Dağ, M. & Karamustafaoğlu, S. (2023). "Maddenin Özellikleri" ünitesi: başarı testi geliştirme ve öğrenci başarısını belirleme. *Milli Eğitim Dergisi*, 52 (237), 221-254. https://doi.org/10.37669/milliegitim.1076053

Demir, M. K. (2012). The study of primary school teacher candidates' opinions towards midterm and final examinations. *Ondokuz Mayıs Universitesi Eğitim Fakültesi Dergisi*, 31(2), 193-21

Ercan, İ., & Kan, İ. (2004). Ölçeklerde güvenirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi, 30*(3), 211-216.

Ercan, M., & Çalışkan, İ. (2023). Araştırma-sorgulama ve beceri temelli fen bilimleri testi geliştirme çalışması-elektrik devreleri ünitesi örneği. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (58), 2922-2947. https://doi.org/10.53444/deubefd.1336581

Erden, B. (2020). Türkçe, matematik ve fen bilimleri dersi beceri temelli sorularına ilişkin öğretmen görüşleri. *Academia Eğitim Araştırmaları Dergisi, 5*(2), 270-292.

Eren, A. A., Önal, N. T., & Büyük, U. (2020). Elementler ve bileşikler konusu için geçerli ve güvenilir bir başarı testi geliştirme çalışması. *Pearson Journal, 5*(6), 152-167.

Erhan, G., Doğan, O., & Haser, Ç. (2024). A scale for seventh and eighth grade students' attitudes towards skill-based mathematics questions. *Turkish Journal of Education, 13*(3), 199-216. https://doi.org/10.19128/turje.1398979

Ermiş, F., & Karaman, İ. (2022). Development of achievement test for "force and energy" chapter in science course. *Cukurova University Faculty of Education Journal, 51*(2), 1278-1314. https://doi.org/10.14812/cuefd.1059186

Ersoy, E. & Bayraktar, G. (2018). İlkokul 4. Sınıf Matematik Dersi "Ondalık Gösterim" Alt Öğrenme Alanına İlişkin Başarı Testi Geliştirilmesi. *Buca Eğitim Fakültesi Dergisi,* 46, 240-266

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

Gönen, S., Kocakaya, S., & Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 8*(1), 40-57.

Guerra-Reyes, F., Guerra-Dávila, E., Naranjo-Toro, M., Basantes-Andrade, A., & Guevara-Betancourt, S. (2024). Misconceptions in the learning of natural sciences: A systematic review. *Education Sciences*, 14(5), 497. https://doi.org/10.3390/educsci14050497

Iliaki, G., Velentzas, A., Michailidi, E., & Stavrou, D. (2018). Exploring the music: a teaching-learning sequence about sound in authentic settings. *Research in Science & Technological Education*, 37, 218 – 238.

İdin, Ş., & Aydoğdu, C. (2016). Kuvvet ve hareket ünitesi başarı testi geçerlik ve güvenirlik araştırması. *Eğitim Bilim ve Teknoloji Araştırmaları Dergisi, 1*(1), 14-33.

İlhan, N. & Hoşgören, G. (2017). Fen Bilimleri dersine yönelik yaşam temelli başarı testi geliştirilmesi: Asit baz konusu. *Fen Bilimleri Öğretimi Dergisi,* 5 (2), 87-110.

Karabulut, H., Tosunbayraktar, G., & Kariper, A. (2022). Ortaokul öğrencilerinin beceri temelli (yeni nesil) fen bilimleri sorularına yönelik görüşlerinin incelenmesi. *EDUCATIONE*, 1(2), 301-320.

Karadağ, N. (2014). Açık ve uzaktan eğitimde ölçme ve değerlendirme: Mega üniversitelerdeki uygulamalar [Doctoral dissertation]. Anadolu Universitesi.

Karslı, G., Karamustafaoğlu, S. & Kurt, M. (2019). Yenilenen Fen Bilimleri Dersi Öğretim Programına Yönelik 7. Sınıf "Hücre ve Bölünmeler" Ünitesi Başarı Testi: Geçerlik ve Güvenirlik. *Fen Bilimleri Öğretimi Dergisi. 7*(1), 68-98.

Keçeci, G., Yıldırım, P., & Zengin, F. K. (2019). Sistemler akademik başarı testi: Geçerlik ve güvenirlik çalışması. *Ulusal Eğitim Akademisi Dergisi, 3*(1), 96-114.

Khorram, B., Khakbaz, A., & Yousefzadeh, M. (2024). The process, methods and factors affecting the acculturation of new teachers: a systematic review. *Research in Teaching*, 12(2), 58-89.

Kolomuc, A., & Karagölge, Z. (2021). Comparison of the 7th Grade students' accomplishments in skill and acquisition based assessment-evaluation. *Journal of Science Learning*, 4(2), 134-139.

Kurt, A., Aydın M. & Bekereci, Ü. (2023). Achievement Test Development Study for 6th Grade Sound and Characteristics Unit: Validity and Reliability Analysis, *E-International Journal of Educational Research, 14* (3), 20-33

Linn, R. L., ve Gronlund, N. E. (2000). *Measurement and assessment in teaching*. 8th ed. Englewood Cliffs, NJ: Merrill/Prentice Hall.

MoNE (2024). Fen bilimleri dersi öğretim programı (3,4,5,6,7 ve 8. sınıflar) https://mufredat.meb.gov.tr

Meço, G., & Görgülü Arı, A. (2024). Fen Eğitimi için "Vücudumuzdaki Sistemler" Konulu Akademik Başarı Testi Geliştirme Çalışması. *International Journal of Social Science Research, 13*(1), 91-106.

Miller, M. D., Linn, R. L., ve Gronlund, N. E. (2012). *Measurement and assessment in teaching* (11th ed.). Upper Saddle, NJ: Pearson.

Nacaroğlu, O., Bektaş, O., & Kızkapan, O. (2020). Madde döngüleri ve çevre sorunları konusunda başarı testi geliştirme: Geçerlik ve güvenirlik çalışması. *Kastamonu Education Journal, 28*(1), 36-51.

Nitto, A. J., & Brookhart, S. M. (2016). Öğrencilerin eğitsel değerlendirmesi. Çeviri Editörleri: Bıçak, B. Bahar, M. ve Özel, S., Ankara: Nobel Yayıncılık.

Özcan, F., & Orhan, A. T. (2024). İlkokul Üçüncü Sınıf Fen Bilimleri Dersi Kuvvet ve Hareket Ünitesi Akademik Başarı Testi Geliştirme Çalışması. *Eğitim Bilim ve Araştırma Dergisi, 5*(2), 126-154.

Özkan, E. B., & Muştu, Ö. E. (2018). 8. Sınıf Basit Makineler Ünitesine Yönelik Başarı Testi Geliştirme: Geçerlik Ve Güvenirlik Çalışması. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 11*(1), 737-754.

Özkılıç, G. E., Bektas, O., & Karaca, M. (2023). Sindirim Sistemi Ünitesine Yönelik Başarı Testi Geliştirme: Geçerlik ve Güvenirlik Çalışması. *Araştırma ve Deneyim Dergisi, 8*(1), 115-154.

Panjaitan, R. (2016). Implementing a "choose your own adventure" method in teaching basic concepts of sound to primary-school teacher candidate students. *Journal of Physics: Conference Series*, 739.

Pradana, D., Nur, M., & Suprapto, N. (2020). Improving critical thinking skill of junior high school students through science process skills based learning. *Jurnal Penelitian Pendidikan*, 6(2), 166-172. https://doi.org/10.29303/jppipa.v6i2.428

Ramsey, G. (2015). Using sound and music to teach waves. *Journal of the Acoustical Society of America*, 138, 1771-1771.

Rennie, L. J., & Parker, L. H. (1996). Placing physics problems in real-life context: students' reactions and performance. *Australian Science Teachers Journal, 42*(1), 55-59.

Rico, A., González, A., Azula, O., & Aranzábal, J. (2021). Learning difficulties about the sound model: a review of the literature. *Enseñanza de las Ciencias. Revista de investigación y experiencias didácticas. 39(2), 5-23.*

Şan, S., & İlhan, N. (2022). Fen bilimleri dersi beceri temelli sorulara (yeni nesil) yönelik kuramsal ve kavramsal çerçeve. *İnönü Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi, 9*(17), 17-36.

Sabri, S. (2019). *The effects of inquiry-based learning on the development of high school biology students' critical thinking skills in the UAE* [Doctoral dissertation]. The British University in Dubai.

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.

Saylan-kırmızıgül, A., & Kaya, H. (2019). "Aynalarda Yansıma ve Işığın Soğrulması" konusunda geçerliliği ve güvenilirliği sağlanmış başarı testi geliştirme çalışması. *Journal of Theoretical Educational Science, 12*(2), 474-493.

SEPT (1999). American Educational Research Association, American Psychological Association, and National council on measurement in education. *Standards for Educational and Psychological testing*. Washington, DC: American Educational Research Association.

Scriven, M. (1991). Evaluation thesaurus. Sage Publications.

Sevim, S., Uysal, İ., & Demirci, E. (2021). Fen bilimleri dersi 5. sınıf "Işığın Yayılması" ünitesine yönelik başarı testi geliştirme çalışması. *Caucasian Journal of Science, 8*(2), 224-246.

Sidiq, Y., Ishartono, N., Desstya, A., Prayitno, H. J., Anif, S., & Hidayat, M. L. (2021). Improving elementary school students' critical thinking skill in science through hots-based science questions: A quasi-experimental study. *Jurnal Pendidikan IPA Indonesia*, 10(3), 378-386. https://doi.org/10.15294/jpii.v10i3.30891

Sontay, G. ve Karamustafaoğlu, O. (2020). Fen bilimleri dersi "Güneş, Dünya ve Ay" ünitesine yönelik başarı testinin geliştirilmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 40*(2), 511-551.

Sontay, G., & Karamustafaoğlu, S. (2017). 5. sınıf fen bilimleri dersi "yer kabuğunun gizemi" ünitesine yönelik başarı testi geliştirme. *Fen Bilimleri Öğretimi Dergisi, 5*(1), 62-86.

Soylu, Ü. İ., Karamustafaoğlu, S., & Karamustafaoğlu, O. (2020). 6. Sınıf "Madde ve Isı" ünitesi başarı testi geliştirme: geçerlik ve güvenirlik. *Ihlara Eğitim Araştırmaları Dergisi, 5*(2), 271-293.

Sözen, M., & Bolat, M. (2014). 11–18 yaş öğrencilerin ses hızı ile ilgili sahip oldukları kavram yanılgılarının belirlenmesi. *Ondokuz Mayis University Journal of Education Faculty, 33*(2), 505-523.

Syah, F.A. (2025). *Analisis miskonsepsi siswa berbantuan instrumen tes diagnostik five-tier untuk topik gelombang*, [Doctoral dissertation]. Universitas Negeri Jakarta

Turan, F., & Timur, B. (2023). Güneş sistemi konusuna yönelik beceri temelli başarı testi geliştirilmesi. *SDU International Journal of Educational Studies, 10*(2), 88-112. https://doi.org/10.33710/sduijes.1366550

Turgut, M. F. (1995). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Yargıcı Matbaası.

Ünsal, S., & Kaba, S. (2022). The characteristics of the skill based questions and their reflections on teachers and students. *Kastamonu Education Journal, 30*(2), 273-282. https://doi.org/10.24106/kefdergi.753717

Wallen, N. E., & Fraenkel, J. R. (2013). *Educational research: A guide to the process*. Routledge.

Yuzuak, A. V., & Recepkethüda, R. (2022). Evaluation of skill-based questions according to science education teachers in Turkey. *Science Education International*, *33*(1), 122-132. https://doi.org/10.33828/sei.v33.i1.13

Appendix 1. Expert Opinions

Item No	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Final Decision
1	Understand	Understand	Understand	Understand	Understand	Understand	Understand
2	Understand	Understand	Understand	Understand	Understand	Understand	Understand
3	Remember	Remember	Remember	Remember	Remember	Remember	Remember
4	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze
5	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze
6	Understand	Understand	Understand	Understand	Understand	Understand	Understand
7	Remember	Remember	Remember	Remember	Remember	Remember	Remember
8	Create	Create	Create	Create	Create	Understand	Create
9	Remember	Remember	Remember	Remember	Remember	Understand	Remember
10	Understand	Understand	Understand	Remember	Understand	Understand	Understand
11	Evaluate	Evaluate	Evaluate	Analyze	Evaluate	Evaluate	Evaluate
12	Remember	Remember	Remember	Remember	Remember	Understand	Remember
13	Evaluate	Evaluate	Evaluate	Understand	Evaluate	Evaluate	Evaluate
14	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze	Analyze
15	Create	Create	Create	Create	Create	Create	Create
16	Remember	Remember	Remember	Remember	Remember	Remember	Remember
17	Remember	Remember	Remember	Remember	Remember	Remember	Remember
18	Understand	Understand	Understand	Understand	Understand	Understand	Understand

Item No	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Final Decision
19	Apply	Apply	Apply	Analyze	Apply	Apply	Apply
20	Apply	Apply	Apply	Analyze	Apply	Apply	Apply
21	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate
22	Understand	Understand	Understand	Understand	Understand	Understand	Understand
23	Remember	Remember	Remember	Remember	Remember	Remember	Remember
24	Remember	Remember	Remember	Remember	Remember	Remember	Remember
25	Remember	Remember	Remember	Remember	Remember	Remember	Remember
26	Understand	Understand	Understand	Understand	Understand	Understand	Understand
27	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate	Evaluate
28	Understand	Understand	Understand	Understand	Understand	Understand	Understand

Appendix 2. Example Questions

1. Example question in the application stage according to Bloom's taxonomy.

Ali is curious whether the sound level he hears differs depending on the medium in which the objects are located.

Therefore, as shown in pictures I, II, and III, Ali takes a domino piece and listens to the sounds produced by tapping it: in Figure 1, in the air; in Figure 2, inside a container filled with water; and in Figure 3, on a table. Which conclusion did Ali reach from his experiment?

A) Sound travels faster in a liquid medium.

B) The speed of sound propagation varies depending on the medium.

C) Sound is heard the same in every medium.

D) The sound level is higher in the air.

2. Example question in the evaluation step according to Bloom's taxonomy.

In the science class, the teacher set up the apparatus shown in the figures so that students could observe the propagation characteristics of sound in different media.

Figure 1

Figure 2

Figure 3

In these setups, the speaker used as the sound source is located as follows: in Figure 1, on a table; in Figure 2, inside a glass container filled with air; and in Figure 3, inside an open container filled with water. During the experiment, the speakers produced the same sound wave at the same intensity simultaneously. However, listeners placed at equal distances from the setups heard the sounds differently depending on the medium.

According to the experiment:

I. Sound propagates in a vacuum.

II. The speed of sound in water is greater than the speed of sound in air.

III. When the medium through which sound is transmitted changes, the sound heard remains the same.

Which of the above statements is correct?

A) Only I

B) Only II

C) I and II

D) I, II and III