

Developing an Achievement Test on the Subject of ‘Circulatory System’ for Sixth Graders[†]

Nurdan Akgun, Seyda Gul

Ataturk University, Erzurum, Turkey

Abstract: *This study aims to develop a test to measure sixth grade students’ success in ‘Circulatory System’. For this purpose, a draft test containing 22 questions was prepared by the researchers. And then, the draft test was submitted to expert opinion and examined in terms of language and content. Following expert opinions, the 22 multiple-choice questions were administered to 128 seventh graders attended to two different secondary schools. Item analysis was performed in the data and thus, two questions were removed from the draft test. According to the findings, the overall item difficulty value of the 20-question final test was calculated as 0.654, and the item discrimination value was calculated as 0.54. KR-20 reliability coefficient of the test was 0.71. Seven of the questions are at the remembering level. Additionally, there were eight questions at the understanding level and five questions at the analysis level according to Bloom’ revised taxonomy.*

Science Insights Education Frontiers 2024; 20(2):3247-3262.

DOI: 10.15354/sief.24.or526

How to Cite: Akgun, N., & Gul, S. (2024). Developing an achievement test on the subject of ‘circulatory system’ for sixth graders. Science Insights Education Frontiers, 20(2):3247-3262.

Keywords: *Achievement, Circulatory System, Test Development, Sixth Graders*

About the Author: Nurdan Akgun, Graduate School of Educational Science, Ataturk University, Erzurum, Turkey, E-mail: nurdanakgun1907@gmail.com, ORCID: <https://orcid.org/0009-0004-3104-7183>

Seyda Gul, Biology Education Division, Department of Mathematics & Science Education, Kazım Karabekir Faculty of Education, Ataturk University, Erzurum, Turkey, E-mail: seydagul@atauni.edu.tr, ORCID: <https://orcid.org/0000-0003-4005-2158>

Correspondence to: Dr. Seyda Gul at Ataturk University of Turkey.

Conflict of Interests: None

¶ This article was produced from the master's thesis of the first author. The authors have no competing interests to declare that are relevant to the content of this article. All authors have participated in the research and in the article preparation. All authors have approved the final article.

Ethics Approval: All procedures performed in the study were in accordance with the ethical standards of Atatürk University. Ethic approvals were obtained with the decision (reference no: E-56785782-050.02.04-2300297174) of the Social and Human Sciences Ethics Committee.

© 2024 Insights Publisher. All rights reserved.



Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed by the Insights Publisher.

Introduction

TODAY, the aim of education is to raise individuals who are enterprising, self-expressive, productive, problem-solving, open to new ideas, and able to think critically. In this perspective, the way to teach students information and today's technology depends on the education and training programs provided in schools (Buyruk, 2018; Tutkun & Aksoyalp, 2010). A correct understanding of the outcomes that are intended to be given to students occurs during the teaching process. The extent to which the teaching goals have been achieved, that is, how successful the teaching process is, is understood through measurement and evaluation (Açıkgoz & Karşı, 2015; Yıldız et al., 2019). In other words, the students' achievement and permanent behaviours are evaluated through measurement and evaluation process (Üçüncü & Sakiz, 2020).

Measurement is the symbol equivalent of certain features that we need in every aspect of our lives. In order to evaluate education, learners must be measured based on a criterion. We can get the answer to how much the achievements determined in education have been achieved by measuring. The quality of the measurement tool depends on its validity and reliability (Öztemel, 2018). Due to the need to create a qualified measurement tool, the "measurement and evaluation" course has begun to be taught as a compulsory course in teacher training faculties in our country (Özkan, 2006).

One of the most important issues in performing measurement and evaluation precisely and accurately is the selection of measurement tools. This is also valid for science education. If the academic achievements of students are to be measured in education, one of the measurement tools that should be used is achievement tests (Kargin & Gül, 2021). Achievement tests are instruments developed to measure the knowledge and skills gained by the student through a certain learning process. The essays, quizzes, open-book tests, true-false tests, multiple choice tests, homework and projects are used to measure and evaluate student success (Ayvaci & Durmuş, 2016). The most preferred among these measurement tools are multiple choice tests. These tests are often used in transition to a higher education institution in our country (Temizkan & Sallabaş, 2015). In multiple choice tests, the number of options can generally vary between three and five, depending on the student's level. Only one of these options is correct, and the others contain incorrect answers (Kargin & Gül, 2021). These tests are very reliable and can be easily applied in a short time, so they are also very useful for teachers. For this reason, it is frequently preferred in measurement and evaluation processes. In addition, since its scoring is objective, it provides an advantage in terms of providing reliable results (Ayvaci & Durmuş, 2016; Çardak & Selvi, 2018; Singh & Rosengrant, 2003; Timur et al., 2019). Of course, mul-

multiple choice tests have some disadvantages as well as these advantages (Temizkan & Sallabaş, 2015). These features are shown in **Figure 1**.

One of the most important features that distinguishes multiple choice tests from other testing techniques is that they provide a lot of data for statistical studies. One of the positive features of multiple choice questions in education is that they can be applied to all age groups, from primary school level to higher education level. It can be said that its reliability is high because the results obtained by different raters are close to each other (Akbulut & Çepni, 2013). Multiple choice tests, which are difficult to prepare but easy to score, are prepared to suit the age of the student. In our country, the number of options for multiple choice tests is preferred as three for primary schools, four for secondary schools, five for high schools and higher education. Since it is possible to ask questions about each learning outcome in these exams, content validity can also be ensured. However, it is difficult to write questions about higher-level cognitive skills. Written items can generally measure lower-level cognitive features such as knowledge, comprehension and application level (Küçükahmet, 2002).

There are researches on developing achievement tests in different subjects of chemistry, physics and biology in science courses in our country (Bolat & Karamustafaoğlu, 2019; Demir et al., 2016; İlhan & Hoşgören, 2017; Keçeci et al., 2019; Kenan & Özmen, 2014; Nacaroğlu et al., 2020; Saraç, 2018; Şener & Taş, 2017; Yalinkilic & Gul, 2023). However, regarding biology, only an achievement test developed by Gülbahar (2023) was found on the circulatory system in the science curriculum updated in 2018 in Turkey. When the development process of this test is examined, the distribution of the questions to the learning outcomes is not specified, and on the other hand, their place in the renewed Bloom Taxonomy has not been examined.

Researchers state that when evaluating student performance in the teaching process, instruments that measure students' understanding and comprehension levels should be used rather than instruments that direct students to memorize. At this point, the renewed Bloom Taxonomy appears as a classification in which different thinking stages are listed (Gündüz, 2009). Taxonomy is an approach that allows classifying the learning outcomes expected to be achieved by students as a result of an education (Bümen, 2006). Therefore, it can be said that there is a need to develop a new achievement test on the sixth grade Circulatory System, which is compatible with the achievements of the 2018 science course curriculum (Ministry of National Education [MoNE], 2018) and takes into account the Revised Bloom Taxonomy. Therefore, in this research, Circulator System Achievement Test (CSAT) was developed by considering the MoNE (2018) sixth grade Science Course Curriculum.

Advantages	Disadvantages
<ul style="list-style-type: none">•The number of questions can be increased as desired.•It has high reliability and content validity.•There is a low chance of error in the results.•It is easy to collect data and analyze it.•It can be applied at the input, process or result stages of education.•It ensures efficient use of time.	<ul style="list-style-type: none">•Preparing distractors is difficult.•Higher-level cognitive skills are difficult to measure.•The number of options should be appropriate to the student level.•Students are less likely to identify which part of the topic they do not understand.

Figure 1. Advantages and Disadvantages of Multiple Choice Tests.

Purpose of the Research

This study aims to develop an achievement test on 'Circulatory System' in the sixth grade science course curriculum. For this aim, answers were sought to the following research questions.

1. *Is the CSAT developed to determine the students' achievement levels on the Circulatory System valid?*
2. *Is the CSAT developed to determine the students' achievement levels on the Circulatory System reliable?*

Method

This study was conducted with the survey method. The sample group research consisted of 128 (65 females, 63 males) 7th grade students from two primary schools randomly selected in a city in the east Turkey. 87 of the students (41 females, 46 males) participated in the study from the first school and 41 (24 females, 17 male) from the second school. According to this, 50.1% of the students are female and 49.2% male in total.

Development Process of Circulatory System Achievement Test (CSAT)

The development process of the CSAT was continued by considering the stages of the test development predicted in many studies in the literature. The stages of the test development followed in this research are similar to the stages used in the studies of Haladyna (1997) and Kızıkan and Bektaş (2018) was done in a similar way. Accordingly, the stages followed and the procedures performed in the test development process are presented below:

Determining the Purpose and Scope of the Test

The purpose of the CSAT to be developed in the research is to determine the student success level for the 'Circulatory System' subject in the sixth grade 'Systems in Our Body' unit. According to the Science Course Curriculum, there are five learning outcomes for the subject of 'Circulatory System' (MoNE, 2018). These:

- (1) F.6.2.3.1. Explains the functions of the structures and organs that make up the circulatory system using models.
- (2) F.6.2.3.2. Examines the pulmonary and systemic circulation on a diagram and explains their functions.
- (3) F.6.2.3.3. Defines the structure and functions of blood.
- (4) F.6.2.3.4. It refers to blood exchange between blood groups.
- (5) F.6.2.3.5. Evaluates the importance of blood donation for society.

In addition to the above learning outcomes, some basic topics and concepts that are expected to be taught about the 'Circulatory System' subject are also included in the sixth grade curriculum (structures and organs that make up the circulatory system, structure and function of the heart, blood vessels, pulmonary and systemic circulation, blood groups, blood donation, circulatory system). Therefore, CSAT was developed by considering the learning outcomes and concepts included in the curriculum.

Literature Review and Determining Test Questions

After determining the scope of the test by examining the 6th Grade Science Curriculum (MoNE, 2018), a comprehensive literature review was conducted by the researchers. In this process, first of all, the achievement tests developed for 6th Graders on the Circulatory System were examined. As a result of the investigations, achievement tests developed by different researchers were found. For example, expert opinion was sought for the validity of the test developed by Çakmak et al. (2012). Its reliability was determined with data collected from 42 students and the Cronbach Alpha internal consistency coefficient was calculated as 0.79. The pilot application of the circulatory system achievement test developed by Özay Köse and Yıldırım (2020) was conducted with a total of 50 students in the 2017-2018 academic year. After item analysis, a test consisting of 25 questions was developed.

However, the reliability coefficient was not mentioned in the study. Achievement test prepared by Yalçinkaya (2018) was prepared according to the Ministry of Education (2005) Science and Technology Course Curriculum. Validity and reliability studies were provided with expert opinion and data collected from 38 students. As a result of item analysis, an achievement test consisting of 19 multiple-choice questions was developed. Cronbach Alpha reliability coefficient was found to be 0.71. An achievement test was developed by Gülbahar (2023) for the new curriculum updated in 2018. However, when the development process of this test is examined, the distribution of the questions to the learning outcomes is not specified, and on the other hand, their place in the renewed Bloom Taxonomy has not been examined.

As a result, when the achievement tests developed for 6th graders on the Circulatory System are examined, it is seen that some of them did not contain item analysis, a pilot study was conducted with a small number of samples, only expert opinion was consulted for validity, the tests were prepared according to the pre-2018 curriculum, the distribution of the questions according to Bloom's taxonomy was not examined, or the questions in the test were at the lower levels of the cognitive domain. For this reason, it was decided to develop a new achievement test in the study. Thus, different questions were prepared from these studies. In the process of preparing questions, relevant resources (textbooks, different resources in internet, etc.) were examined and a total of 22 multiple choice questions were prepared for the Circulatory System topic.

Preparing the Draft CSAT and Asking for Expert Opinion

As stated before in the research, 22 multiple-choice questions, each with four options, were prepared after reviewing the relevant literature. In the literature, checking content validity is one of the must-do practices regarding the test development process (Tunç & Kılınç Alpat, 2015). Because to what extent the questions in the measurement tool represent the scope of the subject is an issue that must be taken into consideration. One of the frequently used methods for the content validity is to obtain expert opinion (Çalık & Ayas, 2003; Treagust, 1988). For this purpose, the draft form of CSAT, which includes 22 multiple-choice questions, was presented to the opinion of a total of six experts, including five academicians and one science teacher. After being examined by experts in terms of features in terms of language, scope, appearance, content, etc., the test was given its final form.

In the research, in addition to expert opinion, to ensure the content validity of the test, a specification table was prepared according to the Bloom's Revised Taxonomy and the questions were analysed by the same

experts. Thus, it was determined which level of the cognitive domain the 22 questions in the test fell into in Bloom's Taxonomy (**Table 1**).

When **Table 1** is examined, seven out of 22 questions are at the remembering level, eight at the understanding level, one at the application level and six at the analysing level.

Findings

During the scope validity studies of the research, the draft CSAT was finalized in line with expert opinions. Then, the item analysis and reliability analysis phase was started. According to Bryman and Cramer (2001), the number of samples should be at least five times the number of items in the scale. Therefore, in this research, the draft test was administered to 128 students during one class hour. Afterwards, item analysis was performed on the data obtained from the students. At the stage of item analysis, values of item difficulty-item discrimination were calculated (**Table 2**).

According to Rao et al. (2016) and Deborah et al. (2021), questions with $p_j > 70\%$ are classified as 'easy', questions with $0.30 < p_j < 0.70$ are classified as medium difficulty, and questions with $p_j < 0.30$ are classified as 'difficult' in multiple choice tests. On the other hand, questions with $r_j \leq 0.30$ are weakly discriminative; questions with $0.30 < r_j \leq 0.39$ are classified as good level discrimination, and $r_j > 39\%$ are classified as excellent level of discrimination. When the item analysis results in **Table 2** are examined, the item difficulty index of one question (Q-8) fell into the difficult category. The item difficulty index value of the nine questions was evaluated in the very easy category. All other questions are of medium difficulty. When the item discrimination of the questions was examined, the item discrimination index value of two questions (Q-8, Q-18) fell into the weak category. However, four questions are at a good level of discrimination, and all of the remaining questions are at an excellent level of discrimination. In light of these findings, it was decided to remove two questions (Q-8, Q-18). The average item difficulty index for the remaining 20 questions was calculated as 0.65, and the average item discrimination was calculated as 0.54. The KR-20 reliability coefficient of 20-question CSAT is 0.71. When the analysis results are evaluated, it can be said that CSAT is a reliable test with medium difficulty and excellent discrimination. After the item analysis was completed, the 20 questions left in the test were renumbered and the test was given its final form. The distribution of the questions in the final CSAT according to the Revised Bloom taxonomy is shown in **Table 3**.

When **Table 3** is examined, seven out of 20 questions are at the remembering level, eight at the understanding level and five at the analysing level.

Table 1. Specification Table for the Questions in the Draft CSAT.

Learning Outcomes	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Learning Outcome 1	Q-1, Q-2, Q-4	Q-3, Q-6, Q-9, Q-10	-	Q-8	-	-
Learning Outcome 2	-	Q-5, Q-6, Q-7, Q-9, Q-10	-	Q-8	-	-
Learning Outcome 3	Q-11, Q-15	-	-	Q-12, Q-13, Q-14	-	-
Learning Outcome 4	Q-16	Q-19	Q-18	Q-17, Q-21	-	-
Learning Outcome 5	Q-20	Q-19, Q-22	-	Q-21	-	-

Table 2. Item Analysis Results of CSAT.

Items	pj	rj	Items	pj	rjx
Q-1	0.81	0.66	Q-12	0.59	0.51
Q-2	0.69	0.58	Q-13	0.70	0.56
Q-3	0.80	0.65	Q-14	0.79	0.63
Q-4	0.58	0.51	Q-15	0.58	0.46
Q-5	0.78	0.63	Q-16	0.41	0.36
Q-6	0.83	0.66	Q-17	0.39	0.36
Q-7	0.72	0.59	Q-18	0.44	0.29
Q-8	0.23	0.12	Q-19	0.55	0.51
Q-9	0.77	0.62	Q-20	0.34	0.35
Q-10	0.89	0.68	Q-21	0.54	0.59
Q-11	0.80	0.64	Q-22	0.47	0.34
Overall test				0.65	0.54

Table 3. Specification Table for the Questions in the Twenty-Question Final CSAT.

Learning outcomes	Questions*	Bloom's Taxonomy
Learning Outcome 1	Q-1	Remembering
	Q-2	Remembering
	Q-3	Understanding
	Q-4	Remembering
	Q-6	Understanding
	Q-8	Understanding
Learning Outcome 2	Q-9	Understanding
	Q-5	Understanding
	Q-6	Understanding
	Q-7	Understanding
Learning Outcome 3	Q-8	Understanding
	Q-9	Understanding
	Q-10	Remembering
	Q-11	Analyzing
	Q-12	Analyzing
Learning Outcome 4	Q-13	Analyzing
	Q-14	Remembering
	Q-15	Remembering
	Q-16	Analyzing
Learning Outcome 5	Q-17	Understanding
	Q-19	Analyzing
	Q-18	Remembering
	Q-20	Understanding

* Some questions meet more than one learning outcome

Discussion and Conclusion

As stated before, the MoNE makes changes in the curriculum in certain periods in line with the needs of the education process. Therefore, measurement and evaluation tools applied in schools need to be updated in accordance with new curriculum. This study aims at developing a CSAT for the 6th graders.

Achievement tests are one of the most frequently preferred tools to measure students' achievement levels in any subject. In this regard, it is important that achievement tests are valid in terms of scope and structure (Üçüncü & Sakız, 2020). In this research, it can be said that a consistent test with high content validity was developed as a result of the improvements made in line with item analyses and expert opinions. It is highly likely that carrying out the test development process correctly by following the stages of the test development prescribed in the literature will cause this result. As a matter of fact, the stages followed in the literature are similar to this study (Haladyna, 1997; İlhan, 2020; Kargın & Gül, 2021). The stages of the test in the literature are similar to this study (Açıkgöz & Karslı, 2015; Bolat & Karamustafaoğlu, 2019; Haladyna, 1997; Kargın & Gül, 2021; Kızırcapan & Bekaş, 2018). For example; Eren et al. (2020) developed a test on 'elements and compounds' in the 7th grade secondary school science courses by following similar steps to this research. As a result of reliability and validity analysis, they developed a 27-item test for the subject 'elements and compounds'. Çardak and Selvi (2019) performed a similar study for the 'teaching principles and methods' course. It was carried out by following a 10-stage process, including the determining the purpose of developing the achievement test, determining the objectives and behaviours of the course, creating the specification table, writing multiple choice test items for critical behaviours, presenting the specification table and test items to expert opinion, obtaining student opinions for the first test draft, developing the draft test, the application (first trial application), item analysis with the data obtained from the trial application, and calculation of the statistics of the second trial application and the final test. After the validity and reliability analysis, it was concluded that this multiple-choice test can be used in scientific studies and courses related to 'Teaching Principles and Methods' course subjects. As a result, it can be said that this research overlaps with some studies in the literature in terms of test development steps (Yalinkilic & Gul, 2023). However, since the questions in this research were examined by a small number of experts, this can be considered a limitation for the study. Because Akbulut and Çepni (2013) and Üçüncü and Sakız (2020) suggested that in order to ensure content validity, it should be determined whether the harmony between expert opinions is valid or not.

Considering the stages of the test development in the research, firstly a literature review was conducted. Thus, 22 questions, each with four options, were prepared in line with the learning outcomes in the Science Course Curriculum. After expert opinions, the questions were placed in the specification table. In the literature, while preparing the draft test, it is recommended to write three items for each outcome in the specification table (Atılğan, 2013; Yalinkilic & Gul, 2023). In this study, an attempt was made to increase the content validity CSAT by paying attention to this criterion. After content validity studies, the draft test containing 22 questions was administered to 128 students for item analysis. At stage of item analysis, item difficulty-item discrimination indexes were calculated and the test was reduced to 20 questions. The average item difficulty of the CSAT was 0.65, and the average item discrimination was 0.54. It can be said that these results show that the test is of medium difficulty and perfectly discriminatory (Assimi et al., 2022). In addition to this positive result, although multiple choice tests have many advantages such as application and scoring (Ayvacı & Durmuş, 2016), they also have disadvantages such as finding the correct answer randomly. In order to minimize this disadvantage, it is recommended that open-ended questions be included in achievement tests in addition to multiple-choice questions (Saylan-Kırmızıgül, & Kaya, 2019). Therefore, the fact that there are only multiple choice questions in the achievement test developed in this study is a limitation for this study. However, the multiple choice questions in the test can be of great benefit to teachers in terms of saving time, ease of scoring, objectivity and preparation for exams that measure achievement. On the other hand, although multiple choice questions are not fully effective in determining misconceptions, the answers given can identify deficiencies or distortions in the subject (İdin & Aydoğdu, 2016; Özkan & Eryılmaz Muştu, 2018; Şen & Eryılmaz, 2011).

Bloom taxonomy is frequently used in test development processes in the field of education. This situation is explained by the advantage of the cognitive domain classification in Bloom's taxonomy in ensuring that the purposes of the items are clear and observable (Ayvacı & Türkdoğan, 2009). Supporting this situation, Özkan and Yadigaroglu (2020) stated that when developing a test, it is necessary to prepare the questions by taking the renewed Bloom Taxonomy into consideration. Thus, the questions can be classified more clearly and in more detail.

The findings of the research showed that seven out of 20 questions in the CSAT are at the remembering level, eight at the understanding level and five at the analysis level. One of the two questions eliminated as a result of item analysis is at the analysis stage (Q-8) and the other is at the application stage (Q-18). These questions may have been found difficult by most students because they were in high knowledge level of Bloom's taxonomy. As a matter of fact, item difficulty and discrimination values show that these

questions are outside the desired limits. In addition to the item analysis findings, the KR-20 reliability coefficient of the CSAT was 0.71. This value shows that the test is a reliable test.

The findings of the research also showed that all of the questions in the final CSAT cover all the learning outcomes in the curriculum and are distributed evenly among the learning outcomes. Accordingly, it can be said that the content validity of the achievement test is high. In addition, it can be predicted that the test will provide detailed information about the students' learning levels and give accurate results about the general concepts of the subject.

It is envisaged that this test, whose reliability and validity has been verified, can be used by educators as a measurement tool in measurement and evaluation studies on the subject of 'Circulatory System'. In addition, this test can be used by science education researchers in experimental studies as a data collection tool to compare the achievements of student groups.

As a result, taking into account the findings and limitations of the study, the following suggestions are offered for future research:

- As it is known, multiple choice tests are not sufficient to measure high-level knowledge of the cognitive domain in Bloom's taxonomy. As a matter of fact, in this research, questions for the create level were not prepared in the draft test. Since high-level information is mostly tested with open-ended questions, different types of questions such as open-ended questions can be prepared in addition to multiple choice questions in similar achievement tests in the future.
- Research suggests that two- or three-stage tests measure knowledge better. Therefore, two or three-stage multiple choice tests can be developed in future research.
- In future research, new achievement tests can be developed by using larger samples and applications in schools at different levels.

References

- Açıkgöz, M., & Karlı, F. (2015). Alternatif ölçme-değerlendirme yaklaşımları kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenirlik analizi [Validity and reliability analysis of achievement test which developed for work and energy, using alternative measure-evaluation techniques]. *Amasya Education Journal*, 4(1):1-25.
- Akbulut, H. İ., & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir? İlköğretim 7.sınıf kuvvet ve hareket ünitesi [How to develop an achievement test for a unit?: A study for grade 7 force and motion unit]. *Amasya Education Journal*, 2(1):18-44.

- Assimi, E., Idrissi, R.J., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, DOI: <https://doi.org/10.1080/00219266.2022.2134175>
- Atılğan, H. (2013). Eğitimde Ölçme ve Değerlendirme (6. Baskı) [Measurement and Evaluation in Education (6th Edition)] Ankara: Anı Yayıncılık.
- Ayvacı, H. Ş., & Türkođan, A. (2010). Yeniden yapılandırılan Bloom taksonomisine göre fen ve teknoloji dersi yazılı sorularının incelenmesi [Examination of science and technology course written questions according to the restructured Bloom taxonomy]. *Journal of Turkish Science Education*, 7(1):13-25. DOI: <https://doi.org/10.17121/ressjournal.302>
- Ayvacı, H. Ş., & Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: Isı ve sıcaklık başarı testi geçerlik ve güvenilirlik araştırması [An achievement test development study: heat and temperature achievement test validity and reliability research]. *Ondokuz Mayıs University Journal of Education Faculty*, 35(1):87-102. DOI: <https://doi.org/10.7822/omuefd.35.1.8>
- Bolat, A., & Karamustafaođlu, S. (2019). Vücudumuzdaki sistemler ünitesi başarı testi geliştirme: Geçerlik ve güvenilirlik [Development achievement of test on the unit "systems in our body": validity and reability]. *Gazi Journal of Education Sciences*, 5(2):131-159. DOI: <https://doi.org/10.30855/gjes.2019.05.02.008>
- Bryman, A., & Cramer, D. (2001). Quantitative Data Analysis with SPSS Release on for Windows. USA: Routledge; Philadelphia, PA.
- Buyruk, H. (2018). Gelişen teknolojiler, değişen işgücü nitelikleri ve eğitim [Developing technologies, changing labour qualities and education]. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 8(14):599-632. DOI: <https://doi.org/10.26466/opus.404223>
- Bümen, N. T. (2006). Program geliştirmede bir dönüm noktası: Yenilenmiş Bloom taksonomisi [A revision of the Bloom's Taxonomy: A turning point in curriculum development]. *Education and Science*, 31(142):3-14.
- Çakmak, M., Gürbüz, H., & Kaplan, H. (2012). Dolaşım sistemimiz konusunda uygulanan kavram haritalarının öğrencilerin akademik başarısına etkisi [The impact of the concept maps used for our circulatory system on students academic success]. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(10):9-28.
- Çalık, M., & Ayas, A. (2003). Çözümlerde kavram başarı testi hazırlama ve uygulama [Preparing concept success test in solutions and application]. *Pamukkale University Journal of Education*, 2(14):1-17.
- Çardak, Ç. S., & Selvi, K. (2019). Test development process for the instructional principles and methods course. *Mediterranean Journal of Educational Research*, 12(26):379-406. DOI: <https://doi.org/10.29329/mjer.2018.172.19>
- Deborah, A. C., Temitope, B., & Peter, A. O. (2021). Comparison of item difficulty and discrimination of pre and post university entrance examinations in Nigeria. *The Universal Academic Research Journal*, 3(1):1-9.
- Demir, N., Kızılay, E. & Bektaş, O. (2016). 7. sınıf çözümler konusunda başarı testi geliştirme: Geçerlik ve güvenilirlik çalışması [Development of an achievement test about solutions for 7th graders: A validity and reliability study]. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 10(1):209-237. DOI: <https://doi.org/10.17522/nefemed.52947>
- Eren, A. A., Tanık Önal, N., & Büyük, U. (2020). Elementler ve bileşikler konusu için geçerli ve güvenilir bir başarı testi geliştirme çalışması [Valid and reliable achievement test development study about element and compounds]. *Pearson Journal of Social Sciences & Humanities*, 6(6):151-167. DOI: <https://doi.org/10.46872/pj.52>
- Gülbahar, S. N. (2023). İstasyon tekniğinin 6. sınıf öğrencilerinin dolaşım sistemi konusundaki akademik başarılarına ve fen bilimleri dersine karşı tutumlarına etkisi. Master thesis, Dicle University, Diyarbakır.
- Gündüz, Y. (2009). İlköğretim 6, 7 ve 8. sınıf fen ve teknoloji sorularının ölçme araçlarına ve Bloom'un Bilişsel Alan Taksonomisine göre analizi [Analysis of primary school 6, 7 and 8. grades science

- and technology questions according to measurement scales and Bloom's taxonomy of the cognitive domain]. *Journal of the Faculty of Education of Van Yüzüncü Yıl University*, 6(2):150-165.
- Haladyna, T. M. (1997). Writing Test Items To Evaluate Higher Order Thinking. London: Allyn & Bacon.
- İdin, Ş. & Aydoğdu, C. (2016). Kuvvet ve hareket ünitesi başarı testi geçerlik ve güvenirlik araştırması [Validity and reliability study for the unit achievement test of force and motion]. *Journal of Research in Education, Science and Technology*, 1(1):14-33.
- İlhan, N., & Hoşgören, G. (2017). Fen bilimleri dersine yönelik yaşam temelli başarı testi geliştirilmesi: Asit baz konusu [Developing of context based achievement test towards science: Acids and bases]. *Fen Bilimleri Öğretimi Dergisi*, 5(2):87-110.
- İlhan, A. (2020). How to develop achievement test for geography teaching? *International Social Sciences Studies Journal*, 6(64):2702-2723. DOI: <https://doi.org/10.26449/sss.2377>
- Kargın, P. D., & Gül, Ş. (2021). Development of an achievement test for sixth grade body systems and health unit. *Ihlara Journal of Educational Research*, 6(1):1-26. DOI: <https://doi.org/10.47479/ihed.729412>
- Keçeci, G., Yıldırım, P., & Zengin, F. K. (2019). Sistemler akademik başarı testi: Geçerlik ve güvenirlik çalışması [Systems academic achievement test: Validity and reliability study]. *Ulusal Eğitim Akademisi Dergisi*, 3(1):96-114. DOI: <https://doi.org/10.32960/uead.514188>
- Kenan, O., & Özmen, H. (2014). Maddenin tanecikli yapısına yönelik iki aşamalı çoktan seçmeli bir testin geliştirilmesi ve uygulanması [Developing and implementing a two-tier multiple choice test for the particulate nature of matter]. *Eğitim ve Öğretim Araştırmaları Dergisi*, 3(3):371-378.
- Kızılcapan, O., & Bektaş, O. (2018). Fen eğitiminde başarı testi geliştirilmesi: Hücre bölünmesi ve kalıtım örneği [Development of an achievement test in science education: Example of cell division and inheritance]. *International Journal of Educational Sciences*, 2(1):1-18.
- Küçükahmet, L. (2002). Öğretimde Planlama ve Değerlendirme (13. Baskı) [Planning And Evaluation in Teaching (13th Edition)]. Ankara: Nobel Yayın Dağıtım.
- Ministry of National Education [MoNE] (2018). Fen bilimleri dersi öğretim programı (İlkokul ve Ortaokul 3, 4, 5, 6, 7 ve 8. Sınıflar) [Science course curriculum (Primary and Secondary School 3rd, 4th, 5th, 6th, 7th and 8th Grades)]. Available at: <https://mufredat.meb.gov.tr/Dosyalar/201812312311937-FEN%20B%C4%B0L%C4%B0MLER%C4%B0%20C3%96%C4%9ERET%C4%B0M%20PROGRAMI2018.pdf>
- Nacaroğlu, O., Bektaş, O., & Kızılcapan, O. (2020). Madde döngüleri ve çevre sorunları konusunda başarı testi geliştirme: Geçerlik ve güvenirlik çalışması [Developing an achievement test for matter cycles and environmental problems: A validity and reliability study]. *Kastamonu Eğitim Dergisi*, 28(1):36-51. DOI: <https://doi.org/10.24106/kefedergi.3374>
- Özay Köse, E., & Yıldırım, T. (2020). Dolaşım sistemi öğretiminde hikâye destekli etkinliklerin ilköğretim öğrencilerinin akademik başarı, tutum ve kalıcılığına etkisi [The effect of story supported activities on academic achievement and retention of primary school students in circulatory system teaching]. *Uluslararası Beşerî Bilimler ve Eğitim Dergisi*, 6(13):68-84.
- Özkan, H. H. (2006). Popüler kültür ve eğitim [Popular culture and education]. *Kastamonu Eğitim Dergisi*, 14(1):29-38.
- Özkan, E. B., & Eryılmaz Muştu, Ö. (2018). 8. sınıf basit makineler ünitesine yönelik başarı testi geliştirme: geçerlik ve güvenirlik çalışması [Developing an achievement test for the unit of simple machines in the 8th grade curriculum: validity and reliability study]. *Hitit University Journal of Social Sciences Institute*, 11(1):737-754. DOI: <https://doi.org/10.17218/hititsosbil.332294>
- Özkan, S. & Yedigaroğlu, M. (2020). Başarı testi geliştirme: asit-baz başarı testi geçerlik ve güvenirlik araştırması [Development of achievement test: A validity and reliability research for acid-base achievement test]. *Turkish Studies-Education*, 15(2):1141-1163. DOI: <https://doi.org/10.29228/TurkishStudies.41734>
- Öztemel, E. (2018). Eğitimde yeni yönelimlerin değerlendirilmesi ve eğitim 4.0 [Evaluation of new trends in education and educa-

- tion 4.0]. *Üniversite Araştırmaları Dergisi*, 1(1):25-30.
- Rao C., Kishan Prasad, H. L., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, 2:201-204. DOI: <https://doi.org/10.4103/2395-2296.189670>
- Saraç H. (2018). Fen Bilimleri dersi 'maddenin değişimi' ünitesi ile ilgili başarı testi geliştirme: Geçerlik ve güvenilirlik çalışması [Development of achievement test about science lesson 'change of matter' unit: Validity and reliability study]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(1):416-445. DOI: <https://doi.org/10.17240/aibuefd.2018-388815>
- Saylan-Kırmızıgül, A., & Kaya, H. (2019). 'Aynalarda yansıma ve ışığın soğrulması' konusunda geçerliği ve güvenilirliği sağlanmış başarı testi geliştirme çalışması [A study on developing a reliable and valid achievement test on 'light reflection in mirrors and light absorption' subject]. *Journal of Theoretical Educational Science*, 12(2):474-493. DOI: <https://doi.org/10.30831/akukeg.426792>
- Singh, C., & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics*, 71(6):607-617. DOI: <https://doi.org/10.1119/1.1571832>
- Şen, H. C. & Eryılmaz, A. (2011). Bir başarı testi geliştirme çalışması: Basit elektrik devreleri başarı testi geçerlik ve güvenilirlik araştırması [An achievement test development study: Reliability and validity investigation of simple electric circuits achievement test]. *Journal of the Faculty of Education of Van Yüzcüncü Yıl University*, 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. DOI: <http://doi.org/10.5539/jel.v6n2p254>
- Temizkan, M., & Sallabaş, M. E. (2015). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması [Comparison multiple choice tests and open ended questions while evaluating reading comprehension skills]. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 30:207-220.
- Timur, S., Doğan, F., İmer-Çetin, N., Timur, B., & Işık, R. (2019). Developing achievement test on cell subject for 6th grade: A validity and reliability study. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 48(2):1202-1219. Available at: <https://dergipark.org.tr/pub/cuefd>
- Treagust, D.F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2):159-169. DOI: <https://doi.org/10.1080/0950069880100204>
- Tunç, T., & Kılınç Alpat (2015). Elektrokimya da geçerlik ve güvenilirliği sağlanmış kimya başarı testinin geliştirilmesi [Developing a valid and reliable chemistry achievement test in electrochemistry]. *Journal of Higher Education and Science*, 5(3):347-360. DOI: <https://doi.org/10.5961/jhes.2015.136>
- Tutkun, Ö. F., & Aksoyalp, Y. (2010). 21. yüzyılda öğretmen yetiştirme eğitim programının boyutları [Dimensions of teacher training curricula in the 21st century]. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24:361-371.
- Üçüncü, G., & Sakız, G. (2020). Başarı testi geliştirme süreci: İlkokul dördüncü sınıf maddeyi tanıyalım ünitesi örneği [The phases of achievement test development: the case of fourth-grade introduction to matter unit]. *Kastamonu Education Journal*, 28(1):82-94. DOI: <https://doi.org/10.24106/kefdergi.3440>
- Yalınkılıç, F., & Gul, S. (2023). Development an achievement test on the subject of "Basic Compounds in the Structure of Living Things". *Science Insights Education Frontiers*, 18(2):2905-2925. DOI: <https://doi.org/10.15354/sief.23.or420>
- Yalçınkaya, I. (2018). Altıncı sınıf seviyesinde argümantasyon odaklı etkinliklerle dolaşım sistemi konusunun öğretiminin akademik başarıya, kavramsal anlamaya ve argümantasyon seviyelerine etkisi. Master thesis, Pamukkale University, Denizli.
- Yıldız, S., Keçeci, G., & Kırbağ-Zengin, F. (2019). Dengeli beslenme akademik başarı testi: geçerlik ve güvenilirlik araştırması [Balanced nutrition academic achievement test: Validity and reliability research]. *Yüzcüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 16(1):848-868. DOI:

Akgun & Gul. (Turkey). Achievement Test of 'Circulatory System' for Sixth Graders.

<http://doi.org/10.23891/efdyyu.2019.144>

Received: 09 November 2023

Revised: 24 November 2023

Accepted: 10 December 2023