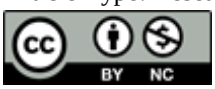
International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

Development of an Achievement Test for the "Seasons and Climate" Unit of the Eighth Grade Science Course in Secondary School

Emine Yurtyapan[1], Ayşe Gül Çirkinoğlu Şekercioğlu[2]
[1]Ministry of National Education, 0000-0003-4630-8601
[2]Balıkesir University. 0000-0001-9474-2977

Article History
Received: 11.07.2023
Received in revised form: 11.10.2023
Accepted: 06.12.2023
Article Type: Research Article



To cite this article:

Yurtyapan, E., and Çirkinoğlu Şekercioğlu, A. G. (2023). Development of an achievement test for the "Seasons and Climate" unit of the eighth grade science course in secondary school. *International Journal of Contemporary Educational Research, 10*(4), 893-916. https://10.52380/ijcer.2023.10.4.501

International Journal of Contemporary Educational Research
Volume 10, Number 4, December 2023, 893-916 ISSN: 2148-3868

Development of an Achievement Test for the "Seasons and Climate" Unit of the Eighth Grade Science Course in Secondary School

Emine Yurtyapan[1], Ayşe Gül Çirkinoğlu Şekercioğlu [2]
[1] Ministry of National Education
[2]Balıkesir University

Abstract

This study aims to develop a valid and reliable achievement test to measure the achievement of eighth-grade students in the "Seasons and Climate" unit of the science course. The research was carried out using a survey model. The study group consists of 298 eighth-grade students studying in various cities in Turkey in the 2020–2021 academic year. The content validity of the prepared questions was examined with a table of specifications. An achievement test of 25 questions was obtained by making item analyses with the data obtained from the pre-pilot study. After the pilot study of this test, the reliability analysis of the test was made, and the KR-20 internal consistency coefficient was 0.88; the KR-21 internal consistency coefficient was determined to be 0.86. In addition, while the mean item difficulty index of the test was 0.61, the mean discrimination index was calculated as 0.50. As a result, a valid and reliable achievement test was obtained for the metacognitive learning outcomes of the "Seasons and Climate" unit in the eighth grade 2018 science curriculum. In addition, since the science curriculum of the Ministry of National Education is updated from time to time according to international developments, it is possible to say that the achievement test developed is for international readers. In the way followed in this study, researchers can develop multiple-choice achievement tests that measure metacognitive learning outcomes for different subjects.
Keywords: Achievement test; Science education; Seasons and climate; Test development

Introduction

In education, the measurement process is carried out through exams to evaluate the knowledge of students in various courses and subjects. For this purpose, many alternative or traditional assessment and evaluation tools can be used. While alternative assessment and evaluation tools are applied to students during the teaching process, traditional assessment and evaluation tools are applied at the end of the teaching process and are mostly preferred for measuring cognitive domain learning. In this study, it was aimed at developing a measurement tool for cognitive domain learning outcomes. For this reason, it is thought that the measurement tool to be developed is more suitable for the measurement tools in the traditional approach.

Traditional cognitive domain assessment and evaluation tools can be handled under three headings: written, oral, and objective exams (Başol, 2019). Each of these exam types has strengths and weaknesses. Particularly, the subjectivity in the scoring of the written and oral exams is an important negative feature in terms of scoring reliability. For this reason, written and oral exams are not preferred in exams where large groups of students will be placed in a higher education institution to measure their cognitive domain success. On the other hand, since objective exams are read with an optical reader or perforated answer key, they are highly reliable (Başol, 2019). Therefore, it is thought that it would be more appropriate to design the measurement tool, which is aimed at being developed within the scope of this study, as an objective test. Questions used in objective exams are true-false, fill-in-the-blank (complementary), matched, and multiple-choice test items. The most commonly used of these questions are multiple-choice tests (Demir et al., 2016; Ogan Bekiroğlu, 2004).The fact that the chance factor is high in finding answers to true-false, matched questions, and the fact that questions with fill-in-the-blank measure only knowledge-level behaviors affect the validity negatively. For this reason, multiple-choice test questions are used in many national and international objective exams in which students' cognitive achievement is measured (Akbulut & Çepni, 2013; Pressley et al., 1997). Considering all these features related to the questions used in objective exams, it is thought that it would be appropriate to use multiple-choice

[1]Corresponding Author: *Emine Yurtyapan, emineyurtyapan@hotmail.com*

questions in the measurement tool that will be developed to measure the cognitive achievement status of students within the scope of this study.

Studies in the related literature show that multiple-choice tests are one of the most frequently used measurement tools after interviews in revealing thoughts about understanding a certain subject or concept (Kempa, 1986; Ogan Bekiroğlu, 2004). In this respect, according to Özçelik (1998), the measurement tool with the most superior measurement features is the multiple-choice test. On the other hand, in some studies in the literature, it is stated that while multiple-choice tests are more suitable for measuring the learning outcomes of the lower stages of the cognitive domain, such as remembering, understanding, and application, they are insufficient for measuring life skills such as creativity and self-expression (Başol, 2019; Küçükahmet, 2002). Considering that the majority of the competencies and learning outcomes in the science course secondary school curriculum of the Ministry of National Education [MoNE] (2018) are high-level, measuring these learning outcomes with multiple-choice tests may constitute an important limitation in terms of validity. Therefore, while designing an objective exam in which high-level learning outcomes consisting of multiple-choice questions will be measured, preparing the questions according to Bloom's metacognitive steps (analysis, synthesis, and evaluation) may be effective in minimizing this limitation (Küçükahmet, 2002). In addition, the questions in the multiple-choice tests are suitable for the purpose, scope, and learning outcomes of the course, and classifying them according to the Bloom taxonomy also increases the quality of the measurement tool in terms of reliability and validity (Kızkapan ve Bektaş, 2018). Therefore, in this study aiming to design a measurement tool with high validity and reliability by using multiple-choice questions, questions were classified according to Bloom's Taxonomy, and mostly questions measuring metacognitive skills were included in the test.

In the related literature, there are many achievement tests prepared for the Science course (Boz et. al., 2022; Çirkinoğlu Şekercioğlu & Yılmaz Akkuş, 2019; Dağ & Karamustafaoğlu, 2023; Doğru & Çepni, 2023; Karamustafaoğlu & Tutar, 2016; Sontay & Karamustafaoğlu, 2020; Türk, 2015; Varzikioğlu, 2023). Achievement tests in some of these studies are about astronomy (Çirkinoğlu Şekercioğlu & Yılmaz Akkuş, 2019; Karamustafaoğlu & Tutar, 2016; Sontay & Karamustafaoğlu, 2020; Türk, 2015; Varzikioğlu, 2023). Astronomy subjects are included in the "Earth and Universe" learning area at all grade levels in the MoNE 2018 secondary school science curriculum. When the studies in the related literature are examined, most of the achievement tests developed for astronomy topics in the "Earth and Universe" learning area are at the fifth, sixth, and seventh grade levels (Çirkinoğlu Şekercioğlu & Yılmaz Akkuş, 2019; Karamustafaoğlu & Tutar, 2016; Sontay & Karamustafaoğlu, 2020; Türk, 2015; Varzikioglu, 2023). When the MoNE (2018) science curriculum is examined, the "Seasons and Climate" unit is included within the scope of the "Earth and Universe" learning area at the eighth grade level. This unit is focused on the topics "Formation of Seasons" and "Climate and Air Movements." Within the scope of "The Formation of the Seasons," it is intended that the students discover the causes of the seasons by estimating. The main factors in the formation of seasons are the annual motion of the Earth and its axis tilt. As a result of the tilt of the Earth's axis, the change in the angle of incidence of the Sun's rays on the Earth throughout the year causes seasons to occur. However, the change in the angle of incidence of the Sun's rays on the Earth throughout the year basically causes the formation of seasons and also causes many other situations, such as day and night periods, changes in the length of the shadow, periastron, apastron, and equinox. Therefore, all these situations mentioned within the scope of the "Formation of Seasons" topic are touched upon. Another topic included in the "Seasons and Climate" unit is "Climate and Air Movements." Climate and weather movements (events) are two related concepts. Due to the similarities between the two concepts, they are often confused by students or considered synonymous. For this reason, the similarities and differences between these two concepts are emphasized in the subject of "Climate and Air Movements." Climate can be defined as the average air movement over large areas over many years. Weather events occur as a result of changes in factors such as temperature, humidity, and pressure. Wind, rain, snow, etc. are weather events. The meanings of the sciences of climatology and meteorology are given by making comparisons in terms of similarities and differences in the subject of "climate and weather movements." In this way, branches of science that deal with two different concepts, such as climate and weather movements, can be learned. In addition, environmental events such as climate change, global warming, and the greenhouse effect are also addressed within the scope of this subject. When the relevant literature is examined, it is seen that measurement tools for different grade levels are mostly developed by considering these subjects separately (Akşit & Şahin, 2011; Bolat & Altınbaş, 2018; Uzunöz & Buldan, 2012). When the studies on the formation of the seasons are examined, it is seen that mostly open-ended or closed-ended questions are asked in the developed tests (Bolat & Altınbaş, 2018; Gülen, 2019; Türk & Kalkan, 2015; Türk et al., 2016). In some of these studies and some studies on general astronomy, multiple-choice questions about the formation of the seasons were also used by turning them into open-ended questions (Gülen, 2019; Schoon, 1992). In the study conducted by Bolat and Altınbaş (2018), using the parallel mixed method, it was tried to determine the understanding of the prospective teachers studying in different departments about the subject of seasons. In the

study conducted by Gülen (2019), the achievements of the eighth grade students in the "Seasons and Climate" unit were determined with six open-ended questions, and the student learning outcomes were interpreted. In these studies, it has been observed that there are many misconceptions among the students about the formation of the seasons. In this respect, it can be said that most of the studies on the formation of seasons using open-ended questions are aimed at identifying misconceptions (Bolat & Altınbaş, 2018; Gülen, 2019; Öksüz & Güven Demir, 2019; Schoon, 1992; Sneider et al., 2011; Yilmaz & Bulunuz, 2019). Therefore, in the related literature, there are a limited number of studies in which achievement tests consisting of multiple-choice questions on the subjects of "Seasons Formation" and "Climate and Weather Movements" within the scope of the eighth grade "Seasons and Climate" unit in the science curriculum of MoNE (2018) are available (Birgin & Özcan, 2022; Geren, 2022; Yanardağ, 2021). In the study conducted by Birgin and Özcan (2022), a multiple-choice achievement test was tried to be developed to measure the knowledge of eighth grade students about the formation of the seasons. However, in the achievement test developed by Birgin and Özcan (2022), there are no questions on the subject of "Climate and Weather Movements" within the scope of the "Seasons and Climate" unit. On the other hand, in the studies conducted by Geren (2022) and Yanardağ (2021), two different multiple-choice achievement tests were developed, including all the topics of the "Seasons and Climate" unit. However, in the achievement tests of Geren (2022) and Yanardağ (2021), there are sample questions of the High School Entrance Examination [HSEE] and learning outcome comprehension test questions, which are considered to be highly valid and reliable in various sources of MoNE. It is stated that due to the COVID-19 pandemic, while the achievement tests were developed in the studies of Geren (2022) and Yanardağ (2021), a pilot study could not be conducted and only expert opinion was taken. In addition, questions in the achievement tests developed by Geren (2022) and Yanardağ (2021) were selected without any cognitive classification. Considering all these features, it can be said that the achievement tests developed for the "Seasons and Climate" unit in the literature have limitations in terms of validity and reliability. Therefore, it is thought that the "Seasons and Climate" achievement test developed within the scope of this study will contribute to the relevant literature by eliminating these deficiencies.

Problem Statement

The problem statement of this research, "Is the achievement test developed for the "Seasons and Climate" unit of the eighth grade science lesson valid and reliable?" has been determined.Depending on the determined problem statement, the sub-problems of the research are presented below.

Sub-Problems of the Research

1. Is the achievement test developed for the "Seasons and Climate" unit of the middle school eighth grade science lesson valid?
2. Is the achievement test developed for the "Seasons and Climate" unit of the middle school eighth grade science lesson reliable?

Purpose of the Research

The purpose of this study is to develop a valid and reliable achievement test for the eighth grade "Seasons and Climate" unit within the scope of the MoNE (2018) science curriculum.

Method

In this section, information about the research model, sample, the development process of the data collection tool, and the analysis of the data are included.

Model of the Research

This study was carried out in accordance with the survey model since it was aimed at developing a valid and reliable achievement test for the eighth grade "Seasons and Climate" unit of secondary school. Survey studies are studies carried out by the researcher to determine the current situation without any intervention (Creswell, 2012). This method allows making predictions about the universe as a result of the analysis of the data obtained from the sample (Özmen & Karamustafaoğlu, 2019). In this study, the pilot achievement test developed by the researchers in accordance with the learning outcomes comprehension test of the "Seasons and Climate" unit was applied to a sample of eighth grade secondary school students. In this way, the validity and reliability were calculated by analyzing the data obtained for each item. As a result, the study was carried out in accordance with

the survey model, since it was aimed at obtaining a valid and reliable achievement test to measure the academic achievement of the eighth grade "Seasons and Climate" unit.

Sample

While developing the achievement test for the research, two pilot studies were made. The sample of the study consists of 298 secondary school eighth grade students: 223 in the pre-pilot study and 75 in the pilot study. While determining the sample, the "maximum diversity sampling" method, one of the purposive sampling methods, was used. Purposeful sampling allows for an in-depth investigation of information-rich situations in line with the purpose of the study (Büyüköztürk et al., 2015). As a result of this study, it is aimed at developing a valid and reliable "Seasons and Climate" achievement test applicable to secondary school eighth grade students. In order for the achievement test to be obtained at the end of the research to be generalizable to all eighth grade students, it is necessary to apply this test to groups that are similar but different from each other during the test development process. As a matter of fact, in maximum diversity sampling, it is essential to work by determining different situations similar to the problems examined in the universe (Büyüköztürk et al., 2015). As a result, the purpose of choosing the sample in this way is to reflect the sample diversity at the highest level for the achievement test to be applied (Yıldırım & Şimşek, 2016). For this reason, it was thought that it would be appropriate to determine the study group with maximum variation sampling, one of the purposive sampling methods. The distribution of the students in the sample according to the provinces of education is presented in Table 1.

Table 1. Distribution of the students in the sample according to the provinces of education

Province	Number of Students (N)	Percentage (%)
Kocaeli	179	60.07
Balıkesir	21	7.05
Bursa	13	4.36
İstanbul	12	4.03
Hatay	6	2.01
Muğla	6	2.01
Denizli	5	1.68
Kahramanmaraş	4	1.34
Mersin	4	1.34
Ankara	4	1.34
Yalova	3	1.01
Gaziantep	3	1.01
Tokat	3	1.01
Bartın	3	1.01
Antalya	3	1.01
Aksaray	2	0.67
Niğde	2	0.67
Zonguldak	2	0.67
Siirt	2	0.67
Konya	2	0.67
İzmir	2	0.67
Samsun	2	0.67
Nevşehir	2	0.67
Şanlıurfa	1	0.34
Adana	1	0.34
Çorum	1	0.34
Düzce	1	0.34
Sakarya	1	0.34
Afyon	1	0.34
Kırşehir	1	0.34
Kırıkkale	1	0.34
Sivas	1	0.34
Bitlis	1	0.34
Rize	1	0.34
Ordu	1	0.34
Kayseri	1	0.34
Total	298	100

Developing the Seasons and Climate Achievement Test

The "Seasons and Climate" academic achievement test was developed by considering Güler's (2021) test development steps. Therefore, the development stages of the "Seasons and Climate" academic achievement test are summarized as follows:

Determining the Purpose of the Test

While preparing the test, it was aimed to take into account the learning outcomes of the "Seasons and Climate" unit in the eighth grade "Earth and Universe" subject area in the MoNE (2018) Science Curriculum, to take into account the level of the students, and to benefit the teachers.

Determining the Scope of the Test

The scope of the test was determined according to the information on the eighth grade "Seasons and Climate" unit in the Science Curriculum of the MoNE (2018). General information on the "Seasons and Climate" unit is given in Table 2.

Table 2. General information on seasons and climate units

Unit Name	Subject Area Name	Numbers of Learning Outcomes	Course Hours	Percentage of Learning Outcomes (%)
Seasons and Climate	Earth and the Universe	3	14	9.7

When the information in Table 2 is examined, there are three learning outcomes belonging to the Seasons and Climate unit within the scope of the Earth and Universe subject area. The recommended time for teaching these learning outcomes is 14 course hours.
The subjects, concepts, suggested duration, learning outcomes, and warnings of the Seasons and Climate Unit are given in Table 3.

Table 3. Subjects, concepts, suggested duration, learning outcomes, and warnings of the seasons and climate unit

Subjects	Concepts	Suggested Duration	Learning Outcomes	Warnings
F.8.1.1. Formation of the Seasons	-Earth's axis of rotation -Orbital plane -Heat Energy -Seasons	Eight lesson hours	F.8.1.1.1. It makes predictions about the formation of the seasons.	a. It is mentioned that the Earth is the axis of rotation. b. The relationship between the Earth's axis of rotation and the plane of its orbit around the Sun is mentioned. c. The effect of the amount of energy falling on a unit surface of light on the seasons is mentioned.
F.8.1.2. Climate and Weather Movements	-Climate -Climatology -Climatologist -Global climate change	Six lesson hours	F.8.1.2.1. Explain the difference between climate and weather events. F.8.1.2.2. It says that climate science (climatology) is a branch of science, and experts working in this field are called climate scientists (climatologists).	

At this stage, where the scope of the test is determined, the type, number, and application time of the questions are decided. The questions in the test were prepared with four options in accordance with the multiple-choice

question type. While preparing the questions, the learning outcomes in the 2018 Science Curriculum (MoNE, 2018) and the warnings about the learning outcomes in Table 3 were taken into account. In addition, the contents of the eighth grade textbook (Yancı; 2019), Education and Informatics Network [EIN], were examined. By examining the national and international studies (Bolat & Altınbaş, 2018; Sneider et al., 2011) in the related literature, it was tried to determine the concepts and learning outcomes that students had difficulty learning within the scope of the Seasons and Climate unit. In the study conducted by Sneider et al. (2011), they draw attention to the importance of asking questions about the results along with the reasons for the formation of the seasons in determining the knowledge level of the students about the seasons. Therefore, in this study, which aims to develop a valid and reliable achievement test for the "Seasons and Climate" unit, care was taken to include questions about the causes and consequences of the formation of the seasons. An item pool consisting of 75 questions with four options for these learning outcomes and concepts was created by making use of various national exam preparation books in EIN. However, this study was carried out in the 2020–2021 academic year, when distance education was available due to the COVID-19 pandemic. Therefore, the pre-pilot study of the achievement test was decided to be carried out by the researchers online via the Google Form. Considering situations such as the pre-pilot study being conducted online and the questions being aimed at metacognitive learning outcomes, it was seen that 75 questions were not appropriate according to the grade level and attention span of the students. For this reason, 49 questions were determined from the 75-item pool by the researchers before the pilot study.

Validity of the Test

A table of specifications has been prepared for the content validity of the achievement test (see Table 11). The learning outcomes in the MoNE (2018) curriculum for the eighth grade "Seasons and Climate" unit are metacognitive domain learning outcomes. In addition, these learning outcomes have general expressions. Therefore, in order to make the learning outcomes of the questions more specific, the sub-learning outcomes for the questions were written by the researchers while preparing the specification table and classified according to Bloom's Taxonomy.

Then, 49 questions were determined by the researchers in accordance with the learning outcomes from the item pool consisting of 75 questions from all levels, but mostly by considering metacognitive domain skills. In order to ensure content validity, expert opinion was obtained in two stages for the classification made according to Bloom's taxonomy and face validity. In both stages, opinions were received from a total of four experts, two science teachers, and two faculty members in the science education department. The first expert opinion is for 49 questions determined by the researchers out of 75 questions. As a result of the suggestions and corrections from the expert opinions, a preliminary achievement test of 49 questions was created for the pre-pilot study. After the pre-pilot study, item analyses were carried out. As a result of the analysis, it was seen that some items should be removed from the test or improved. In addition, the number of items was reduced to 25 as a result of the item analysis carried out by the researchers, since 49 questions were too many to apply in a class hour. While reducing the number of items in the test, the researchers paid special attention not to affect the distribution of the items according to Bloom's taxonomy in terms of content validity. Therefore, a second expert opinion was needed. The second expert opinion was taken for the final version of the academic achievement test, consisting of 25 questions. Some formal and verbal corrections were made in the 11th, 15th, 18th, 26th, 45th, and 46th items in order to better understand the questions with the suggestions from these expert opinions. Thus, the academic achievement test for the eighth grade "Seasons and Climate" unit, consisting of 25 questions in total, was made ready.

Pre-Pilot Study

This academic achievement test development study was carried out in the 2020–2021 academic year, when distance education was held due to the COVID-19 pandemic. For this reason, a pre-pilot was conducted online with 223 eighth grade students to determine whether the questions in the achievement test were understandable by the students before the pilot study and whether the time to answer the questions was sufficient during the application of the test. Students have the right to take this online exam, which consists of 49 questions, and the exam duration is determined to be 120 minutes. As a result of the pre-pilot study, item analyses were made, and 1st, 2nd, 3rd, 4th, 7th, 9th, 13th, 14th, 17th, 19th, 20th, 21th, 23th, 24th, 30th, 32th, 33th, 35th, 37th, 38th, 40th, 41th, 47th, and 49th items were excluded from the test. After these test items were removed from the test, the second expert opinion was taken. According to expert opinions, the "Seasons and Climate" achievement test with 25 questions was made ready for pilot study by making formal and verbal adjustments in the 11th, 15th, 18th, 26th, 45th, and 46th items.

Pilot Study

The "Seasons and Climate" achievement test, consisting of 25 multiple-choice questions, was applied to 78 eighth grade students in the 2022–2023 academic year. The duration of the exam is determined as one lesson hour. As a result of the analysis, the tests of three students were not analyzed due to incorrect or incomplete marking. As a result, the test answers of 75 students were included in the analysis.

Reliability and Item Analysis

The "Seasons and Climate" achievement test consists of multiple-choice test questions. While analyzing the data obtained from the developed achievement test, one point was given for correct answers and zero points for wrong answers or blanks. While the "Seasons and Climate" achievement test was being developed, two applications were made as a preliminary pilot study and a pilot study, so item and reliability analysis were performed twice in order to compare the first and final versions of the test. KR-20 and KR-21 internal consistency coefficients were calculated in the reliability analysis of the questions in the developed achievement test. In this study, "item discrimination index" and "item difficulty index" were calculated for each item in the test by using two different methods in Classical Test Theory for item analysis. It is a measure of being able to distinguish those who know the feature that is wanted to be measured with a test item from those who do not (Güler, 2021). The rate of correct answers to each item in a test is defined as the "item difficulty index" (Hasançebi et al., 2020). While the simple method (lower-upper group method) was used in the first item analysis, the method described as the "second method" by Güler (2021) was applied in the other. While the simple method (lower-upper group method) is preferred when the number of students is 300–400 or more, the second method can be preferred when 60–70 or less (Güler, 2021). In the simple method (lower-upper group method), the scores of the students in the achievement test are ordered from the highest score to the lowest score. 27% of the people with the highest scores are defined as the "upper group," while 27% of the people with the lowest scores are determined as the "lower group" (Beuchert & Mendoza, 1979). Item analyses are made by considering these groups. The item difficulty index is indicated by Pj and when the simple method is used, the item difficulty index is calculated with the formula in the Table 4.

Table 4. Item difficulty index formula

Formula	Variables in the formula
$Pj = \left(\frac{n(cu) + n(cl)}{N}\right)$	Pj: Difficulty index of the item $n(cu)$: The number of those who answered the item correctly in the upper group $n(cl)$: The number of those who answered the item correctly in the lower group N: Total number of students in the upper group and lower group

The item discrimination index is shown, and in the simple method, the item discrimination index is calculated with the formula given in Table 5.

Table 5. Item discrimination index formula

Formula	Variables in the formula
$r_{jx} = \left(\frac{n(cu) - n(cl)}{N}\right)$	r_{jx}: Item discrimination index $n(cu)$: The number of those who answered the item correctly in the upper group $n(cl)$: The number of those who answered the item correctly in the lower group N: Number of students in any of the groups

In the second method, the scores of all students to whom the achievement test was applied are included in the calculation. When this method is used, the formula in Table 6 is applied to calculate item difficulty.

Table 6. Item difficulty index formula

Formula	Variables in the formula
$Pj = \left(\frac{n(c)}{N}\right)$	Pj: Difficulty index of the item $n(c)$: Number of people who answered the item correctly N: Total number of students

When the second method is used, the item discrimination index should be calculated using the formula presented in Table 7.

Table 7. Item discrimination index formula

Formula	Variables in the formula
$r_{jx} = \frac{(\underline{X}(d) - \underline{X})}{S_x} \times \sqrt{\frac{pj}{qj}}$	r_{jx}: Item discrimination index $n(cu)$: The number of those who answered the item correctly in the upper group $n(cl)$: The number of those who answered the item correctly in the lower group N: Number of students in any of the groups

Since item discrimination is related to the reliability of the item, it is the most important statistic of an item, and whether the item will be included in the test is decided by looking at the item discrimination index (Güler, 2021). The item discrimination index of the items in the test developed in this study was evaluated according to Table 8 (Akbulut & Çepni, 2013).

Table 8. Item Discrimination Index and Evaluation

Item Discrimination Index	Evaluation
0.40 and higher	Very good item
0,30-0,39	Pretty good item
0,20-0,29	It must be corrected.
0.19 and less	Very weak item. It should be removed from the test.

When the relevant literature is examined, it is desired that the average discrimination index of an ideal test should be as close to +1.00 as possible (Güler, 2021).

The difficulty of the items in the "Seasons and Climate" achievement test was evaluated according to Table 9 (Akbulut & Çepni, 2013).

Table 9. Item Difficulty Index and Evaluation

Item Difficulty Index	Evaluation
0.29 and higher	Difficult
0.30- 0.49	Medium difficulty
0.50- 0.69	Easy
0.70- 1.00	Very easy

In many sources in the related literature, it is stated that the average difficulty of a test is around 0.50 and that there should be questions from all levels in the test (Güler, 2021; Karamustafaoğlu & Tutar, 2016).

After calculating the item discrimination and difficulty indexes in the study, reliability was calculated. There are many formulas for calculating the reliability of a measurement tool. If the items in the measurement tool have only one correct answer, the KR-20 or KR-21 internal consistency coefficient should be preferred to calculate the reliability of the test. Therefore, KR-20 or KR-21 internal consistency coefficients are calculated in studies to develop multiple choice achievement tests (Boz et al., 2022; Yazıcı et al., 2022). For this reason, in this study in which the "Seasons and Climate" achievement test was developed, the reliability of the test was calculated by using both the KR-20 and KR-21 internal consistency coefficients, the formulas of which are given in Table 10.

Table 10. Formulas of internal consistency coefficients KR-20 and KR-21

Formula	Variables in the formula
$KR-20 = \frac{k}{k-1}\left(1 - \frac{\sum {s_j}^2}{{S_x}^2}\right)$	k: Number of items in the test $\sum {s_j}^2$: Sum of item variances ${s_j}^2 = p_j.q_j$: Variance of an item p_j: Item difficulty index $q_j = (1 - p_j)$: Percentage of those who answered the item incorrectly ${S_x}^2$: Variance of the test
$KR-21 = \frac{k}{k-1}\left(1 - \frac{k.\underline{X} - \underline{X}^2}{k.{S_x}^2}\right)$	k: Number of items in the test $\underline{X}$: Arithmetic mean ${S_x}^2$: Variance of the test

Results

In this section, the findings obtained through the validity and reliability study after the application of the multiple-choice achievement test prepared for the "Seasons and Climate" unit in the Earth and Universe learning area of the eighth grade Science course are included.

In order to ensure the content validity of multiple-choice achievement tests, a table of specifications should be prepared. In this study, the cognitive domain learning outcomes of the related unit in the MoNE (2018) secondary school science curriculum were examined and classified according to Bloom's taxonomy. However, since these learning outcomes are metacognitive learning outcomes and contain very general skill expressions, sub-learning outcomes were written by the researchers according to Bloom's taxonomy and associated with the questions. The table of specifications prepared in this way is presented in Table 11.

Table 11. Specification table

Learning Outcomes	Knowledge	Understanding	Application	Analysis	Synthesis	Evaluation
F.8.1.1.1. It makes predictions about the formation of the seasons.						
It expresses that the seasons are formed as a result of the earth's axial tilt. (Knowledge)*	1	-	-	-	-	-
It compares the days and nights to be experienced on certain dates by looking at the positions of the countries according to the tropics on the world model. (Application)*	-	-	2,3	-	-	-
It shows the position of the country on the world model based on the events and information experienced on a certain date. (Application)*	-	-	4,5	-	-	-
Explains the events that occur in different hemispheres from the equinox date to the solstice date, depending on the Earth's axial tilt.(Understanding)*	-	6	-	-	-	-
Explains the locations and dates when the sun's rays reach the earth at an oblique angle. (Understanding)*	-	7, 40	-	-	-	-
It compares the events that can be experienced in different hemispheres by looking at the angle of incidence of sunlight in the tropics on the model. (Analysis)*	-	-	-	8, 36	-	-
Establishes a relationship between the average precipitation graph and the model showing the Earth's movement around the Sun. (Analysis)*	-	-	-	9	-	-
Designs an experiment to show the effect of the change in the angles of the sun's rays falling on the earth on the formation of the seasons. (Synthesis)*	-	-	-	-	10	-
It questions the effect of the distance between them on the formation of the seasons during the movement of the Earth around the Sun. (Evaluation)*	-	-	-	-	-	11

Evaluates the events that may occur in the absence of the Earth's axial tilt. (Evaluation)*	-	-	-	-	-	15
Draws a graph of daylight hours in a year using tabular data on a country's solstice and equinox dates. (Synthesis)*	-	-	-	-	34	-
Interprets and compares two graphs showing the variation of the angle of incidence of sunlight according to hours. (Analysis)*	-	-	-	35	-	-
It creates a model that shows the location of the cities by using the data in the graphs showing the day and nighttime on different dates. (Synthesis)*	-	-	-	-	37	-
It establishes a relationship between the data in the shadow length change graph of an object during the day and the date and location. (Analysis)*	-	-	-	38	-	-
Evaluates the results of the experiment designed for the effect of the angle of incidence of sunlight on temperature. (Evaluation)*	-	-	-	-	-	39
It refers to the solstice and equinox dates in different hemispheres. (Knowledge)*	41, 42					
Expresses the gases that make up the structure of the atmosphere and the greenhouse effect. (Knowledge)*	43					
Question the reasons for the formation of the seasons with an experiment. (Evaluation)*						45
It creates a model that shows the location of the regions by using the data in the graph showing the amount of energy per unit surface on the date of the solstice. (Synthesis)*					46	
Explains the effect of the change in the angle of incidence of the sun's rays on the temperature depending on the axial tilt of the earth. (Understanding)*		47				
Relates the seasonal temperature values of countries located in different hemispheres to the world model and the angles of incidence of the sun's rays. (Analysis)*				48		
F.8.1.2.1. Explain the difference between climate and weather events.						
It shows how factors affecting wind formation are used in the design of some technological tools. (Application)*	-	-	12	-	-	-
Designs experiments or models to show wind formation and direction. (Synthesis)*	-	-	-	-	13 ,1 4	-

Interprets by comparing the data given in the precipitation-temperature graph or in the form of a table. (Analysis)*	-	-	-	17, 18, 19, 20	-	-
Evaluates the results of the experiment designed for the formation of the greenhouse effect. (Evaluation)*	-	-	-	-	-	21
Evaluates the results of the experiment designed for the formation of weather events. (Evaluation)*	-	-	-	-	-	22
Shows the factors affecting precipitation patterns, wind formation, direction, and intensity on a map, model, table, or figure. (Application)*	-	-	23, 24, 25			
It creates a meteorological warning map in line with the data.(Synthesis)*	-	-	-	-	26	-
Establishes example sentences describing the characteristics of climate and weather. (Understanding)*	-	27, 32, 33	-	-	-	-
Evaluates the results of the experiment designed for the factors affecting the wind intensity. (Evaluation)*	-	-	-	-	-	28, 29
Explain the effect of temperature, humidity, and pressure on the formation of weather events. (Understanding)*	-	30	-	-	-	-
Explain the difference between dew and frost formation. (Understanding)*	-	31	-	-	-	-
Explain the consequences of global climate change.(Understanding)*	-	44,49	-	-	-	-
F.8.1.2.2. It says that climate science (climatology) is a branch of science, and experts working in this field are called climate scientists (climatologists).						
It refers to the experts working in the field of climate science as "climatologists." (Knowledge)	16	-	-	-	-	-

* Added learning outcomes later

When Table 11 is examined, it is seen that some of the learning outcomes are found in the eighth grade "Seasons and Climate" unit in the Science course curriculum of the MoNE (2018). Other learning outcomes are the sub-learning outcomes developed by the researchers for the subjects and concepts of the "Seasons and Climate" unit, depending on these learning outcomes. Only one of these sub-learning outcomes was written by dividing the behaviors indicated by the learning outcomes in the MoNE (2018) Science Curriculum, and the others were written by the researchers by associating the unit content with the learning outcomes in the curriculum." F.8.1.2.2. It says that climate science (climatology) is a branch of science, and experts working in this field are called climate scientists." The learning outcome in the MoNE (2018) curriculum has been shortened and arranged as a sub-acquirement. This sub-learning outcome has been written as "It refers to the experts working in the field of climate science as "climatologists"." On the other hand, the sub-learning

outcomes written by the researchers considering the unit content and the learning outcomes in the curriculum are as follows:

- It expresses that the seasons are formed as a result of the earth's axial tilt.
- It compares the days and nights to be experienced on certain dates by looking at the positions of the countries according to the tropics on the world model.
- It shows the position of the country on the world model based on the events and information experienced on a certain date.
- Explains the events that occur in different hemispheres from the equinox date to the solstice date, depending on the Earth's axial tilt.
- Explains the locations and dates when the sun's rays reach the earth at an oblique angle.
- It compares the events that can be experienced in different hemispheres by looking at the angle of incidence of sunlight in the tropics on the model.
- Establishes a relationship between the average precipitation graph and the model showing the Earth's movement around the Sun.
- Designs an experiment to show the effect of the change in the angles of the sun's rays falling on the earth on the formation of the seasons.
- It questions the effect of the distance between them on the formation of the seasons during the movement of the Earth around the Sun.
- Evaluates the events that may occur in the absence of the Earth's axial tilt.
- Draws a graph of daylight hours in a year using tabular data on a country's solstice and equinox dates.
- Interprets and compares two graphs showing the variation of the angle of incidence of sunlight according to hours.
- It creates a model that shows the location of the cities by using the data in the graphs showing the day and nighttime on different dates.
- It establishes a relationship between the data in the shadow length change graph of an object during the day and the date and location.
- Evaluates the results of the experiment designed for the effect of the angle of incidence of sunlight on temperature.
- It refers to the solstice and equinox dates in different hemispheres.
- Expresses the gases that make up the structure of the atmosphere and the greenhouse.
- Questions the reasons for the formation of the seasons with an experiment.
- It creates a model that shows the location of the regions by using the data in the graph showing the amount of energy per unit surface on the date of the solstice.
- Explains the effect of the change in the angle of incidence of the sun's rays on the temperature depending on the axial tilt of the earth.
- Relates the seasonal temperature values of countries located in different hemispheres to the world model and the angles of incidence of the sun's rays.
- It shows how factors affecting wind formation are used in the design of some technological tools.
- Designs experiments or models to show wind formation and direction.
- Interprets by comparing the data given in the precipitation-temperature graph or in the form of a table.
- Evaluates the results of the experiment designed for the formation of the greenhouse effect.
- Evaluates the results of the experiment designed for the formation of weather events.
- Shows the factors affecting precipitation patterns, wind formation, direction, and intensity on a map, model, table, or figure.
- It creates a meteorological warning map in line with the data.
- Establishes example sentences describing the characteristics of climate and weather.
- Evaluates the results of the experiment designed for the factors affecting the wind intensity.
- Explain the effect of temperature, humidity, and pressure on the formation of weather events.
- Explain the difference between dew and frost formation.
- Explain the consequences of global climate change.

Considering that these sub-learning outcomes are necessary for the content validity of the subject, they were written in line with expert opinions and added to the specification table. Then, item analysis of the achievement test was made, and item difficulty and item discrimination were calculated and presented in Table 12.

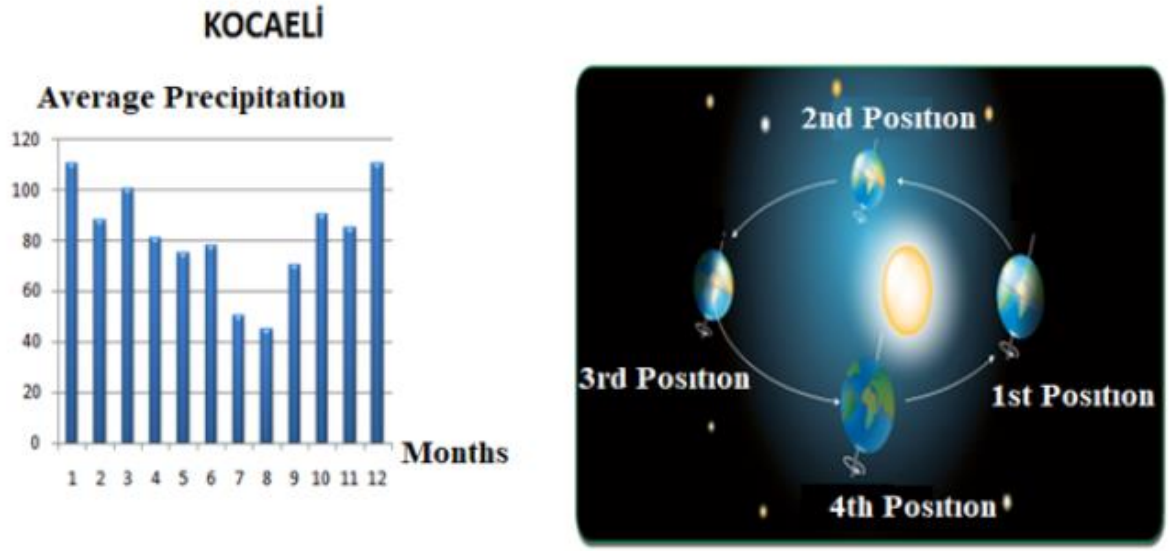
Table 12. Item Difficulty (p) and Discrimination Indexes (r) in the Result of Achievement Test Item Analysis

Item	Upper group	Lower group	Item difficulty index (p)	Evaluation	Item discrimination index (r)	Evaluation
1	55	17	0.6	Easy	0.63	Very good item
2	56	15	0.59	Easy	0.68	Very good item
3	58	26	0.7	Very easy	0.53	Very good item
4	42	13	0.46	Medium difficulty	0.48	Very good item
5	57	12	0.58	Easy	0.75	Very good item
6	46	7	0.44	Medium difficulty	0.65	Very good item
7	43	13	0.47	Medium difficulty	0.5	Very good item
8	58	11	0.58	Easy	0.78	Very good item
9	21	12	0.28	Difficult	0.15	Very weak item
10	43	17	0.5	Easy	0.43	Very good item
11	43	9	0.43	Medium difficulty	0.57	Very good item
12	56	10	0.55	Easy	0.77	Very good item
13	8	20	0.23	Difficult	-0.2	Very weak item
14	39	13	0.43	Medium difficulty	0.43	Very good item
15	57	8	0.54	Easy	0.82	Very good item
16	58	10	0.57	Easy	0.8	Very good item
17	41	11	0.43	Medium difficulty	0.5	Very good item
18	45	9	0.45	Medium difficulty	0.6	Very good item
19	52	8	0.5	Easy	0.73	Very good item
20	60	9	0.58	Easy	0.85	Very good item
21	47	20	0.56	Easy	0.45	Very good item
22	44	13	0.48	Medium difficulty	0.52	Very good item
23	48	12	0.5	Easy	0.6	Very good item
24	54	8	0.52	Easy	0.77	Very good item
25	54	16	0.58	Easy	0.63	Very good item
26	51	18	0.58	Easy	0.55	Very good item
27	57	13	0.58	Easy	0.73	Very good item
28	36	9	0.38	Medium difficulty	0.45	Very good item
29	59	12	0.59	Easy	0.78	Very good item
30	51	10	0.51	Easy	0.68	Very good item
31	38	9	0.39	Medium difficulty	0.48	Very good item
32	56	12	0.57	Easy	0.73	Very good item
33	32	10	0.35	Medium difficulty	0.37	Pretty good item
34	55	13	0.57	Easy	0.7	Very good item
35	15	20	0.29	Difficult	-0.08	Very weak item
36	49	15	0.53	Easy	0.57	Very good item
37	27	10	0.31	Medium difficulty	0.28	It must be corrected.

38	56	12	0.57	Easy	0.73	Very good item
39	55	14	0.58	Easy	0.68	Very good item
40	50	17	0.56	Easy	0.55	Very good item
41	54	13	0.56	Easy	0.68	Very good item
42	59	10	0.58	Easy	0.82	Very good item
43	47	15	0.52	Easy	0.53	Very good item
44	53	16	0.58	Easy	0.62	Very good item
45	60	12	0.6	Easy	0.8	Very good item
46	44	13	0.48	Medium difficulty	0.52	Very good item
47	47	14	0.51	Easy	0.55	Very good item
48	54	12	0.55	Easy	0.7	Very good item
49	35	15	0.42	Medium difficulty	0.33	Pretty good item

Table 12 shows item numbers, the number of correct answers in the upper and lower groups, item difficulty indexes, and item discrimination indexes. According to the difficulty indexes of the items in the test, one is very easy, thirty-one is easy, fourteen is middle difficulty, and three is difficult. In addition, it is seen that forty-three of the items have very good item discrimination, two are pretty good, one needs to be corrected, and three are very weak. It was determined that the difficulties of the 9th, 13th, and 35th items in the test were difficult. These items, which are described as difficult in terms of item difficulties, also have low discrimination. As a result of the item analysis, the 9th, 13th, and 35th items were excluded from the test because their discrimination indexes were below 0.19. Item 37th, whose discrimination index is between 0.20 and 0.29, needs to be corrected. In order to correct this item, the distractors need to be rewritten. However, after these corrections, it is usually necessary to perform item analysis by re-application for the final test. It may not always be possible to make such an application. In such cases, if the relevant items do not affect the distribution of learning outcomes according to Bloom's taxonomy, they can be removed from the test. Item 37th, which needs to be corrected in this study, was excluded from the test because it did not change the distribution of learning outcomes according to Bloom's taxonomy. Some detailed explanations about these items that were excluded from the test are given below. The 9th item, which was asked to determine how accurately the relationship between the average precipitation graph and the model showing the movement of the Earth around the Sun can be established, is given in Figure 1.

The distribution chart of the last 35 years of average precipitation in Kocaeli province by months and the various positions of the Earth while orbiting the Sun are given below.



Which locations on Earth are located in the month with the highest average precipitation in Kocaeli?

A) 1st and 2nd **B)** 2nd and 3rd **C)** 3rd and 4th **D)** 4th and 1st

Figure 1. Item 9 that should be removed from the test

The answers given by the students in the upper and lower groups to item 9 given in Figure 1 are presented in Table 13.

Table 13. The answers given to item 9th

Groups	Options			
	A	B	C	D*
Upper group (60)	37	2	0	21
Lower group (60)	15	17	16	12

**Correct answer*

According to Table 13, it is seen that most of the students in the upper and lower groups of the 9th item were mistaken by choosing option A. Therefore, most of the students could not distinguish between the 4th and 1st positions of the World at the beginning of December, when the most precipitation was seen, according to the graph. The fact that most of the students in the lower and upper groups marked the distractor in option A, and especially the students in the upper group marked this option more, indicates that the 9th item is not distinctive.

Item 13th, which is a question of designing an experiment for the demonstration of wind formation and direction of movement, is presented in Figure 2.

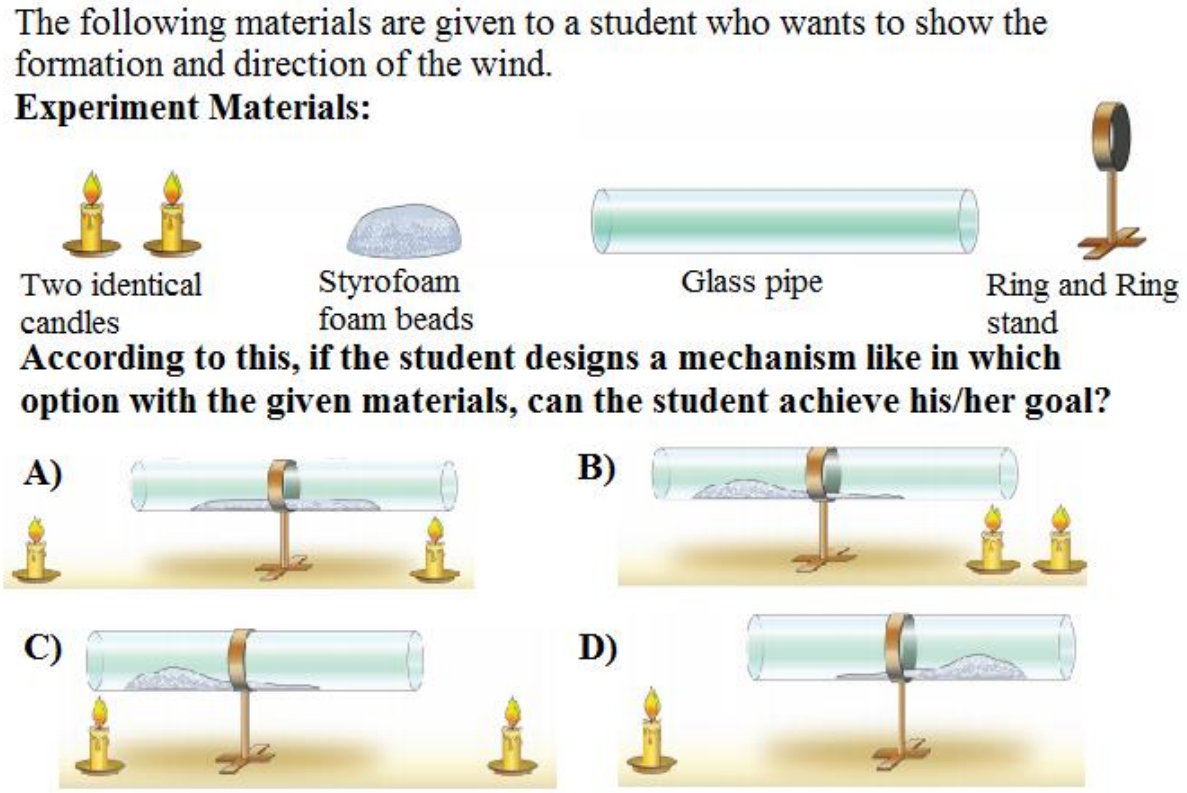


Figure 2. Item 13 should be removed from the test

The answers given by the students in the upper and lower groups to item 13th in Figure 2 are presented in Table 14.

Table 14. The answers given to item 13th

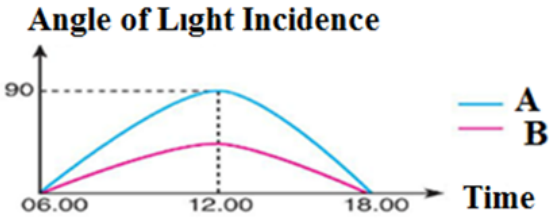
Groups	Options			
	A*	B	C	D
Upper group (60)	8	13	38	1
Lower group (60)	20	15	15	10

**Correct answer*

When Table 14 is examined, it is seen that most of the students in the upper and lower groups made mistakes by marking the B and C options for the 13th item. When the experimental setups in these marked options are examined, the candles for wind formation are correctly positioned to create a temperature difference at both ends of the pipes. However, in the B and C options, the styrofoam foams are not evenly distributed inside the pipe. This is experimentally wrong. Because in order for an experiment to yield clear results, effectiveness must be investigated by manipulating one variable, while other variables must be controlled by keeping them constant. This variable that is manipulated in the experiment is called the independent variable, and the other variables that are kept constant are called the control variable. The independent variable whose effectiveness was investigated and manipulated in the question is temperature. Since the distribution of the styrofoam foam inside the pipe, which will show the wind direction, is a control variable, it should be equally distributed to both ends of the pipe, as in option A. Therefore, according to Table 14, it can be said that most of the students in the lower and upper groups correctly know the information that the wind occurs from cold to hot. However, these students could not distinguish that there should be only one independent variable while conducting an experiment. In addition, the fact that the number of students in the upper group who marked the distractor in option C is higher than the number of students in the lower group, and the number of students in the upper group who marked the correct answer to question A is less than the number of students in the lower group, are the findings supporting that item 33 is not distinctive and should be removed from the test.

The 35th item, which is asked to determine how accurately it is possible to interpret two graphs showing the change of the angle of incidence of sunlight according to hours and compare them with each other, is presented in Figure 3.

35) The time-dependent graph of the angles of sunlight reaching cities A and B on the same day is given below.



Accordingly, regarding cities A and B,

I. On this date, the air temperature in city A is higher than the air temperature in city L.

II. City A is in the Southern Hemisphere, city B is in the Northern Hemisphere.

III. On this date, the day time in city A is longer than the night time.

which of these judgments is definitely true?

A) Only I **B)** I and II **C)** I and III **D)** I, II and III

Figure 3. Item 35 that should be removed from the test

The answers given by the students in the upper and lower groups to item 35 in Figure 3 are presented in Table 15.

Table 15. The answers given to item 35th

Groups	Options			
	A*	B	C	D
Upper group (60)	15	1	42	2
Lower group (60)	20	14	14	12

**Correct answer*

When Table 15 is examined, it is seen that most of the students in the lower and upper groups made a mistake by choosing option C. Therefore, most of the students could not distinguish that it is not possible to make any judgments about the duration of the day and night by only looking at the angles of incidence of the sun's rays in the same day without knowing the location of the cities. In particular, the fact that the number of students in the upper group who marked the distractor in option C is higher than the number of students in the lower group and that the number of students in the upper group who marked the correct answer to question A is less than the number of students in the lower group are the findings showing that item 35 should be excluded from the test as it is not distinctive.

Figure 4 presents the 37th item asked to determine how accurately a model showing the location of the cities is constructed by using the data in the graphs showing the day and night durations on different dates.

37) The night times in city A are shown in Graph-I, and the daytimes in city B are shown in Graph-II at the beginning of the season.

According to this, which of the models showing the position of cities A and B on the face of the Earth to the Sun on December 21 is given correctly?

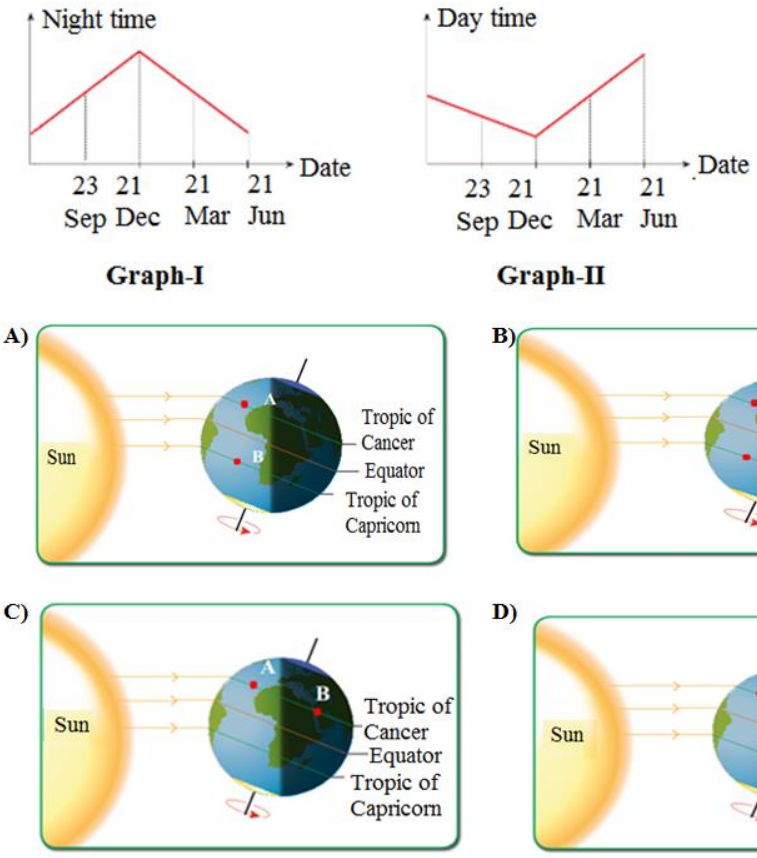


Figure 4. Item 37 that needs to be corrected

The answers given by the students in the upper and lower groups to item 37 in Figure 4 are presented in Table 16.

Table 16. The answers given to item 37

Groups	Options			
	A	B	C	D*
Upper group (60)	13	6	14	27
Lower group (60)	13	18	19	10

**Correct answer*

When Table 16 is examined, it is seen that the 37th item is not distinctive, and most of the students in the upper and lower groups made a mistake by choosing option C. In particular, most of the students in the upper group, which we can define as successful, could not distinguish that it is not possible to make any judgments about the duration of the day and night by only looking at the angles of incidence of the sun's rays on the same day without knowing the location of the cities. In addition, the fact that the number of students in the upper and lower groups who marked the distractor in option A was equal is another remarkable finding. This shows that the distractors in options A and C should be rewritten, and item 37 should be corrected.

As a result of the item analysis, the 9th, 13th, 35th, and 37th items were excluded from the test. However, considering the fact that the exam duration is one class hour and the attention span of the students is taken into consideration, it has been observed that the number of questions is high despite the removal of the specified items from the test. For this reason, in addition to the four items removed from the test, twenty more items that did not affect the distribution of learning outcomes according to Bloom's taxonomy were determined and excluded from the test. As a result, a total of twenty-four items were removed from the test. Items removed from

the test: 1st, 2nd, 3rd, 4th, 7th, 9th, 13th, 14th, 17th, 19th, 20th, 21st, 23rd, 24th, 30th, 32nd, 33rd, 35th, 37th, 38th, 40th, 41st, 47th, and 49th items. There are 25 questions in total in the achievement test. A second expert opinion was needed by the researchers in order to examine the questions in this achievement test, which consists of 25 questions, in order to affect the distribution of the learning outcomes of the items according to Bloom's taxonomy and to examine the questions figuratively and linguistically. In line with the suggestions received from these expert opinions, the 8th grade "Seasons and Climate" achievement test, consisting of a total of 25 questions, was made ready for pilot study by making some formal and verbal corrections in the 11th, 15th, 18th, 26th, 45th, and 46th items. The KR-20 and KR-21 reliability coefficients were calculated as a result of the pilot study made with 75 eighth grade students and presented in Table 17.

Table 17. Statistical results of the seasons and climate unit achievement test

	Initial state of the test	Final version of the test
Seasons and climate achievement test number of questions	49	25
Number of applied persons (N)	223	75
KR-20	0.93	0.88
KR-21	0.93	0.86
Average item difficulty (P_{jx})	0.50	0.61
Average item discrimination (r_{jx})	0.58	0.50

Discussion and Conclusion

This study was carried out to develop a valid and reliable achievement test for the eighth grade "Seasons and Climate" unit within the scope of the science curriculum (MoNE, 2018).

The most important feature of a measurement tool is its validity. There are many types of validity, such as content validity, construct validity, predictive validity, and face validity. However, content validity is a more important type of validity than the others (Demircioğlu, 2007). For this reason, one of the most important conditions for ensuring the validity of a measurement tool is to ensure the content validity of the developed measurement tool (Yiğittir & Çalışkan 2013). Content validity expresses to what extent the questions in the measurement tool can measure the targeted acquirement universe (Büyüköztürk et al., 2015; Fraenkel & Wallen, 2006; cited in Yiğittir & Çalışkan, 2013). In order to ensure the content validity of a measurement tool, the questions must be in accordance with the specification table in terms of the learning outcomes it measures (Baykul, 2000). For this reason, a table of specifications was prepared for the validity of the Seasons and Climate Achievement Test, and scope validity was tried to be ensured. The table of specifications prepared in this study was prepared by taking into account the learning outcomes in the 2018 Science Curriculum (MoNE, 2018) and classified according to Bloom's Taxonomy. In the relevant literature, it is stated that expert opinion is a necessary prerequisite for ensuring content validity (Karslı & Ayas, 2013). For this reason, in order to ensure content validity in the study, the opinions of three science education specialist lecturers, two science teachers, and a Turkish teacher were taken. The scope validity of the "Seasons and Climate" achievement test, which was prepared as a result of these studies, was ensured.

When the studies on the subjects of "Formation of Seasons" and "Climate and Air Movements" in the eighth grade seasons and climate unit in the MoNE (2018) science curriculum are examined, it is seen that there are some limitations in terms of validity and reliability (Birgin & Özcan, 2022; Geren, 2022; Yanardağ, 2021). The achievement test in the study of Birgin and Özcan (2022) was developed to determine the level of knowledge of eighth grade students about "the formation of the seasons." In the study conducted by Birgin and Özcan (2022), questions for validity were classified using the one-dimensional cognitive domain taxonomy developed by Bloom (1956) and updated by Haladyna (1997) in the context of higher-order thinking skills. In addition, in the study of Birgin and Özcan (2022), item analysis and internal consistency coefficients were calculated by making a pilot study for the reliability study. Although a valid and reliable achievement test has been obtained as a result of all the studies, the test developed by Birgin and Özcan (2022) is not suitable for measuring success for all the subjects in the "Seasons and Climate" unit in terms of content validity since it only covers the subject of "The Formation of the Seasons." In the study conducted by Geren (2022), it was aimed to determine the perception, perception skills, cause-effect relationship, cause-effect relationship skills, and exam successes of eighth grade students regarding the concepts in the "Seasons and Climate" unit and to investigate the relationship between these variables. Additionally, determining whether these variables differ according to gender was also investigated. Therefore, many scales (perception, perception skill, cause-and-effect relationship,

etc.) were developed by Geren (2022) for the concepts in the "Seasons and Climate" unit within the scope of the research. One of these is the Seasons and Climate exam. On the other hand, in the study conducted by Yanardağ (2021), it was aimed at investigating the effect of the flipped classroom application on the academic success and learning retention of eighth grade students for the "Seasons and Climate" unit. For this reason, the "Seasons and Climate" academic achievement test was developed and used by Yanardağ (2021). When the academic achievement tests developed by Geren (2022) and Yanardağ (2021) were examined, it was seen that they covered all subjects of the "Seasons and Climate" unit. However, only expert opinion was taken for content validity in both achievement tests developed by Geren (2022) and Yanardağ (2021).In addition, while Geren (2022) and Yanardağ (2021) were developing the achievement tests, the HSEE sample questions in various sources of the MoNE and the questions in the learning outcomes comprehension tests were used without changing them. In order to ensure the reliability of the achievement tests in the study of Geren (2022) and Yanardağ (2021), a pilot study and item analysis could not be performed. For this reason, although the achievement tests developed by Geren (2022) and Yanardağ (2021) were developed for all subjects in the "Seasons and Climate" unit, it can be said that they are limited in terms of validity and reliability.

It is an important feature of a measurement tool to be reliable as well as valid. Şencan (2005) states that scientists refer to reliability as a lower limit value of validity. Therefore, it is possible to say that reliability is a prerequisite for validity. Reliability can be defined as a measurement tool accurately measuring the feature to be measured and giving similar results in different samples selected from the same population at different places and times (Şencan, 2005). Therefore, reliability may change as the sample changes (Capraro & Capraro, 2002; Henson et al., 2001). Therefore, while developing a measurement tool, the reliability of the measurements should be calculated (Capraro et al., 2001; Thompson, 1994). In this study, which was carried out to develop an achievement test for the Seasons and Climate unit, in order to ensure reliability, the pre-pilot study was carried out online for the eighth grade students in secondary school, and the pilot study was carried out in writing in the classroom environment. Item analysis and KR-20 and KR-21 internal consistency coefficients were calculated with the data obtained from both pilot studies.

As a result of the pre-pilot study, both the KR-20 and KR-21 reliability coefficients of the 49-question "Seasons and Climate" achievement test were calculated at 0.93. According to Büyüköztürk (2013), a reliability coefficient of 0.70 and above is sufficient for the reliability of the measurement results. In this respect, it is possible to say that the reliability of the first version of the achievement test is sufficient. In addition, in the first version of the Seasons and Climate Achievement Test, the average item difficulty was calculated as 0.50, and the average discrimination index was calculated as 0.58. In a test, the average item difficulty should be around 0.50 and the average discrimination index should be above 0.30 (Çepni, et al., 2008; Tekin, 2010). In this respect, the first version of the Seasons and Climate Achievement Test is at the desired level in terms of average difficulty and average discrimination index. However, according to the item analyses made through the data obtained from the pre-pilot study, the 9th, 13th, and 35th items should be removed from the test because the discrimination index is below 0.30, and the 37th item should be corrected because the discrimination index is 0.28. However, since there were other items in the test that could measure the learning outcomes measured by these items, all of the 9th, 13th, 35th, and 37th items were excluded from the test. Considering that the duration of the exam is one lesson and the attention span of the students is very high, the number of questions is very high in terms of the applicability of the exam. For this reason, twenty more items that did not affect the distribution of learning outcomes according to Bloom's taxonomy were determined and removed from the test. As a result, a total of twenty-four items were removed from the test. Items removed from the test: 1st, 2nd, 3rd, 4th, 7th, 9th, 13th, 14th, 17th, 19th, 20th, 21st, 23rd, 24th, 30th, 32nd, 33rd, 35th, 37th, 38th, 40th, 41st, 47th, and 49th items. By removing the specified items from the test, an achievement test of 25 questions was obtained. Expert opinion was needed to remove the items specified by the researchers, to affect the distribution of the learning outcomes according to Bloom's taxonomy, and to control the questions figuratively and linguistically. In line with the suggestions received from these expert opinions, the 8th grade "Seasons and Climate" achievement test, consisting of a total of 25 questions, was made ready for pilot study by making some formal and verbal corrections in the 11th, 15th, 18th, 26th, 45th, and 46th items. As a result of the pilot study, the KR-20 reliability coefficient was calculated at 0.88 and the KR-21 reliability coefficient at 0.86. The reliability coefficient values in the final version of the seasons and climate achievement test decreased compared to the first version. This may be due to the decrease in the number of test items. As a matter of fact, the number of items in a measurement tool is a factor affecting its reliability (Şencan, 2005). Although the reliability coefficient values of the test decrease due to the decrease in the number of items in the test, the KR-20 and KR-21 values in the final version of the test are above 0.70. KR-20 and KR-21 values above 0.70 indicate that the measurement results are statistically reliable. Therefore, it is possible to say that the Seasons and Climate Achievement Test with 25 questions developed within the scope of this study is a reliable test. In addition, in the final version of the Seasons and Climate Achievement Test, the average item difficulty was calculated at 0.61

and the average discrimination index was calculated at 0.50. The fact that the average item difficulty is around 0.50 and the average discrimination index is above 0.30 in the developed test (Çepni et al., 2008; Tekin, 2010) shows that the discrimination and difficulty levels of the test are sufficient.

As a result of the analyses made in this study, an achievement test was developed for the whole eighth grade "Seasons and Climate" unit, consisting of 25 items with questions about different levels of Bloom's Taxonomy and statistically sufficient validity and reliability.

Suggestions

In the context of the findings obtained from the studies carried out to ensure the validity and reliability of the "Seasons and Climate" achievement test, some suggestions are presented below.

- Since this achievement test for the Seasons and Climate unit was developed with a survey model, it can be applied to 8th grade students in secondary school in terms of generalizability to the universe. In this respect, it can be used for measurement and evaluation purposes.
- The learning outcomes in the 2018 secondary school science curriculum are metacognitive learning outcomes. Therefore, with the "Seasons and Climate" achievement test developed, it was aimed at measuring more metacognitive achievements. For this reason, the questions and the determined outcomes were classified according to Bloom's Taxonomy, and mostly questions in the application, analysis, synthesis, and evaluation steps were used in the test. In addition, multiple-choice questions, which are easy to apply to large masses but not suitable for measuring metacognitive skills, have been made suitable for measuring high-level skills in this way. In this way, researchers can develop multiple-choice achievement tests for different subjects and achievement tests that measure metacognitive achievements.
- While developing the achievement test for the "Seasons and Climate" unit, the achievements in the MoNE science curriculum were taken as a basis. Since MoNE's science curriculum is updated from time to time according to international developments, it is possible to say that the achievement test developed is aimed at international readers. In this sense, the developed achievement test can be translated into different languages and adapted.
- After the adaptation work is done, the academic achievements of the students in different countries for "Seasons and Climate" can be determined and compared with those of the students in Turkey.

Acknowledgements or Notes

This study was carried out within the scope of the doctoral thesis of Emine YURTYAPAN, who is the corresponding author of the article.

Author (s) Contribution Rate

Emine Yurtyapan: Literature review, Achievement test development, Data collection, Data Analysis, Application, Writing and Editing
Ayşe Gül Çirkinoğlu Şekercioğlu: Methodology, Consulting, Audit, Review and Editing

Conflicts of Interest

There is no conflict of interest.

Ethical Approval (only for necessary papers)

Ethical permission (15.09.2023-E.293638 and 14.09.2023-E.293277) was obtained as a result of the decision of the Scientific and Engineering Sciences Ethics Commission of Balıkesir University Institute of Science and Technology at its meeting dated 12.09.2023 and numbered 2023/5.

References

Akbulut, H. İ., & Çepni, S. (2013). How to develop an achievement test for a unit?: A study for grade 7 force and motion unit. *Amasya Education Journal*, *2*(1), 18-44. Retrieved from https://dergipark.org.tr/en/pub/amauefd/issue/1728/21171

Akşit, F., & Şahin, C. (2011). The effect of active learning on academic achievement and attitudes in geography teaching. *The Western Anatolia Journal of Educational Sciences (WAJES)*, *2*(4), 1-26.Retrieved from https://dergipark.org.tr/en/pub/baebd/issue/3339/46229

Başol, G. (2019). *Assessment and evaluation in education*. (6th ed.). Ankara: Pegem Academy Publishing.

Baykul, Y. (2000). *Measurement in education and psychology: Classical test theory and its application*. Ankara: Student Selection And Placement Center (SSPC) Publications.

Beuchert, A. K., & Mendoza, J. L. (1979). A Monte Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement,* 109-117. Retrieved from https://www.jstor.org/stable/1434454

Birgin, A., & Özcan, H. (2022). Development of an instrument for measuring 8th grade students' knowledge about formation of seasons. *Trakya University Journal of Social Science*, *24*(1), 305-326. Retrieved from https://dergipark.org.tr/tr/pub/trakyasobed/issue/70085/817792

Bolat, M., & Altınbaş, A. (2018). Determination of field information about seasons of science and social studies teacher candidates by parallel mixed. *OPUS International Journal of Society Researches*, *9*(16), 950-983. Retrieved from https://dergipark.org.tr/en/pub/opus/article/435788

Boz, C., Bostan Sarıoğlan, A., & Özcan, H. (2022). Development of an achievement test on chemical bonds. *Manisa Celal Bayar University Journal of the Faculty of Education*, *10*(2), 149-166. Retrieved from https://dergipark.org.tr/en/pub/mcbuefd/issue/74952/1065432#article_cite

Büyüköztürk, Ş. (2013). *Handbook of data analysis for social sciences.* (8th ed.). Ankara: Pegem Academy Publishing.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö.E., Karadeniz, Ş. & Demirel, F. (2015). *Scientific Research Methods*. (19 th ed.). Ankara: Pegem Academy Publishing.

Capraro, R. M., & Capraro, M. M. (2002). Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study. *Educational and Psychological Measurement*, *62*(4), 590-602. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c651761767aaa2a28722d309b95411c119e8f16d

Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement error of scores on the mathematics anxiety rating scale across studies. *Educational and Psychological Measurement,61*(3), 373-386. Retrieved from https://psycnet.apa.org/doi/10.1177/00131640121971266

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education, Inc.

Çepni, S., Bayrakçeken, S., Yılmaz, A., Yücel, C., Semerci, Ç., Köse, E., Sezgin, F., Demircioğlu, G. & Gündoğdu, K. (2008). *Assessment and Evaluation*. Ankara: Pegem Academy Publishing.

Çirkinoğlu Şekercioğlu, A.G. & Yılmaz Akkuş, G. (2019). The effect of drama method to the 7th grade students' success in solar system and beyond: Space puzzle unit. *Abant İzzet Baysal University Journal of Education Faculty*, *19*(1), 125-146.Retrieved from https://doi.org/10.17240/aibuefd.2019.19.43815-538314

Dağ, M., & Karamustafaoğlu, S. (2023). "Properties of matter" unit success test development and determining student success. *Journal of National Education*, *52*(237), 221-254. Retrieved from https://doi.org/10.37669/milliegitim.1076053

Demir, N., Kızılay, E., & Bektaş, O. (2016). Development of an achievement test about solutions for 7th graders: A validity and reliability study. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *10*(1), 209-237. Retrieved from https://dergipark.org.tr/en/download/article-file/224370

Demircioğlu, G. (2007). *Validity and reliability*, In E. Karip (Ed.), *Quantification and consideration*, Ankara: PegemA Publishing.

Doğru, M., & Çepni, S. (2023). A comparative study of traditional multiple choice and context-based achievement test preparation: 7th grade light and matter interaction unit. *Journal of Science Mathematics Entrepreneurship and Technology Education*, *6*(1), 74-101. Retrieved from https://dergipark.org.tr/en/pub/fmgted/issue/75754/1215035#article_cite

Geren, M. (2022). *Evaluation of students' perception, cause-effect relationship, skills and exam success about the "Seasons and climate unit" conceptions.* (Unpublished Master's Thesis), Sakarya University, Sakarya.

Gülen, S. (2019). Analysis of eighth grade students' academic achievements in seasons and climate unit. *Journal of Educational And Instructional Studies in the World*. *9*, 1-11.

Güler, N. (2021). *Assessment and Evaluation in Education.* (15th. ed.). Ankara: Pegem Academy Publishing.

Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Allyn & Bacon, 160 Gould Street, Needham Heights, MA 02194-2310

Hasançebi, B., Terzi, Y., & Küçük, Z. (2020). Distractor analysis based on item difficulty index and item discrimination index. *Journal of Science, Gümüşhane University*, *10*(1), 224-240. Retrieved from https://doi.org/10.17714/gumusfenbil.615465

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and psychological Measurement*, *61*(3), 404-420. Retrieved from https://doi.org/10.1177/00131640121971284

Karamustafaoğlu, S., & Tutar, M. (2016). Developing an achievement test for the unit "Our World, Moon, and Our Life Source Sun" for 6th grade. *Innovations and Quest for Quality in Educational Sciences* (ss. 303-320), Ö. Demirel ve S. Dinçer (Ed.), Pegem Academy Publishing, Ankara. Retrieved from http://dx.doi.org/10.14527/9786053183563b2.020

Karslı, F., & Ayas, A. (2013). A study on the development of a test about measurement of science and technology teachers' science process skills. *Journal of Turkish Science Education*, *10*(2), 66-84. Retrieved from https://www.tused.org/index.php/tused/article/view/291

Kempa, R. (1986). *Assessment in Science.* Cambridge University Press, Cambridge, London.

Kızkapan, O., & Bektaş, O. (2018). Development of an achievement test on seventh grade structure and properties of matter unit: validity and reliability study. *The Journal of International Lingual, Social and Educational Sciences*, *4(*2), 186-202. Retrieved from https://doi.org/10.34137/jilses.431971

Küçükahmet, L. (2002). *Instructional Planning and Assessment*. (13th ed.). Ankara: Nobel Publishing and Distribution.

MoNE (2018). *Science Curriculum for Primary Education Institutions (Elementary and Middle Schools) (Grades 3, 4, 5, 6, 7, and 8)*. Ankara: MoNE Publishing House.

Ogan Bekiroğlu, F. (2004). *How successful are they? Classical and alternative assessment and evaluation methods and applications in physics* (1st ed.). Ankara: Nobel Publishing and Distribution.

Öksüz, Y., & Güven Demir, E. (2019). Comparison of open ended questions and multiple choice tests in terms of psychometric features and student performance. *Hacettepe University Journal of Education*, *34*(1), 259-282. Retrieved from http://efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/2725-published.pdf

Özçelik, D.A. (1998). *Assessment and evaluation*. Ankara: Student Selection And Placement Center (SSPC) Publications.

Özmen, H. & Karamustafaoğlu, O. (2019). *Research methods in education.*(1st ed.). Ankara: Pegem Academy Publications.

Pressley, M., Yokoi, L., Van Meter, P., Van Etten, S., & Freebern, G. (1997). Some of the reasons why preparing for exams is so hard: What can be done to make it easier?. *Educational Psychology Review*, *9*, 1-38. Retrieved from https://doi.org/10.1023/A:1024796622045

Schoon, K. J. (1992). Students' alternative conceptions of earth and space. *Journal of Geological Education*, *40*(3), 209-214. Retrieved from https://doi.org/10.5408/0022-1368-40.3.209

Şencan, H. (2005). *Reliability and validity in social and behavioural measurements*. (1st ed.). Seçkin Publishing.

Sneider, C., Bar, V. & Kavanagh, C. (2011). Learning about seasons: A guide for teachers and curriculum developers. *Astronomy Education Review*, *10*(1). Retrieved from https://www.learntechlib.org/p/109098/.

Sontay, G., & Karamustafaoğlu, O. (2020). Developing achievement test for science course "Sun, earth and moon" unit. *Gazi University Journal of Gazi Faculty of Education*, *40*(2), 511-551. Retrieved from https://dergipark.org.tr/en/pub/gefad/issue/56462/670701

Tekin, H. (2010). *Assessment and evaluation in education.* (20th ed.). Ankara: Yargı Publishing House.

Türk, C. & Kalkan, H. (2015). The effect of planetariums on teaching specific astronomy concepts. *Journal of Science Education and Technology*, *24*(1), 1-15. Retrieved from https://doi.org/10.1007/s10956-014-9516-6

Türk, C. (2015). *Efficiency of the teaching of astronomy through models*. (Unpublished Doctoral Dissertation), Ondokuz Mayıs University, Samsun.

Türk, C., Kalkan, H., Kıroğlu, K. & Ocak İskeleli, N. (2016). Elementary school students' mental models about formation of seasons: A cross sectional study. *Journal of Education and Learning*, *5*(1), 7-30. Retrieved from http://files.eric.ed.gov/fulltext/EJ1085430.pdf

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54,* 837-847.

Uzunöz, A., & Buldan, İ. (2012). Devoloping an achievement test for atmosphere and climate unit of naturel systems on secondary education geography lesson. *Kastamonu Education Journal*, *20*(1), 291-312. Retrieved from https://dergipark.org.tr/en/pub/kefdergi/issue/48696/619526

Varzikioğlu, D. (2023). *The effect of teaching enriched with astronomical activities and educational games on astronomical success and attitude.* (Unpublished Master's Thesis), Gazi University, Ankara.

Yanardağ, H. (2021). *The effects of flipped classroom practices on 8th grade students' academic achievement, attitudes and retention in teaching the seasons and the climate unit.*(Unpublished Master's Thesis), Dicle University, Diyarbakır.

Yancı, M. V. (2019). *The science textbook for middle school and imam hatip middle school 8*. (1st ed.). Ankara: SDR Dikey Publishing.

Yazıcı, F., Gül, Ş., & Sözbilir, M. The development of an achievement test towards "Reproduction, growth and development in plants and animals" subject. *Journal of Kazım Karabekir Education Faculty*,*46*, 11-19. Retrieved from https://doi.org/10.5152/AUJKKEF.2022.837043

Yıldırım, A. & Şimşek, H. (2016). *Qualitative research methods in the social sciences.* (10th ed.). Ankara: Seçkin Publications.

Yiğittir, S. & Çalışkan H. (2013). Content validity analysis of questions on the field of social studies in the level assessment exam (SBS). *National Education*, 197, 145-156. Retrieved from https://dergipark.org.tr/en/pub/milliegitim/issue/36167/406567

Yilmaz, G., & Bulunuz, M. (2019). Evaluating the effect of formative assessment based learning on students'understanding regarding basic astronomical concepts. *European Journal of Education Studies*, *6*(5), 211-235. Retrieved from https://doi.org/10.5281/zenodo.3374852