Id: 1524161

Participatory Educational Research (PER)
Vol.11(6), pp. 189-215, November 2024
Available online at http://www.perjournal.com
ISSN: 2148-6123
http://dx.doi.org/10.17275/per.24.86.11.6

Investigation of the Items Prepared by Teachers Who Received Training on Item Preparation Measuring Higher-Order Thinking Skills

Ahmet BOLAT*
Sungurlu Vocational School, Hitit University, Çorum, Türkiye
ORCID: 0000-0002-3581-2899

Sevilay KARAMUSTAFAOĞLU
Faculty of Education, Amasya University, Amasya, Türkiye
ORCID: 0000-0002-2852- 7061

Article history
Received:
29.07.2024

Received in revised form:
12.09.2024

Accepted:
28.09.2024

Key words:
Item preparation, Higher-order thinking skills, Science education, In-service training

This study was conducted to determine how the in-service training program for science teachers affects teachers' competencies in preparing items measuring higher order learning levels. The study was conducted with the exploratory special case design, which is one of the special case study models. The study group consisted of 8 science teachers who participated in the training program implemented within the scope of the study. The training program is an 8-week question preparation training that measures higher-order thinking skills. The training was carried out theoretically and practically by science education, assessment and evaluation and curriculum experts. During the study, the teachers prepared items to measure higher order thinking skills related to the Earth and Universe learning area in the science curriculum for six months. A 21-item checklist developed by the researcher was used as a data collection tool in the study. The data obtained in the scope of the program were analyzed descriptively. The analyses were reported as frequency and percentage, and the results were visualised with graphs. At the end of the program, it was determined that 43% of the items prepared by the teachers measured higher-order thinking skills. This finding suggests that the training program positively affected the teachers' competence in writing items that measure higher-order thinking skills. At the end of the study, a fair number of suggestions were made, the most important being the widespread implementation of the training program.

Introduction

To adapt to rapidly changing and developing science and technology, individuals need to enhance their thinking skills. The main task of education should be to cultivate individuals with HOTS by teaching ways of accessing knowledge rather than merely imparting knowledge. This approach enables the development of strategies to more effectively address the existing problems of humanity (Chaffee, 1994). In developing higher-order thinking skills (From here on, it will be expressed as HOTS.), it is important to transfer information learned in the classroom to daily life (Brookhart, 2010). To develop transfer skills, complete learning must be achieved. When complete learning is realized, the information transferred to daily life

* Correspondency: ahmetbolat@hitit.edu.tr

can be used to solve individual and social problems. This leads to the development of problem-solving skills, an important component of HOTS. Similarly, individuals who develop problem-solving skills also enhance their creative thinking skills (Yenilmez & Yolcu, 2007). The development of creative thinking enables individuals to create value-added products for their country.

The philosophy of science education has shifted from rote learning to the acquisition of real-life skills. In its report published in 2007, the US National Research Council (NRC) emphasized that science-literate individuals not only understand natural events but also use them to create new evidence (NRC, 2007). The Organisation for Economic Cooperation and Development (OECD) carries out multifaceted studies to develop and measure students' HOTS in higher education (Güneş, 2012). The current curricula in Turkey place more emphasis on the development and measurement of HOTS than previous curricula (MoNE, 2018). Despite this increased emphasis, the results of PISA (Program for International Student Assessment) show that Türkiye is not yet at the desired level in these skills (Yüksel, 2022). Therefore, Turkey needs to make more progress in developing students' HOTS.

HOTS involves collecting the necessary data on a particular subject, evaluating the data, and reaching a judgment based on this evaluation (Ritchhart & Perkins, 2005). The first comprehensive study on this subject was conducted by Costa (2001). Costa's study identified four key HOTS: problem-solving, creative thinking, decision-making, and critical thinking (Costa, 2001). Additionally, Swartz and Parks (1994) included analytical thinking skills as part of HOTS. Overall, HOTS generally include five types: problem-solving, creative thinking, decision-making, critical thinking, and analytical thinking. Teachers, who are among the most important components of education, have responsibilities in developing students' HOTS.

Teachers can develop HOTS through in-class discussions, activities or tasks that solve real-life problems, and items that require HOTS (Miri, David & Uri, 2007; Antonio & Prudente, 2024). However, study shows that the items asked by teachers in both central exams and classroom practices in Turkey mostly measure lower-order thinking skills (Akpınar & Ergin, 2006; Arı & İnci, 2015; Ayvacı & Türkdoğan, 2009; Cansüngü Koray & Yaman, 2002; Dindar & Demir, 2006; İnceçam, Demir & Demir, 2006; Keskin & Aydın, 2011; Koray, Altunçekiç & Yaman, 2005; Mutlu, Uşak & Aydoğdu, 2003; Özcan & Akcan, 2010; Yıldız, 2015, Wilson & Narasuman, 2020). Similarly, international studies indicate that most items prepared by teachers are insufficient in measuring HOTS (Abdullah et al., 2016; Abosalem, 2016; Boyd, 2008; Driana & Ernawati, 2019; Ernst-Slavit & Pratt, 2017; Marso & Pigge, 1988; Mitana, Muwagga & Ssempala, 2018).

It is argued that using items that encourage students to think and item in both central exams and in-class assessment and evaluation practices will improve their HOTS (Karabulut, 2017). However, HOTS are complex both conceptually and practically. Due to their nature, teachers are concerned about teaching and measuring these skills (Lewis & Smith, 1993; Güleryüz & Erdoğan, 2018). This study aims to examine science teachers' competencies in writing items that measure HOTS in detail. The study is expected to contribute to the literature on teachers' competencies in writing such items.

Abdullah et al. (2016) conducted a study with 196 mathematics teachers and found that their knowledge and practice levels of HOTS were inadequate. The training program applied to teachers positively changed their knowledge and practice levels. Villarroel, Bruna, Brown, and Bustos (2021) provided authentic assessment training to five teachers and examined the effect on the structure of the items prepared by the teachers. They found that the items prepared by the teachers were 1) more open-ended with higher cognitive complexity, 2) more closed and required less memorization, 3) used realistic contexts to measure knowledge in an embedded way, and 4) were more consistent with other items in the tests, which improved the curriculum alignment of the items. Yip (2004) studied the competence of 14 biology teachers who participated in a two-year in-service training program to measure HOTS. Yip found that the items asked by the teachers at the end of the program mostly measured HOTS. Yurdakul, Başokçu, and Yazıcılar (2020) organized a professional development program for 100 secondary school mathematics teachers to prepare items measuring HOTS. They determined that there was a satisfactory increase in teachers' competence in writing items measuring HOTS compared to before the program.

As explained above, there are a limited number of studies in both international literature and in-service training programs that examine the effect of in-service training programs on teachers' competencies in preparing items that measure HOTS. In the national literature, there is no model proposed to improve teachers' competence in preparing items that measure HOTS. This study aims to test and finalize a training model on this subject, which is not found in the national literature.

In this study, an answer to the item "At the end of the in-service training program for science teachers, how are the teachers' competencies in preparing items that measure high-level learning levels?" was sought. The sub-problems related to this main problem are presented below:

(1) At the end of the applied training program, how is the competence of science teachers to prepare items that measure higher-order learning levels at the 5th, 6th-,7th and 6thgrade level?
(2) At the end of the implemented training program, how is the competence of science teachers to prepare items that measure higher-order learning levels at all grade levels?

Method

Research Design

The study was conducted by the special case study model. Special case studies involve in-depth examination of the subject being studied (Punch, 2013). Special case studies are of three types: explanatory, exploratory, and descriptive case studies (Yin, 2003). The case study applied in the scope of this study is an exploratory case study. Exploratory case studies are typically applied to determine the effectiveness of an implemented program (Davey, 2009). To determine the effectiveness of the training program implemented within this study, item preparation and development studies were carried out with 8 science teachers participating in the program. Through a checklist developed by the researcher, the items developed by the teachers were evaluated against 21 different criteria. This allowed teachers to examine their strengths and weaknesses in writing items that measure HOTS in depth by the end of the training program.

To ensure internal validity, detailed reports on how the results were reached were provided. For external validity, the conclusions reached at the end of the special case study were also reported in detail. Detailed process reporting in case studies ensures internal validity, while generalizing results to the theory or model instead of the universe ensures external validity (Gillham, 2000). To ensure reliability, the rater reliability of the checklist used as a data collection tool was confirmed. Checklist reliability is ensured through rater reliability (Landis & Koch, 1977). For this study, three raters, including the researcher, coded the checklist consisting of the data obtained. The flow chart of the research is presented in Figure 1.

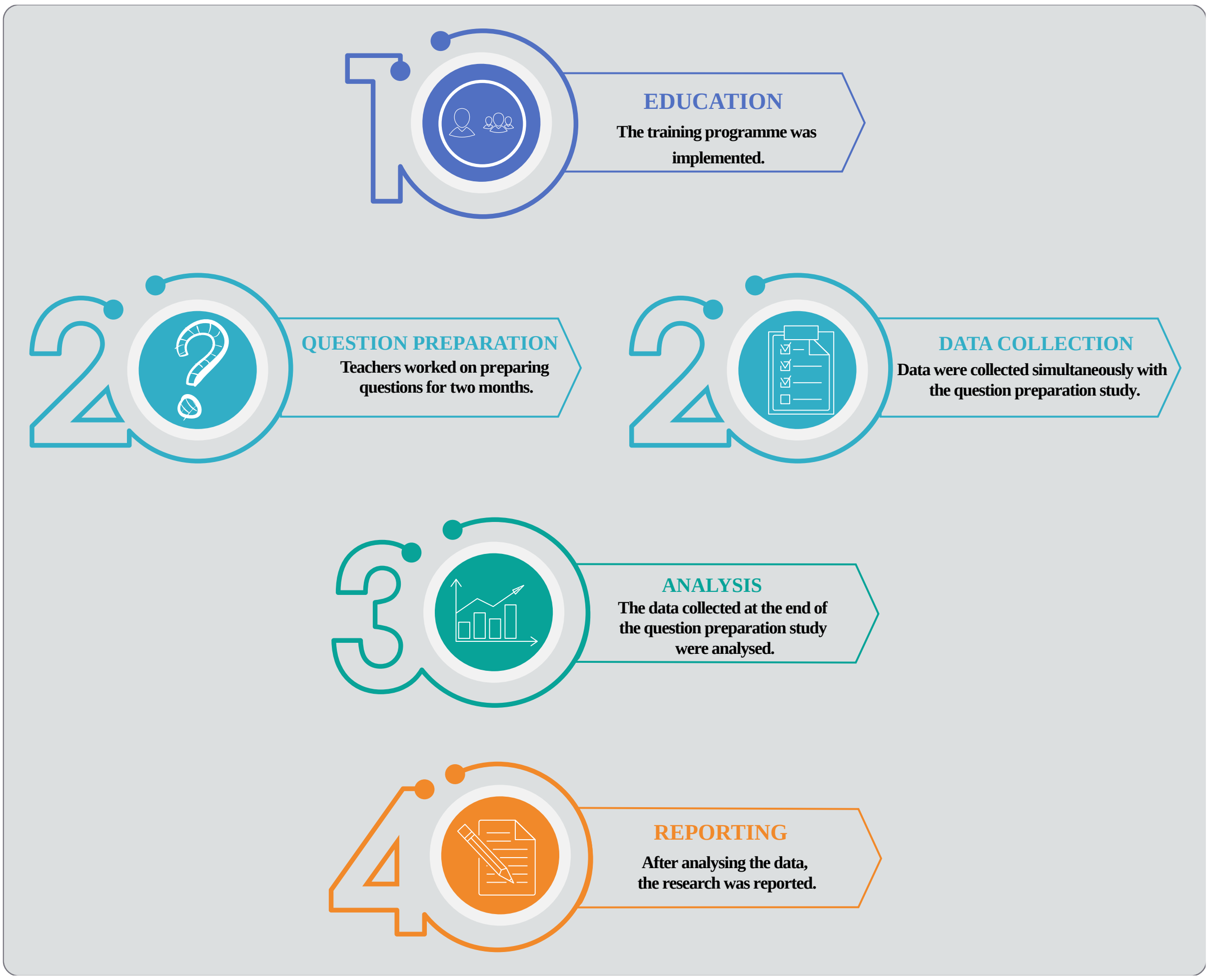


Figure 1. Flowchart of the research process

Sample

This study is the second phase of the first author's doctoral dissertation study. In the first phase, the effect of the training program on 25 teachers' self-efficacy in preparing questions measuring higher order thinking skills was determined. In this phase, which is the second phase, teachers' self-efficacy was divided into four segments according to the score value. In the study, by the maximum diversity technique, two teachers from each of the four different segments, whose self-efficacy scores ranged from low to high based on the scale scores applied in the previous stage of the study, were included in the sample. The maximum diversity technique ensures that the sample consists of individuals with similar, varying, or different features related to the problem (Baltacı, 2018). The study group consisted of a total of 8 teachers selected on the basis of volunteerism.

Education Program

The training program was developed by the researcher by the Taba Program Development Model. The stages of the program creation are presented in Figure 2.

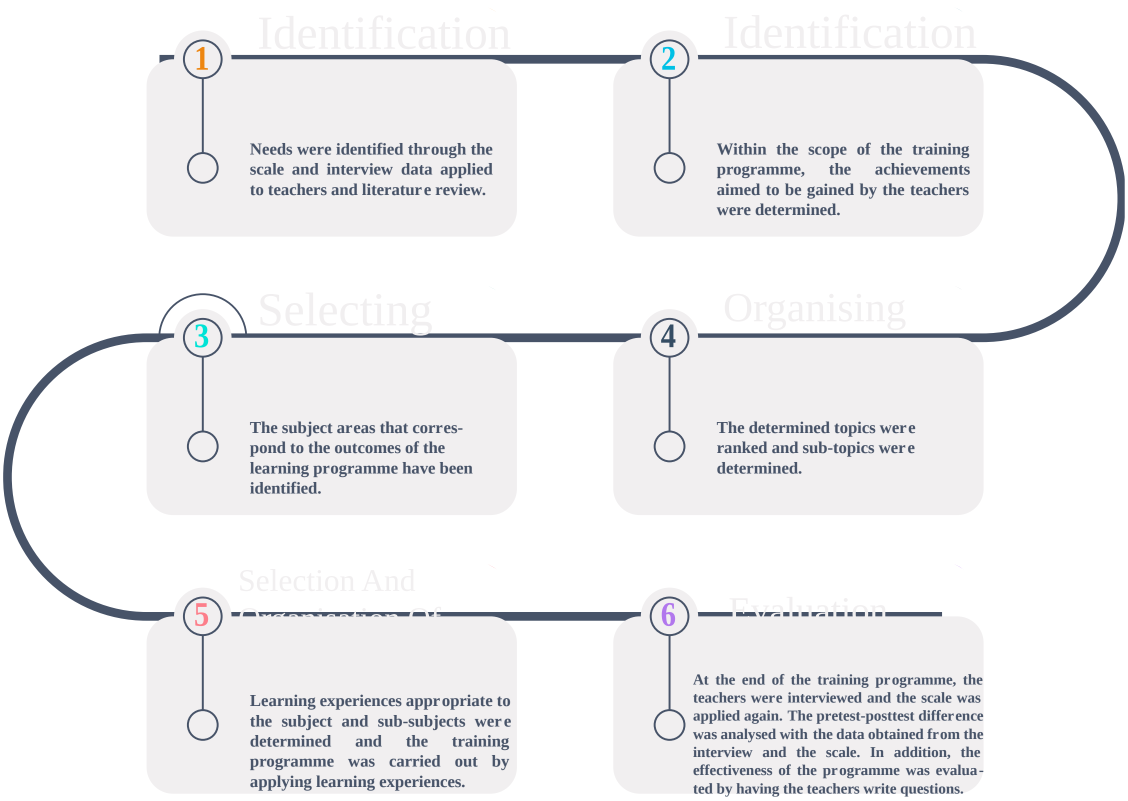


Figure 2. Training program development process for measuring HOTS

In the scope of the study, a needs analysis was conducted, and the objectives of the training program were determined based on the data obtained. In line with the determined objectives, five acquisitions were prepared. These outcomes are as follows:

(1) Explains the basic concepts and principles of measurement and evaluation in education.
(2) Explains high-level thinking skills.
(3) Prepares items appropriate to the outcomes in the Science Curriculum.
(4) Prepares items that measure HOTS.
(5) Prepares a measurement tool for HOTS, including the basic stages of the test planning process.

The main topics were determined to achieve these outcomes. The main topics included "Basic Concepts and Principles of Measurement and Evaluation in Education," "HOTS," "Item Development," " Item Development Measuring HOTS," and "Basic Stages of the Test Planning Process." Subsequently, main headings were listed, and subheadings were created. The main topics and subtopics are presented in Table 1.

Table 1. Main and sub-topics of the training program

Main Topics	Subtopics
HOTS	Critical Thinking. Analytical Thinking. Creative Thinking. Problem Solving. Decision Making. Taxonomies HOTS according to taxonomies
Substance Development	Gain Analysis Skill Analysis Identifying real life skills to be measured. Preparing the specification table. Open-ended items and issues to be considered in writing. Open-ended item writing. Holistic and Analytical Rubrics Analysing and Criticising Sample Open-ended Items. Points to be Considered During the Examination of Open-Ended Items How to Score with Rubric Rules to be Considered by Lower and Upper Scorers What is a multiple-choice item? Multiple choice items and issues to be considered in writing.
Item Development Measuring HOTS	Points to be considered during article writing Relationship between acquisition and skill concepts and skill selection studies Determining the level of items according to Bloom's taxonomy. Analysing, evaluating, writing item by the steps of creation. Multiple-choice item writing studies suitable for acquisition and skill. Writing open-ended items appropriate to the attainment and skill Holistic and Analytical Rubric writing studies Examination of the written items by measurement and evaluation experts, field experts and language experts. Proofreading Selection and Placement of Items in the Trial Form.
Key Stages of the Test Plan Process	Designing a Test Plan Purpose of the Test Contents of the Test Designing a Test Plan

Training on the topics created was implemented through distance education due to the Covid-19 global pandemic. In the first week of the training, training was given on basic concepts in measurement and evaluation, in the second week on higher order thinking skills. In the third week, training was given on the achievements and skills in the science curriculum, and in the fourth week, training was given on preparing open-ended questions that measure higher-order thinking skills. In the fifth week of the training, training was given on multiple-choice question preparation, and in the sixth week, training was given on multiple-choice question preparation measuring higher-order thinking skills. Test development training was given in the seventh week. In the eighth week, with the participation of measurement and evaluation experts, language experts and science education experts, the questions prepared by the teachers were evaluated and feedback was given to the teachers about the questions. The trainings were given by faculty members working in the fields of science education, measurement and evaluation, and instructional programs. A language expert also took part in the trainings. Each field expert provided training in two stages. The first stage was theoretical, and the second stage was practical. During the practical sessions, items were prepared by the theoretical training given in the first stage. Each item was evaluated by the field expert, and feedback was provided to the teachers. The training program lasted two months.

At the end of the training program, the scale was applied again, and interviews were conducted. Additionally, the quality of the items prepared by teachers over six months was evaluated using the checklist developed by the researcher. The evaluation results were presented to the teachers as feedback. In this way, the effectiveness of the program was tested.

Data Collection Tool

Item evaluation checklist measuring HOTS

During the study process, science teachers prepared items to measure HOTS. These items were based on the 5th, 6th, 7th, and 8th grade acquisitions of the Earth and Universe learning area. This part of the study focused on the Earth and Universe learning area because it is taught at all grade levels in schools. The learning domain of Earth and Universe includes 7 objectives at the 5th grade level, 5 objectives at the 6th grade level, 10 objectives at the 7th grade level, and 3 objectives at the 8th grade level, totaling 25 objectives. Each teacher prepared one item per objective, resulting in a total of 25 items per teacher. In total, the 8 participating teachers prepared 200 items over the course of two months.

To facilitate this process, an online class was created in which teachers and the researcher participated. Teachers were tasked with writing items through this online platform. The researcher evaluated the items using a checklist and provided feedback to the teachers during online meetings.

A checklist was developed by the researcher to evaluate the prepared items. Checklists are known as lists of performance criteria (Russel & Airaisan, 2012). Their main purpose is to determine whether the students can perform specified tasks, with scoring done in a binary manner such as present-absent or observed-not observed. The initial version of the checklist contained 27 items and was reviewed by 3 field experts, 1 curriculum development expert, and 2 measurement and evaluation experts. Based on their feedback, the checklist was refined to its final form, which consists of 21 items across four sections. The checklist consists of four sections and 21 items. These are the skills required to measure "the general characteristics, analysis, evaluation and creation steps that questions measuring higher order thinking skills should carry". The general characteristics included in the list are bias, appropriateness of the words to the level of the student, appropriateness to the spelling rules, necessity and compatibility of the visuals for the question, clarity and comprehensibility of the question, scientific accuracy, avoiding unnecessary information, avoiding unnecessary information, the expressions used do not give clues to the correct answer, containing context, the context should include a real life situation and supported by visuals. These features include the standards that should be followed in the preparation of multiple-choice questions put forward by Downing (2005). Standards were set so that the questions in the analyzing step include the skills of comparison, classification, establishing part-whole relationships, and establishing cause and effect relationships. Criteria were established to include the skills of explaining the causes of events, prediction, generalization and decision-making skills of the questions at the evaluation level. Criteria were created to measure the problem solving skills of the questions at the creation level. In the creation of these criteria, the standards set by Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths, and Wittrock (2018) were taken into consideration. The sections of the checklist are presented in Table 2.

Table 2. Sections and items of the checklist

Category	Assessment Criteria
General Features	Bias Conformity of the words to the level of the student Conformity to the spelling rules Necessity and harmony of the visuals for the item Clarity and comprehensibility of the item Scientific accuracy Avoiding unnecessary information Avoiding unnecessary information, the expressions used do not give clues to the correct answer Containing context, the context should include a real-life situation and be supported by visuals.
Analysing	Comparison skill Classification skill Sorting skill Establishing part-whole relationship skill Establishing cause and effect relationship skill
Evaluation	Explaining the causes of events Skill Prediction skill Generalisation skill Decision-making skill
Created	Problem solving skill

In the creation of these criteria, the standards determined by Anderson et all. (2018:39) were taken into consideration. The list was scaled as "Observed" and "Not Observed" according to whether the item exhibited the desired feature.

Analysing the Data

The checklist developed by the researcher, consisting of 21 items, was used to determine whether the items prepared by the participant teachers met the sub-skills requiring HOTS. When coding the checklist, "Observed" responses were coded as "1" and "Not Observed" responses were coded as "0". In the table titles where the results of the analyses are presented, N refers to the total number of items, frequency refers to the number of items that meet the relevant item in the total number of items, and percentage (%) refers to the percentage ratio of the frequency in N.

Before coding, five items about the Earth and Universe learning domain prepared by randomly selected teachers were coded by three experts, including the researcher, and rater reliability was calculated. One of the raters, other than the researcher, is a science education field expert, and the other is a measurement and evaluation field expert. The raters were informed about the checklist beforehand, and instructions were prepared and presented for their use during coding. Fleiss' Kappa Coefficient of Agreement was calculated for 105 items coded by each of the three raters. Fleiss' Kappa Coefficient provides information about scoring reliability by statistically calculating the coefficient of agreement between raters when there are more than two raters (Fleiss, 1971). Fleiss' Kappa Coefficient takes values between -1.00 and +1.00. The coefficient is interpreted as follows: "a coefficient less than 0.00 indicates poor agreement due to chance, 0.00-0.20 indicates insignificant agreement, 0.21-0.40 indicates poor agreement, 0.41-0.60 indicates moderate agreement, 0.61-0.80 indicates substantial agreement, and 0.81-1.00 indicates almost perfect agreement" (Landis and Koch, 1977). The calculation was made with the open-source and free R software. As a result of the calculation, Fleiss' Kappa Coefficient was calculated as 0.52. This value shows that the scoring reliability is generally moderate. After ensuring scoring reliability, the actual scoring process began. During the scoring process, each item prepared by the teachers was coded according to the 21 criteria in the checklist. The scores of "1" and "0" as a result of coding were transferred to the Microsoft Excel program. Descriptive analyses of the items prepared by the teachers were made first according to the grade levels and then according to the total number of items. The analyses were reported as frequency and percentage, and the results of the analyses were visualized with graphs.

Results

Findings related to the first sub-problem

According to the science curriculum, there are seven acquisitions at the 5th grade level. These acquisitions are as follows:

Table 3. Acquisition s at the 5th grade level

Acquisition Code	Acquisition Code
F.5.1.1.1.	Explains the properties of the Sun.
F.5.1.1.2.	Prepares a model comparing the size of the Sun with the size of the Earth.
F.5.1.2.1.	Explains the properties of the Moon.
F.5.1.2.2.	Discusses the idea that living things can live on the Moon.
F.5.1.3.1.	Explains the rotation and orbital movements of the Moon.
F.5.1.3.2.	Explains the relationship between the phases of the Moon and its orbit around the Earth.
F.5.1.4.1.	Prepares a model representing the movements of the Sun, Earth, and Moon relative to each other.

Each teacher in the study group prepared a total of 7 items by the purpose of the study, with one item from the 5th grade acquisitions of the Earth and Universe learning area of the secondary school Science Curriculum. The items they prepared were analyzed using the " Item Evaluation Checklist Measuring HOTS."

The percentages of the level of the items prepared by the study group teachers from the 5th grade acquisitions of the Earth and Universe learning domain are presented in Figure 3.

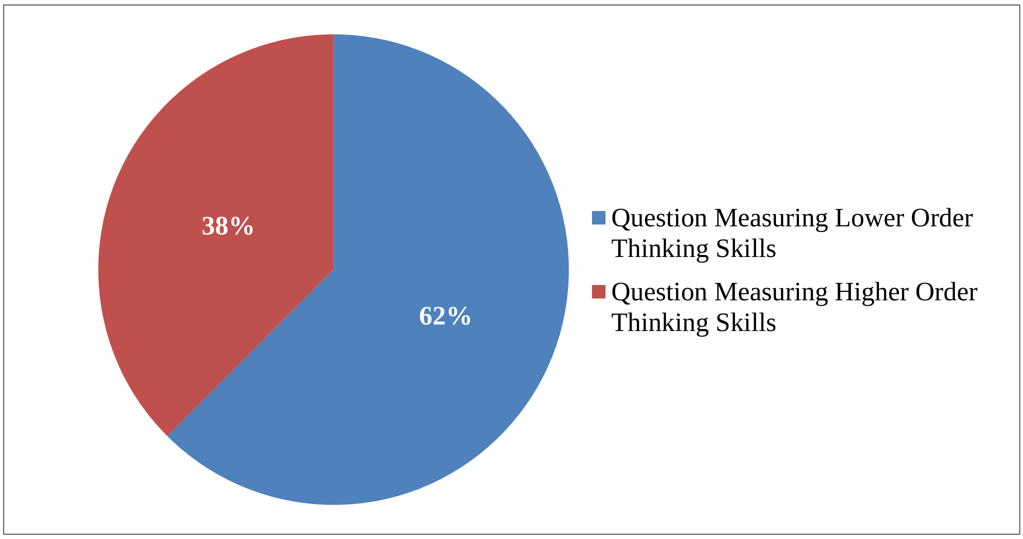


Figure 3. Distribution of items written for 5th grade acquisitions according to cognitive level

When Figure 3. is analyzed, it is understood that 35 (62%) of the total 56 items prepared by the study group teachers from the 5th grade acquisitions of the Earth and Universe learning domain consisted of items measuring lower-level thinking skills, and 21 (38%) of them consisted of items measuring HOTS.

An example of a item measuring HOTS prepared by teachers at the fifth-grade level is presented in Figure 4.

After the teacher covered the topic of "Properties of the Sun" in the lesson, she asked her students to find or make models to show these properties they had learnt.

For the task given by the teacher, Tufan brought oranges to the class as shown in the visual above and explained the similar aspects of the Sun's properties and the model as follows.

I. The shape of the orange is similar to the shape of the Sun.

II. The peel of the orange is similar to the layered structure of the Sun.

III. The dark spots on the orange are similar to the cold regions of the Sun.

According to this, which of Tufan's statements are correct?

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

Figure 4. Example of a 5th grade item that measures HOTS

An example of a item measuring lower-order thinking skills prepared by teachers at the fifth-grade level is presented in Figure 5.

The student, who is doing research on the structure of the Moon, notices deep pits on the surface of the Moon in all the pictures and observations of the Moon. He wonders why and asks his teacher.

Which of the following could be the answer he got from his teacher?

A) Weather events do not occur on the Moon.
B) The temperature difference between day and night is high on the Moon.
C) The Moon reflects the light it receives from the Sun.
D) Because the moon's atmosphere is so thin.

Figure 5. Example of a 5th grade item that measures lower order thinking skills

According to the science curriculum, there are five acquisitions at the 6th grade level. These acquisitions are as follows:

Table 4. Acquisition s at the 6th grade level

Acquisition Code	Acquisition Code
F.6.1.1.1.	Compares the planets in the solar system with each other.
F.6.1.1.2.	Creates a model sorting the planets in the solar system according to their proximity to the Sun.
F.6.1.2.1.	Predicts how a solar eclipse occurs.
F.6.1.2.2.	Predicts how a lunar eclipse occurs.
F.6.1.2.3.	Creates a model representing solar and lunar eclipses.

By the purpose of the study, each teacher in the study group prepared one item from the 6th grade acquisitions of the Earth and Universe learning area of the secondary school Science Curriculum. Since there are 5 objectives at the 6th grade level, each teacher prepared 5 items, totaling 40 items. The items they prepared were analyzed using the " Item Evaluation Checklist Measuring HOTS."

The percentages of the level of the items prepared by the study group teachers from the 6th grade acquisitions of the Earth and Universe learning domain are presented in Figure 6.

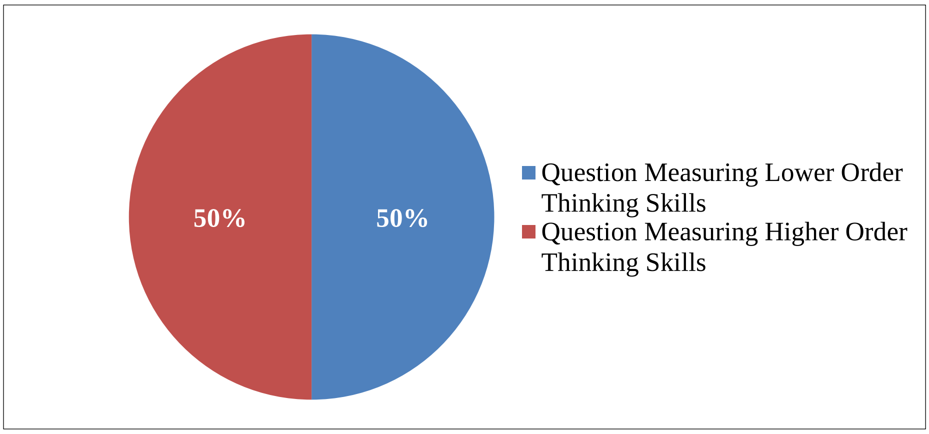


Figure 6. Distribution of items written for 6th grade acquisitions according to cognitive level

An example of a item measuring HOTS, prepared by teachers at the sixth-grade level, is presented in Figure 7.

Ahmet glued the pictures of the Sun at the centre of the wall clock, the Earth at the end of the hour hand and the Moon at the end of the minute hand without affecting the operation of the clock.

About the operation of this clock;

I. A lunar eclipse occurs 24 times a day
II. 2 solar eclipses occur in one day
III. Sun, Earth and Moon cannot be in the same direction except during eclipses

Which of the statements are true?

A) Only I. C) I. and II.

B) I. and III. D) I., II. and III.

Figure 7. Example of a 6th grade item that measures HOTS

An example of a item measuring lower-order thinking skills, prepared by teachers at the sixth-grade level, is presented in Figure 8.

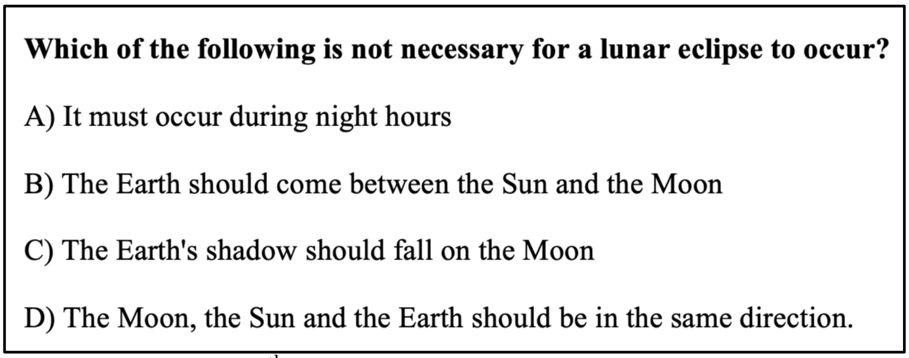


Figure 8. Example of a 6th grade item that measures lower order thinking skills

According to the science curriculum, there are ten acquisitions at the 7th grade level. These acquisitions are as follows:

Table 5. Acquisitions at the 7th grade level

Acquisition Code	Acquisition Code
F.7.1.1.1.	Explains space technologies.
F.7.1.1.2.	Expresses the causes of space pollution and predicts the possible consequences of this pollution.
F.7.1.1.3.	Explains the relationship between technology and space exploration.
F.7.1.1.4.	Explains the structure of the telescope and what it does.
F.7.1.1.5.	Makes inferences about the importance of telescope in the development of astronomy.
F.7.1.1.6.	Prepares and presents a simple telescope model.
F.7.1.2.1.	Recognises the star formation process.
F.7.1.2.2.	Explains the concept of star.
F.7.1.2.3.	Explains the structure of galaxies.
F.7.1.2.4.	Explains the concept of universe.

Each teacher in the study group prepared one item from the 7th grade acquisitions of the Earth and Universe learning domain of the secondary school Science Curriculum, by the purpose of the study. Since there are 10 acquisitions at the 7th grade level, each teacher prepared 10 items, resulting in a total of 80 prepared items. The items they prepared were analyzed using the " Item Evaluation Checklist Measuring HOTS."

The percentage ratios related to the level of the items prepared by the study group teachers from the 7th grade acquisitions of the Earth and Universe learning domain are presented in Figure 9.

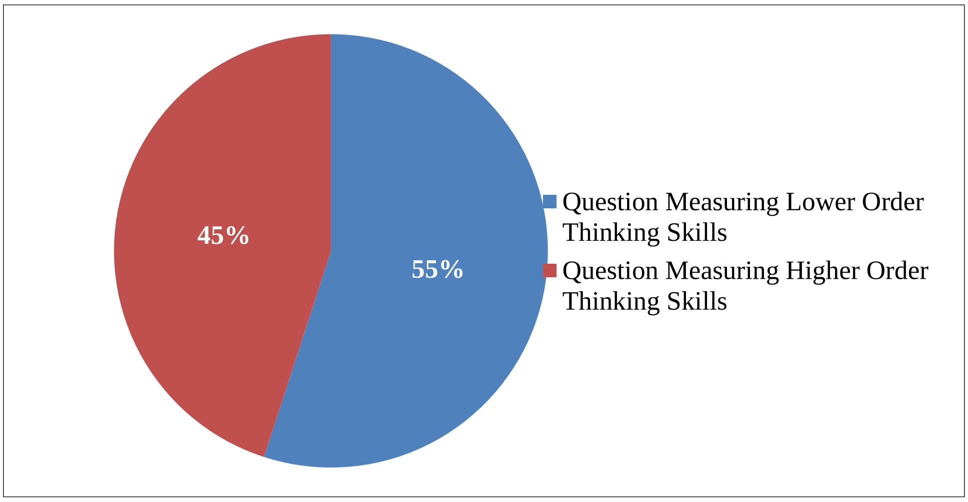


Figure 9. Distribution of the items written for the 7th grade acquisitions according to cognitive level

When Figure 9. is analyzed, 44 (55%) of the total 80 items prepared by the study group teachers from the 7th grade acquisitions of the Earth and Universe learning domain consisted of items measuring lower-level thinking skills, and 36 (45%) of them consisted of items measuring HOTS.

An example of an item measuring HOTS prepared by teachers at the seventh-grade level is presented in Figure 10.

Researchers have warned NASA that human-caused space debris in Earth orbit is on the verge of exceeding reasonable limits.

A report by the National Research Council also warns that space junk could cause fatal leaks on shuttles or damage important satellites.

The organisation also calls for international arrangements to limit the amount of junk in space and to use giant magnetic networks.

The junk in Earth's orbit consists of the additional engines of the space shuttles, used and obsolete satellites.

According to the organisation, collisions between orbiting clumps could damage many satellites and increase the amount of garbage.

According to this, which of the following is not one of the situations that may be caused by space pollution?

A) Increased risk for astronauts on the Space Station

B) Climate and earth science research may be interrupted

C) A decrease in the amount of light in orbit may increase global warming

D) Important space applications such as weather forecasting, communications and the Internet may be lost

Figure 10. Example of a 7th grade item that measures HOTS

An example of a item measuring lower-order thinking skills, prepared by teachers at the seventh-grade level, is presented in Figure 11.

When we want to get close-up pictures and information about the planets, moons, comets and asteroids in our Solar System, we send space probes. Some space probes fly close to many planets and moons, studying them. Others orbit a planet or satellite for a closer look. It transmits the data it collects to Earth via radio signals from space. Probes must withstand harsh conditions to collect data. The instruments carried by space probes may include radiometers, magnetometers, television cameras sensitive to infrared, visible and ultraviolet light, micrometres, special detectors for cosmic rays, gamma rays and solar wind.

Which of the following cannot be deduced from the above information alone?

A) Space probes are equipped with advanced technological tools

B) Communication is established between the space probe and the Earth by radio signals

C) Thanks to space probes, we can learn about the depths of the universe

D) Space probes allow us to learn about comets, not only about the travelling ones.

Figure 11. Example of a 7th grade item that measures lower order thinking skills

In the question above, it is clearly stated which spacecraft space probes are used to study. Therefore, reaching the correct answer requires thinking skills in the dimensions of analysis, evaluation and creating. It is enough for students to read and understand the presented paragraph carefully.

According to the science curriculum, there are ten acquisitions at the 8th grade level. These acquisitions are as follows:

Table 6. Acquisition s at the 8th grade level

Acquisition Code	Acquisition Code
F.8.1.1.1.	Makes predictions about the formation of seasons.
F.8.1.2.1.	Explains the difference between climate and weather events.
F.8.1.2.2.	Says that climate science (climatology) is a branch of science and experts working in this field are called climatologists.

By the purpose of the study, each teacher in the study group prepared one item from the 8th grade acquisitions of the World and Universe learning area of the secondary school Science Curriculum. Since there are three objectives at the 8th grade level, each teacher prepared 3 items, totaling 24 items. The items they prepared were analyzed using the " Item Evaluation Checklist Measuring HOTS."

The percentages of the level of the items prepared by the study group teachers from the 8th grade acquisitions of the World and Universe learning domain are presented in Figure 12.

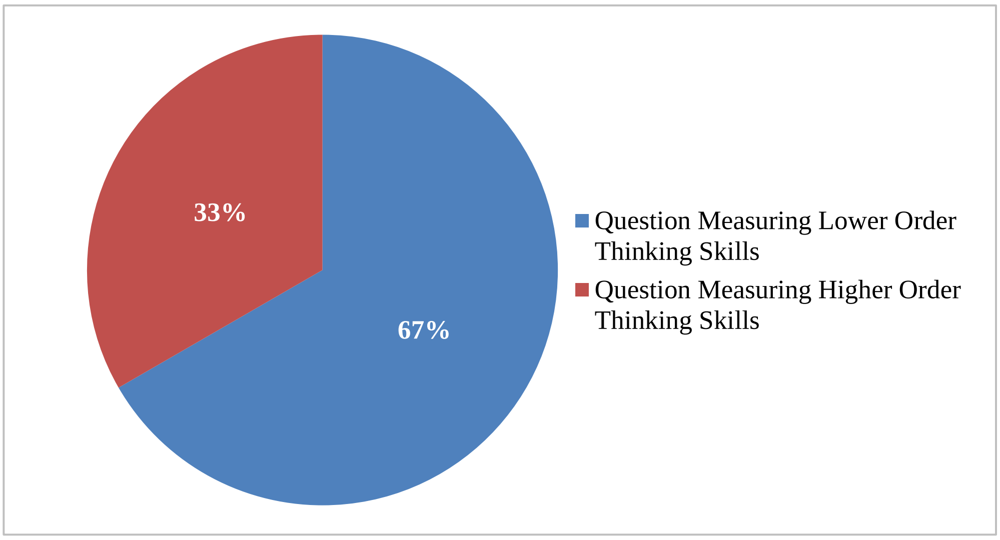
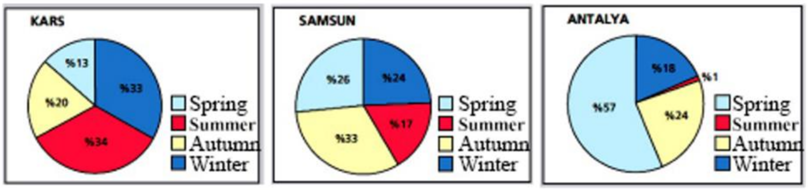


Figure 12. Distribution of the items written for the 8th grade acquisitions according to cognitive level

When Figure 12. is analyzed, 16 (67%) of the total 24 items prepared by the study group teachers from the 8th grade acquisitions of the Earth and Universe learning domain consisted of items measuring lower-level thinking skills, while 8 (33%) of them consisted of items measuring HOTS.

An example of a item measuring HOTS, prepared by teachers at the eighth-grade level, is presented in Figure 13.

In the graphs below, the distribution of precipitation in three provinces according to seasons is given.



According to this.

I. Geographical location is effective on precipitation regime.
II. The graphs were prepared by climatologists.
III. It was created by averaging the weather events in the provinces.

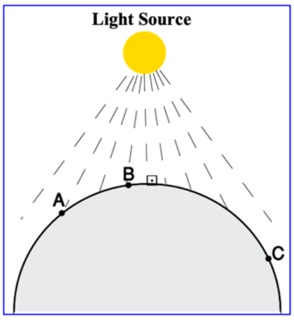
Which of the following statements can be made?

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

Figure 13. Example of an 8th grade item that measures HOTS

An example of a item measuring lower-order thinking skills, prepared by teachers at the eighth-grade level, is presented in Figure 14.

A student prepares the following experimental setup.



After waiting for a while, he measures the temperature values at points A, B and C.

According to this, what is the relationship between the measured temperatures?

A) A=B=C B) A>B>C C) B>C>A D) B>A>C

Figure 14. Example of an 8th grade item that measures lower order thinking skills

Findings related to the second sub-problem

By the purpose of the study, the teachers who formed the study group prepared a item from each acquisition of the World and Universe learning area of the secondary school Science Curriculum. The items they prepared were analyzed using the " Item Evaluation Checklist Measuring HOTS." The results of the analysis are presented in Table 7.

Table 7. Analysis results of the items written for all objectives

Item No	Item	N	Frequency	Percentage (%)
i1	The item is unbiased (gender, cultural, geographical, ethnic, etc.).	200	198	99,00
i2	The wording is appropriate for the student's level.	200	200	100,00
i3	The item follows Turkish spelling rules.	200	200	100,00
i4	Any accompanying image is necessary and relevant to the item.	200	121	60,50
i5	The item is clear, understandable, and concise.	200	194	97,00
i6	The information presented in the item is scientifically accurate.	200	160	80,00
i7	Unnecessary details are avoided in the item.	200	189	94,50
i8	The item does not provide hints towards the correct answer.	200	164	82,00
i9	The item includes context.	200	58	29,00
i10	The context provided in the item reflects a real-life skill or problem.	200	51	25,50
i11	The item's context is supported by a table, graph, figure, or picture.	200	41	20,50
i12	The item assesses the student's ability to compare.	200	35	17,50
i13	The item assesses the student's ability to classify.	200	5	2,50
i14	The item assesses the student's ability to establish a part-whole relationship.	200	4	2,00
i15	The item assesses the student's ability to sort.	200	8	4,00
i16	The item assesses the student's ability to establish cause-and-effect relationships.	200	44	22,00
i17	The item assesses the student's ability to explain the causes of events.	200	22	11,00
i18	The item assesses the student's estimation skills.	200	7	3,50
i19	The item assesses the student's ability to generalize.	200	6	3,00
i20	The item assesses the student's decision-making skills.	200	5	2,50
i21	The item assesses the student's ability to solve a real-life problem with an original approach.	200	2	1,00

N: Total Number of Items; f: Frequency; %: Percentage

When Table 7. is analyzed, in terms of the general characteristics that the items should have, all of the items prepared by the teachers meet the required criteria in the items m2 (conformity to student level) and m3 (conformity to spelling rules) among the evaluation criteria. The criterion met at the highest rate among the criteria of HOTS is m16, which is the criterion of cause-effect relationship, prepared at a rate of 25%. Then, m12, the criterion of comparison skill, was prepared at a rate of 17.5%. Following that, m17, the criterion of explaining the causes of events, was prepared at a rate of 11%. Other criteria prepared include m15 (sorting skill) at 4%, m18 (prediction skill) at 3.5%, m19 (generalization skill) at 3%, m13 (classification skill) at 2.5%, m20 (decision-making skill) at 2.5%, and m14 (establishing part-whole relationship) at 2%. The criterion prepared at the lowest rate (1%) among the items was m21, which is the criterion of problem-solving skill. The percentages related to the level of the items prepared by the study group teachers from all acquisitions of the World and Universe learning domain are presented in Figure 15.

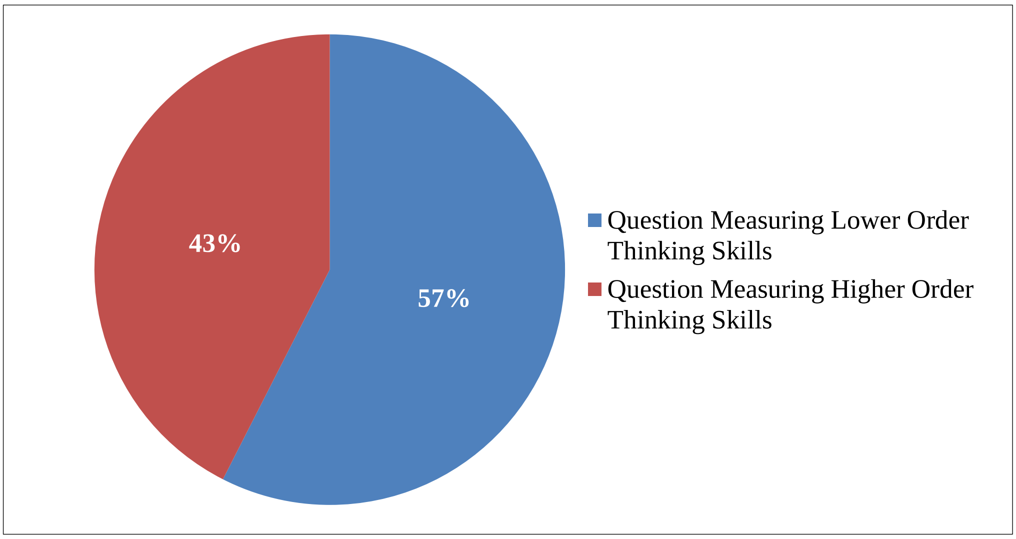


Figure 15. Distribution of items written for all objectives according to cognitive level

When Figure 15. is analyzed, 115 (57%) of the total 200 items prepared by the study group teachers from all acquisitions of the Earth and Universe learning domain were items measuring lower-level thinking skills, while 85 (43%) were items measuring HOTS. The distribution of the prepared items measuring HOTS according to their cognitive level is presented in Figure 16.

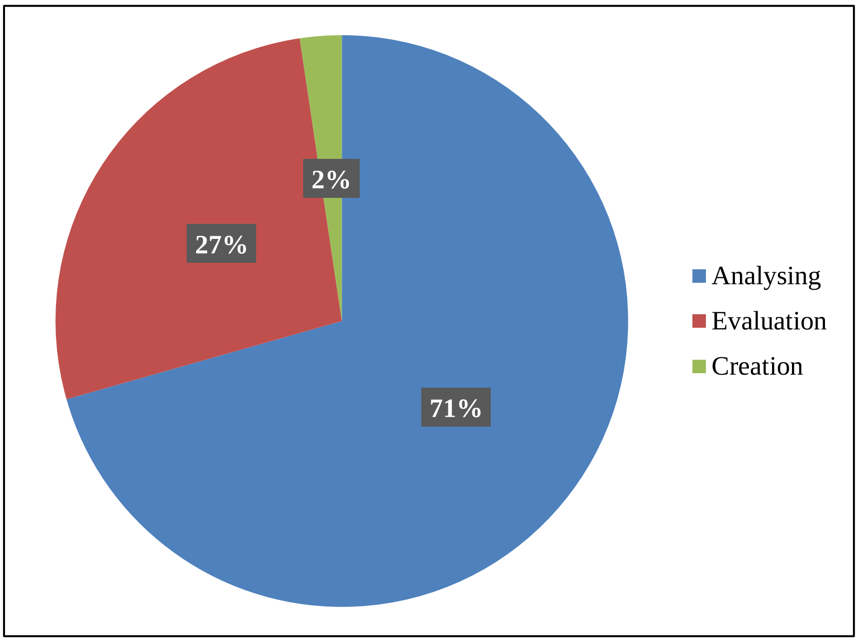


Figure 16. Distribution of items measuring HOTS according to cognitive level

When Figure 16. is analyzed, 60 (71%) of the items measuring HOTS are at the analyzing level, 23 (27%) are at the evaluating level, and 2 (2%) are at the creating level. Figure 17 presents the number of items measuring HOTS according to grade levels.

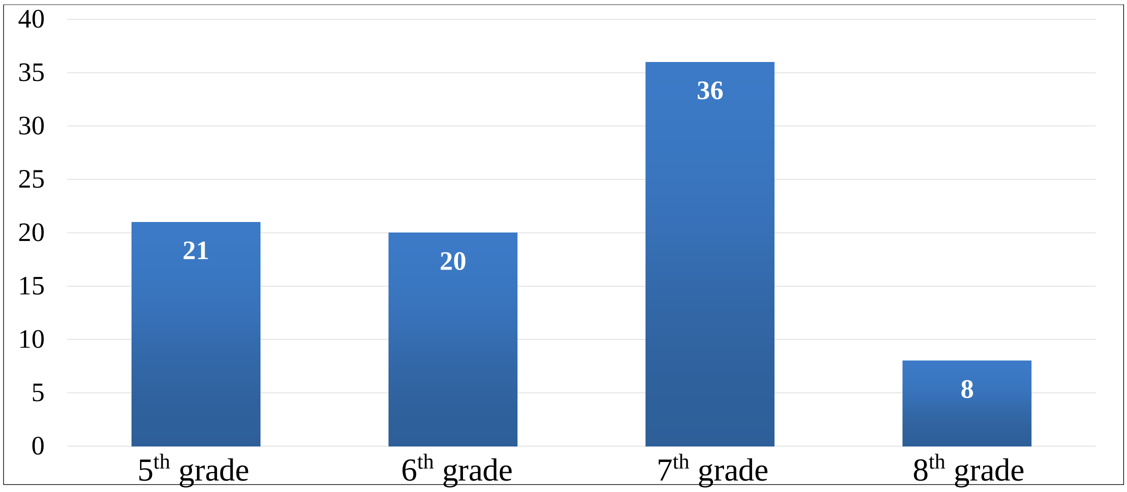


Figure 17. Distribution of the number of items measuring HOTS according to grade level

When Figure 17. is analyzed, the study group teachers prepared 21 items at the 5th grade level, 28 items at the 6th grade level, 29 items at the 7th grade level, and 8 items at the 8th grade level to measure HOTS. However, since the total number of items prepared differed according to the grade levels, it was also evaluated in terms of ratios. The ratio of the prepared items measuring HOTS to the total items is presented in Figure 18.

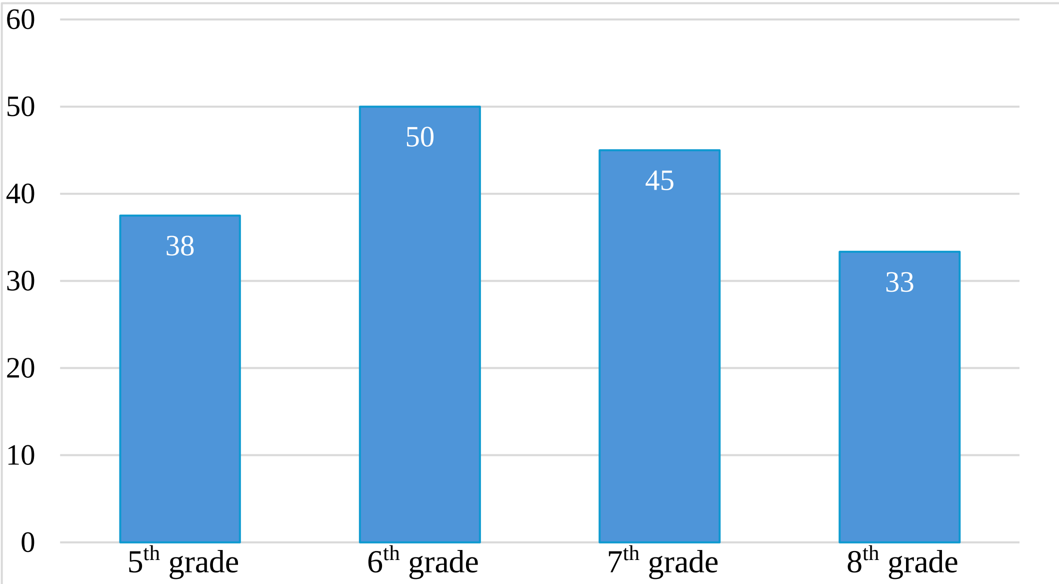


Figure 18. Percentage distribution of the items measuring HOTS according to the items prepared at class level

When analyzing Figure 18., it is evident that items measuring HOTS were prepared at the highest rate for the 6th grade level (50%), followed by the 7th grade level (45%), the 5th grade level (38%), and the 8th grade level (33%).

The type of the item prepared by each teacher according to the acquisitions is presented in Table 8. Items measuring HOTS were coded as 1, while items measuring lower-order thinking skills were coded as 0. The table also includes the frequency and percentage values of items measuring HOTS for each outcome. Additionally, it provides the frequency and percentage of items measuring HOTS prepared by each teacher.

Table 8. Frequency and percentage value of the type of item prepared by each teacher according to the objectives

Grade	Learning Outcome Code		T1	T2	T3	T4	T5	T6	T7	T8	Lower Order		Higher Order	
											f	%	f	%
5	F.5.1.1.1.		0	1	1	0	0	0	0	1	5	62,5	3	37,5
	F.5.1.1.2.		0	0	0	1	0	0	1	0	6	75	2	25
	F.5.1.2.1.		0	1	1	1	1	0	1	0	3	37,5	5	62,5
	F.5.1.2.2.		0	1	1	1	1	0	1	0	3	37,5	5	62,5
	F.5.1.3.1.		0	0	0	1	0	0	1	0	6	75	2	25
	F.5.1.3.2.		0	1	1	1	0	0	0	0	5	62,5	3	37,5
	F.5.1.4.1.		0	1	0	0	0	0	0	0	7	87,5	1	12,5
6	F.6.1.1.1.		0	1	1	0	1	0	0	0	5	62,5	3	37,5
	F.6.1.1.2.		0	1	0	0	1	0	1	0	5	62,5	3	37,5
	F.6.1.2.1.		0	1	1	1	1	0	1	0	3	37,5	5	62,5
	F.6.1.2.2.		0	1	1	1	1	0	1	0	3	37,5	5	62,5
	F.6.1.2.3.		0	1	1	1	0	0	0	1	4	50	4	50
7	F.7.1.1.1.		0	1	1	1	0	1	0	0	4	50	4	50
	F.7.1.1.2.		0	1	1	1	0	0	0	1	4	50	4	50
	F.7.1.1.3.		0	0	1	1	0	0	1	1	4	50	4	50
	F.7.1.1.4.		0	1	0	1	0	0	1	0	5	62,5	3	37,5
	F.7.1.1.5.		0	0	0	1	0	1	1	0	5	62,5	3	37,5
	F.7.1.1.6.		1	1	1	1	0	0	1	0	3	37,5	5	62,5
	F.7.1.2.1.		0	1	1	1	0	1	1	0	3	37,5	5	62,5
	F.7.1.2.2.		0	1	0	1	0	0	1	0	5	62,5	3	37,5
	F.7.1.2.3.		0	1	0	0	0	1	0	0	6	75	2	25
	F.7.1.2.4.		0	0	1	1	0	0	1	0	5	62,5	3	37,5
8	F.8.1.1.1.		0	1	1	0	1	0	1	0	4	50	4	50
	F.8.1.2.1.		0	0	0	0	1	0	0	1	6	75	2	25
	F.8.1.2.2.		0	0	0	0	0	0	1	1	6	75	2	25
Total	Lower Order	f	24	7	10	8	17	21	9	19	115	57,5	85	42,5
		%	96	28	40	32	68	84	36	76				
	Higher Order	f	1	18	15	17	8	4	16	6				
		%	4	72	60	68	32	16	64	24				

T: Teacher, f: Frequency, %: Percentage

When Table 8. is analyzed, it is understood that items measuring HOTS were prepared at the 5th grade level at the highest rate from the objectives "Explains the properties of the Moon" and "Discusses the ideas he/she has produced that living things can live on the Moon". Conversely, the acquisition "Prepares a model representing the movements of the Sun, Earth, and Moon relative to each other" had the lowest rate of items measuring HOTS. At the 6th grade level, items measuring HOTS were prepared most frequently for the objectives "Predicts how a solar eclipse occurs" and "Predicts how a lunar eclipse occurs," while the objectives "Compares the planets in the solar system with each other" and "Creates a model by ordering the planets in the solar system according to their proximity to the Sun" had the lowest rates. For the 7th grade level, items measuring HOTS were most commonly prepared for the objectives "Prepares and presents a simple telescope model" and "Recognizes the star formation process," whereas the objective "Explains the structure of galaxies" had the lowest rate. At the 8th grade level, items measuring HOTS were prepared at the highest rate for the acquisition "Makes predictions about the formation of seasons," while the acquisitions "Explains the difference between climate and weather events" and "Describes climate science (climatology) as a branch of science and identifies experts in this field as climate scientists (climatologists)" had the lowest rates.

Discussion

As part of the study's purpose, a two-month item writing study was conducted with science teachers to assess their competence in creating items that measure HOTS. The items prepared by the teachers were evaluated using a checklist designed for this purpose. According to the evaluation results, out of the 56 items prepared by teachers at the 5th grade level, 21 were identified as measuring HOTS, representing 38% of the total. Comparatively, existing literature reveals varying percentages of higher-order thinking items among those prepared by teachers. For instance, in a study examining items from science courses at the 5th grade level, only 1.15% of 1505 items were categorized as measuring HOTS (Dindar & Demir, 2006). Another study focusing on items prepared by science teachers found that 32.1% of 1592 items assessed were at the higher-order thinking level (Ayvacı & Türkdoğan, 2009). Similarly, a study analyzing items from 327 teachers, including secondary school educators, highlighted that a significant majority were knowledge-based, with few categorized as higher-order thinking items (Marso & Pigge, 1988). Moreover, a study involving over 1400 items by mathematics and science teachers reported that 8% of the items were at the higher-order thinking level (Oescher & Kirby, 1990). In the context of national and international literature, the findings of this study align with existing studies. However, the study suggests that the rate of higher-order thinking items prepared by teachers in this study exceeds those reported in previous literature. This suggests that the training programs and associated interventions implemented as part of this study may positively influence teachers' competence in crafting items that measure HOTS.

Out of a total of 40 items prepared by teachers based on 6th grade acquisitions, 20 (50%) were categorized as measuring lower-level thinking skills, and 20 (50%) were categorized as measuring HOTS. Comparatively, existing literature provides insights into the prevalence of higher-order thinking items among those prepared by teachers in various contexts: In a study analyzing exam items from 13 science teachers at the sixth-grade level using Revised Bloom's Taxonomy (RBT), only 6 (1.2%) out of 543 items were identified as measuring HOTS (Ataş & Güneş, 2020). Another study focusing on pre-service science teachers found that 102 (29.8%) out of 342 items examined measured HOTS (Özcan & Akcan, 2010). A study analyzing exam items by science and technology teachers revealed that 44 (9.1%) out of 1061 items assessed were categorized as measuring HOTS (Tanık & Saraçoğlu, 2011). Boyd (2008) examined items by mathematics teachers at the 8th grade level and reported that 13% of the items assessed measured HOTS. Abdullah et al. (2016), in a study involving 196 mathematics teachers, found inadequacies in preparing items that measure HOTS, particularly in Abu Dhabi. Mitana et al. (2018), studying items for the Primary Completion Examination in Uganda using RBT, found that only 13.2% of the items assessed measured HOTS. When compared with the literature, the findings of this study suggest that teachers participating in this study demonstrated a higher competence in preparing items that measure HOTS.

As a result of analysing the 7th grade level objectives, it was found that 44 (55%) of the total 80 items prepared by the teachers consisted of items measuring lower-level thinking skills and 36 (45%) of them consisted of items measuring higher level thinking skills. When the literature is examined, in the study in which the written exam items prepared by science and technology teachers were analyzed according to the RBT, it was determined that only 37 (6.9%) of the 533 items prepared by the teachers were items measuring HOTS (Gökulu, 2015). In the study in which the written exam items prepared by social studies teachers were analyzed according to the RBT, 60 (11.91%) of the 533 items prepared by the teachers were determined to be items measuring HOTS (Şanlı & Pınar, 2017). In another study examining the items prepared by primary school teachers, it was determined that the items were mostly at the level of recall and comprehension (Driana & Ernawati, 2019). According to the results of a study of 10 mathematics teachers' competencies in assessing HOTS, 9 out of 10 teachers found that they had difficulty in preparing and applying the tool to measure HOTS (Afifah &

Retnawati, 2019). When compared with the national and international literature, it is seen that the teachers in this study have higher rates of preparing items that measure HOTS. This may indicate that the implemented training program contributed positively to teachers' competencies in preparing items that measure HOTS.

Upon analyzing the objectives at the 7th grade level, it was found that out of the total 80 items prepared by teachers, 44 (55%) were categorized as measuring lower-level thinking skills, while 36 (45%) were categorized as measuring HOTS. Comparatively, when examining existing literature: A study analyzing exam items by science and technology teachers using RBT found that only 37 (6.9%) out of 533 items assessed measured HOTS (Gökulu, 2015). In a study analyzing exam items by social studies teachers using RBT, 60 (11.91%) out of 533 items were identified as measuring HOTS (Şanlı & Pınar, 2017). Another study focusing on primary school teachers revealed that the majority of items were at the recall and comprehension levels (Driana & Ernawati, 2019). A study assessing the competencies of 10 mathematics teachers in assessing HOTS found that 9 out of 10 teachers encountered difficulties in preparing and implementing tools to measure HOTS (Afifah & Retnawati, 2019). In comparison to national and international literature, the findings of this study suggest that teachers involved in this study demonstrated higher rates of preparing items that measure HOTS. This suggests that the training program implemented in this study may have positively contributed to teachers' competencies in preparing such items.

Upon analyzing items prepared for objectives across all grade levels, it was found that out of the total 200 items, 115 (57%) were categorized as measuring lower-level thinking skills, while 85 (43%) were categorized as measuring HOTS. Of the items measuring HOTS, 65 (77%) were at the analysis level, 19 (22%) at the evaluation level, and 1 (1%) at the creation level. The distribution of these items by grade level shows that 42% were at the 6th grade level, followed by 24% at the 5th grade level, 22% at the 7th grade level, and 12% at the 8th grade level. These findings are consistent with existing literature. Although the study indicates a positive improvement in teachers' competencies in preparing items measuring HOTS, the rate of such items in the prepared set remains limited to 43%. Similar results have been reported in studies by Kartal & İlgün Dibek (2021), Villarroel et al. (2021), and Yurdakul et al. (2020), which also show that training programs have improved teachers' competencies, yet a significant proportion of items still do not measure HOTS. Contrasting findings can also be found in the literature. For example, Ernst-Slavit and Pratt (2017) found that a majority (72%) of items posed by science teachers in a primary school 4th grade science course were aimed at measuring HOTS in the context of the rocks and minerals unit.

Both national and international literature consistently shows that teachers predominantly craft items assessing lower-level thinking skills. Our findings also reflect this trend, indicating that a majority of items prepared by teachers measure lower-level thinking skills. Several factors contribute to this: central exams primarily emphasize lower-order thinking skills (Arı & İnci, 2015; Keskin & Aydın, 2011), curriculum goals and assessments predominantly focus on lower-order thinking skills (Cangüven et al., 2017; Güven, 2014; Zorluoğlu et al., 2016; Zorluoğlu et al., 2017), concerns about student comprehension difficulties (Güleryüz & Erdoğan, 2018), and textbooks that do not adequately support HOTS (Biber & Tuna, 2017). Another significant factor is teachers' insufficient knowledge and experience in crafting items that measure HOTS.

This study aimed to enhance teachers' competencies in writing items that measure HOTS through a training program designed to increase their knowledge and skills. The findings indicate the success of the training program in achieving this objective. Although a majority of the items prepared by participating teachers assessed lower-level thinking skills, the rate of items measuring HOTS is notably higher compared to findings in both national and international literature. Studies referenced in this study report a combined average of 11.12% for higher-order thinking items in the literature (Ataş & Güneş, 2020; Ayvacı & Türkdoğan, 2009; Dindar & Demir, 2006; Gökulu, 2015; Güleryüz & Erdoğan, 2018; Özcan & Akcan, 2010; Şanlı & Pınar, 2017; Tanık & Saraçoğlu, 2011), while within our study, 43% of all items prepared by teachers measured HOTS. This suggests that the training program had a positive impact on teachers' ability to write items that measure HOTS. Furthermore, the success of our training program is supported by findings in both national and international literature (Abdullah et al., 2016; Ar, 2019; Çepni & Şenel Çoruhlu, 2010; Kartal & İlgün Dibek, 2021; Şenel Çoruhlu et al., 2008; Villarroel et al., 2021; Yip, 2004; Yurdakul et al., 2020). As evidenced by the literature, the training program focused on HOTS implemented by our researchers has been deemed successful.

When the types of items prepared according to the objectives are evaluated, it is understood that the items measuring lower-level thinking skills belong to the objectives that require psychomotor skills such as modelling. Some of the objectives of the items measuring lower-level thinking skills correspond to lower level thinking skills such as recall and comprehension. Some of the items measuring HOTS are those that require HOTS such as prediction and inference. This situation shows that the cognitive level of the outcome accompanies the prepared item. On the other hand, the results obtained from the study show that the teacher's item preparation experience is effective in preparing items that measure HOTS. Similar situation is also supported by the literature (Kurnaz Adıbatmaz ve Kutlu, 2020:117; Üstüner ve Şengül, 2004). When all the results of the study are evaluated, it is revealed that the curriculum outcomes and the teacher's item writing experience affect the competence of preparing items that measure HOTS, and that the applied training program applied in the scope of the study will significantly improve the competence of teachers to prepare items that measure HOTS.

Recommendations

- The training program implemented within this study could be widely adopted to enhance teachers' proficiency in preparing items that measure HOTS.
- Including practices targeting lower skills in in-service programs can help improve teachers' proficiency in preparing items that measure HOTS.
- Encouraging the use of open-ended items, particularly beyond multiple-choice items, in teachers' exams can effectively enhance the measurement of HOTS.
- Organizing comprehensive item writing workshops focused on HOTS, guided by central and local experts, could significantly enhance teachers' experience in item preparation.
- Providing training and guidance for teachers to develop items at the evaluation and creation levels can enhance their competence in preparing items at these higher levels.
- Conducting study based on long-term practices is crucial due to the significant role of teacher experience in item writing.
- In the scope of this study, teachers were not guided on sub-skills for item preparation, leading to some sub-skills being overlooked. Planning to address these sub-skills could improve teachers' competencies in preparing items across all sub-skills.

Acknowledgements or Notes

This study was produced from the doctoral thesis of the first author.

References

Abdullah, A. H., Mokhtar, M., Abd Halim, N. D., Ali, D. F., Tahir, L. M., & Kohar, U. H. A. (2016). Mathematics teachers' level of knowledge and practice on the implementation of higher-order thinking skills (HOTS). *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(1), 3-17. https://doi.org/10.12973/eurasia.2017.00601a

Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education*, *4*(1), 1-11. https://doi.org/ 10.11648/j.ijsedu.20160401.11

Afifah, I. R. N., & Retnawati, H. (2019, October). Is it difficult to teach higher order thinking skills? In *Journal of Physics: Conference Series* (Vol. 1320, No. 1, p. 012098). IOP Publishing.

Akpınar, E., & Ergin, Ö. (2006). Fen bilgisi öğretmenlerinin yazılı sınav sorularının değerlendirilmesi[Evaluation of science teachers' written exam items]. *Milli Eğitim Dergisi[Journal of National Education],* *35*(172), 225-231.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock. (2018). Bloom'un eğitim hedefleri ile ilgili sınıflamasının güncelleştirilmiş biçimi[An updated version of Bloom's classification of educational objectives] (Third Edition). Durmuş Ali Özçelik (Trans.). Ankara: Pegem Academy, XX-115.

Antonio, R. P., & Prudente M. S. (2024). Effects of inquiry-based approaches on students' higher-order thinking skills in science: A meta-analysis. International Journal of Education in Mathematics, Science, and Technology (IJEMST), 12(1), 251-281. https://doi.org/10.46328/ijemst.3216

Ar, M. E. (2019). *Fen Bilimleri Öğretmenlerine Yönelik Geliştirilen Nitelikli Yaşam Temelli Açık Uçlu Soru Hazırlama Kursunun Uygulanması ve Değerlendirilmesi[Implementation and Evaluation of Qualified Life-Based Open-Ended* Item *Preparation Course Developed for Science Teachers].* Yayınlanmamış Yüksek Lisans Tezi[Unpublished Master Thesis], Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü[Uludağ University Institute of Educational Sciences], Bursa.

Arı, A., & İnci, T. (2015). Sekizinci sınıf fen ve teknoloji dersine ilişkin ortak sınav sorularının değerlendirilmesi[Evaluation of the common exam items related to the eighth grade science and technology course]. *Uşak Üniversitesi Sosyal Bilimler Dergisi[Usak University Journal of Social Sciences],* *8*(4), 17-50. https://doi.org/10.12780/uusbd.75570

Ataş, E., & Güneş, P. (2020). Altıncı sınıf fen bilimleri dersi sınav sorularının yeniden yapılandırılmış Bloom taksonomisine göre değerlendirilmesi[Evaluation of sixth grade science exam items according to the restructured Bloom's taxonomy]. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi[Bolu Abant Izzet Baysal University Journal of Faculty of Education], 20* (2), 1066-1078. https://doi.org/10.17240/aibuefd.2020..-632040

Ayvacı, H. Ş., & Türkdoğan, A. (2009). Yeniden yapılandırılan bloom taksonomisine göre fen ve teknoloji dersi yazılı sorularının incelenmesi[Analysing science and technology course written items according to the restructured bloom taxonomy]. *Türk Fen Eğitimi Dergisi[Journal of Turkish Science Education],* *7*(1), 13-25.

Boyd, B. T. (2008). Effects of state tests on classroom test items in mathematics. *School Science and Mathematics*, *108*(6), 251-262.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom* (First Edition). USA: ASCD.

Cangüven, H. D., Öz, O., Binzet, G., & Avcı, G. (2017). Millî Eğitim Bakanlığı 2017 fen bilimleri taslak programının Yenilenmiş Bloom Taksonomisi'ne göre incelenmesi[Examining the draft science program of the Ministry of National Education 2017 according to the Revised Bloom's Taxonomy]. *International Journal of Eurasian Education and Culture*, 2, 2-80.

Cansüngü Koray, Ö. & Yaman, S. (2002). Fen bilgisi öğretmenlerinin soru sorma becerilerinin bloom taksonomisine göre değerlendirilmesi[Evaluation of science teachers' item asking skills according to bloom taxonomy]. *Kastamonu Eğitim Dergisi[Kastamonu Education Journal], 10*(2), 317-324.

Chaffee, J. (1994). Thinking critically (Fourth Edition). Boston: Houghlon Mifflin Company.

Çepni, S., & Çoruhlu, T. Ş. (2010). Alternatif ölçme ve değerlendirme tekniklerine yönelik hazırlanan hizmet içi eğitim kursundan öğretime yansımalar[Reflections on teaching from the in-service training course on alternative assessment and evaluation techniques]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi[Pamukkale University Journal of Education], 28*(28), 117-128.

Çoruhlu, T. Ş., Nas, S. E., & Çepni, S. (2008). Fen ve teknoloji öğretmenleri için alternatif ölçme ve değerlendirme tekniklerine yönelik bir hizmet içi eğitim programından yansımalar: Trabzon örneği[Reflections from an in-service training program on alternative assessment and evaluation techniques for science and technology teachers: Trabzon case]. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi[Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education], 2*(2), 1-22.

Davey, L. (2009). The application of case study evaluations. *Elementary Education Online, 8*(2),1-3.

Dindar, H., & Demir, M. (2006). Beşinci sınıf öğretmenlerinin fen bilgisi dersi sınav sorularının Bloom Taksonomisi'ne göre değerlendirilmesi[Evaluation of fifth grade teachers' science course exam items according to Bloom's Taxonomy]. *Gazi Eğitim Fakültesi Dergisi[Gazi University Journal of Gazi Educational Faculty (GUJGEF], 26*(3), 87-96.

Driana, E., & Ernawati, E. (2019). Teachers' understanding and practices in assessing higher order thnking skills at primary schools. *Acitya: Journal of Teaching and Education*, *1*(2), 110-118. https://doi.org/10.30650/ajte.v1i2.233.

Ernst-Slavit, G., & Pratt, K. L. (2017). Teacher items: Learning the discourse of science in a linguistically diverse elementary classroom. *Linguistics and Education, 40,* 1-10.

Gillham, B. (2000). Case Study Research Methods. *British Library Cataloguing in Publication Data.* Typeset by Paston Pre Press Ltd., Beccles, Suffolk. Printed and bound in Great Britain by TJ International Ltd., Padstow, Cornwall.

Gökulu, A. (2015). Fen ve teknoloji öğretmenlerinin yazılı sınav soruları ile TEOG sınavlarında sorulan fen ve teknoloji sorularının Yenilenmiş Bloom Taksonomisi'ne göre incelenmesi[Examination of science and technology teachers' written exam items and science and technology items asked in TEOG exams according to the Revised Bloom's Taxonomy]. *Route Educational and Social Science Journal*, *2*(2), 434-446.

Güleryüz, H., & Erdoğan, İ. (2018). Orta okul fen bilimleri dersi sınav sorularının Bloom'un bilişsel alan taksonomisine göre değerlendirilmesi: Muş İli örneği[Evaluation of secondary school science exam items according to Bloom's cognitive domain taxonomy: The case of Muş Province.]. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi[The Journal of Social Sciences],* *6*(1), 43-49. https://doi.org/10.18506/anemon.354846.

Güven, Ç. (2014). *6, 7, 8. Sınıflar Fen ve Teknoloji Dersi Öğretim Programındaki Soruların Yenilenmiş Bloom Taksonomisi'ne Göre İncelenmesi[Investigation of the Items in the 6th, 7th, 8th Grades Science and Technology Course Curriculum According to the Revised Bloom's Taxonomy].* Yayımlanmamış Yüksek Lisans Tezi[Unpublished Master Thesis], Ahi Evran Üniversitesi Fen Bilimleri Enstitüsü[Ahi Evran University Institute of Science and Technology], Kırşehir.

Güneş, F. (2012). Öğrencilerin düşünme becerilerini geliştirme[Developing students' thinking skills]. *Türklük Bilimi Araştırmaları[Journal of Turkology Research], 32*, 127-146.

İnceçam, B., Demir, E., & Demir, E. (2018). Ortaokul Öğretmenlerinin Sınıf İçi Ölçme ve Değerlendirmelerde Yazılı Yoklamalarda Kullandıkları Açık Uçlu Maddeleri Hazırlama Yeterlikleri[Secondary School Teachers' Competencies of Preparing Open-Ended Items Used in Written Assignments in Classroom Measurement and Assessment]. *İlköğretim Online[Primary Education Online], 17*(4). https://doi.org/10.17051/ilkonline.2019.506900

Karabulut, A. (2017). MEB'in yayınlamış olduğu örnek pisa sorularının 2015 Türkçe öğretim programı okuma anlama kazanımları çerçevesinde analizi[Analysis of sample Pisa items published by MoNE within the framework of reading comprehension acquisitions of 2015 Turkish curriculum]. *Uluslararası Eğitim Bilim ve Teknoloji Dergisi[International Journal of Educational Science and Technology], 3*(3), 166-174.

Kartal, M., & Dibek, M. İ. (2021). Madde yazma eğitiminin özel okulda çalışan ilköğretim İngilizce öğretmenlerinin maddenin kalitesini değerlendirmeye ilişkin yeterliklerine etkisi[The effect of item writing training on the competences of primary school English teachers working in private schools on evaluating item quality]. *Milli Eğitim Dergisi[Journal of National Education], 51*(236), 3345-3358. https://doi.org/10.37669/milliegitim.991124

Keskin, M. Ö., & Aydın, S. (2011). Seviye Belirleme Sınavı 6. sınıf fen ve teknoloji testinde çıkan biyoloji sorularının revize edilmiş taksonomi'ye göre incelenmesi[Examination of biology items in the 6th grade science and technology test of the Placement Test according to the revised taxonomy]. *Gazi Eğitim Fakültesi Dergisi[Gazi University Journal of Gazi Educational Faculty (GUJGEF], 31*(3), 727-742.

Koray, Ö., Altunçekiç, A., & Yaman, S. (2005). Fen bilgisi öğretmen adaylarının soru sorma becerilerinin Bloom taksonomisine göre değerlendirilmesi[Evaluation of pre-service science teachers' item asking skills according to Bloom's taxonomy]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi[Pamukkale University Journal of Education], 17*(17), 33-39.

Kurnaz Adıbatmaz, B., & Kutlu, Ö. (2020). Bilimsel düşünme becerilerinin ölçülmesi[Measurement of scientific thinking skills] (Birinci Baskı[First Edition]). Ankara:Pegem Akademi[Ankara:Pegem Academy].

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into practice, 32*(3), 131-137.

Marso, R. N. & Pigge, F. L. (1988, Nisan). An analysis of teacher-made tests: Testing practices, cognitive demands, and item construction errors. *The annual meeting of the national council on measurement in education,* New Orleans.

Millî Eğitim Bakanlığı[Ministry of National Education]. (2018). İlköğretim kurumları fen bilimleri dersi öğretim programı[Primary school science curriculum], Talim ve Terbiye Kurulu Başkanlığı[Presidency of the Board of Education], Ankara.

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: a case of critical thinking. *Research in Science Education, 37*(4), 353-369. https://doi.org/10.1007/s11165-006-9029-2

Mitana, J. M. V., Muwagga, A. M., & Ssempala, C. (2018). Assessment of higher order thinking skills: A case of Uganda Primary Leaving Examinations. *African Educational Research Journal, 6*(4): 240-249. https://doi.org/ 10.30918/AERJ.64.18.083

Mutlu, M., Uşak, M., & Aydoğdu, M. (2003). Fen bilgisi sorularının bloom taksonomisine göre değerlendirilmesi[Evaluation of science items according to bloom taxonomy]. *Kırşehir Eğitim Fakültesi Dergisi[Ahi Evran University Journal of Kırşehir Faculty of Education], 4*(2), 87-95.

National Research Council. (2007). Taking Science to School; NRC. Washington, DC: The National Academy Press.

Oescher, J., & Kirby, P. C. (1990, April). Assessing teacher-made tests in secondary math and science classrooms. *The anual meeting of the national council on measurement in education*, Boston.

Özcan, S., & Akcan, K. (2010). Fen bilgisi öğretmen adaylarının hazırladığı soruların içerik ve Bloom taksonomisi'ne uygunluk yönünden incelenmesi[Analysing the items prepared by prospective science teachers in terms of content and appropriateness to Bloom's taxonomy]. *Kastamonu Eğitim Dergisi[Kastamonu Education Journal], 18*(1), 323-330.

Punch, K. F. (2013). Introduction To Social Research: Quantitative and Qualitative Approaches. New York: Sage.

Şanlı, C., & Pınar, A. (2017). Sosyal bilgiler dersi sınav sorularının Yenilenen Bloom Taksonomisi'ne göre incelenmesi[Analysing social studies course exam items according to Bloom's Taxonomy]. *İlköğretim Online[Primary Education Online], 16*(3), 949-959. https://doi.org/10.17051/ilkonline.2017.330234

Tanık, N., & Saraçoğlu, S. (2011). Fen ve teknoloji dersi yazılı sorularının Yenilenmiş Bloom Taksonomisi'ne göre incelenmesi[Analysing science and technology course written items according to the Revised Bloom's Taxonomy]. *TÜBAV Bilim Dergisi[Journal of TUBAV Science], 4*(4), 235-246.

Üstüner, A., & Şengül, M. (2004). Çoktan seçmeli test tekniğinin türkçe öğretimine olumsuz etkileri[Negative effects of multiple choice test technique on Turkish language teaching]. *Fırat Üniversitesi Sosyal Bilimler Dergisi[Journal of Social Sciences], 14* (2), 197-208.

Villarroel, V., Bruna, D., Brown, G. T. L., & Bustos, C. (2021). Changing the quality of teachers' written tests by implementing an authentic assessment teachers' training program. *International Journal of Instruction, 14*(2), 987-1000. https://doi.org/10.29333/iji.2021.14256a

Wilson, D. M., & Narasuman, S. (2020). Investigating teachers' implementation and strategies on higher order thinking skills in school-based assessment instruments. *Asian Journal of University Education*, *16*(1), 70-84. https://doi.org/10.24191/ajue.v16i1.8991

Yenilmez, K., & Yolcu, B. (2007). Öğretmen davranışlarının yaratıcı düşünme becerilerinin gelişimine katkısı[Öğretmen davranışlarının yaratıcı düşünme becerilerinin gelişimine katkısı]. *Manas Üniversitesi Sosyal Bilimler Dergisi[Manas Journal of Social Studies], 9*(18), 95-105.

Yıldız, D. Ç. (2015). The analysis of Turkish course exam items according to re-constructed Bloom's taxonomy. *Gaziantep University Journal of Social Sciences*, *14*(2), 479-497. https://doi.org/10.21547/jss.256771

Yin, R. K. (2003). Case study research design and methods (Third Edition). London: Sage Publications.

Yip, D. Y. (2004). Questioning skills for conceptual change in science instruction. *Journal of Biological Education*, *38*(2), 76-83.

Yurdakul, B., Başokçu, T. O. & Yazıcılar, Ü. (2020). Evaluation of the professional development program for secondary math teachers on item writing related to higher order thinking skills. *Journal of Teacher Education and Educators*, *9*(1), 83-106.

Yüksel, M. (2022). PISA 2018 araştırma sonuçlarına göre ülkelerin bileşik PISA performans sıralaması[Ranking the composite PISA performance of countries according to PISA 2018 research results]. *Muğla Sıtkı Koçman Üniversitesi Eğitim Fakültesi Dergisi* [*MSKU Journal of Education*], *9*(2), 788-821.

Zorluoğlu, S. L., Kızılaslan, A. & Sözbilir, M. (2016). School chemistry curriculum according to revised bloom taxonomy. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *10*(1), 260-279.

Zorluoğlu, S. L., Şahintürk A. & Bağrıyanık K. E. (2017). 2013 yılı Fen Bilimleri Öğretim Programı kazanımlarının Yenilenmiş Bloom Taksonomisi'ne göre analizi ve değerlendirilmesi[Analysis and evaluation of 2013 Science Curriculum outcomes according to the Revised Bloom's Taxonomy.]. *Bartın Üniversitesi Eğitim Fakültesi Dergisi[Bartın University Journal of Faculty of Education],* *6*(1), 1-15. Doi: https://doi.org/10.14686/buefad.267190