Journal name	International e-Journal of Educational Studies
Abbreviation	IEJES
e-ISSN	2602-4241
Founded	2017
Article link	http://doi.org/10.31458/iejes.1314305
Article type	Research Article
Received date	14.06.2023
Accepted date	26.09.2023
Publication date	21.10.2023
Volume	7
Issue	15
pp-pp	568-583
Section Editor	Prof.Dr. Naim UZUN
Chief-in-Editor	Prof.Dr. Tamer KUTLUCA
Abstracting & Indexing	Education Source Ultimate Database Coverage List EBSCO Education Full Text Database Coverage List H.W. Wilson Index Copernicus DRJI Harvard Library WorldCat SOBIAD
Article Name	The Effect of Question Preparation Training Program that Measures Higher Order Thinking Skills on the Self-Efficiency of Science Teachers

Author Contribution Statement

[1] **Ahmet BOLAT**
Lecturer
Hitit University, Turkey

[2] **Sevilay KARAMUSTAFAOĞLU**
Prof.Dr.
Amasya University, Turkey

Conceptualization, literature review, methodology, implementation, data analysis, translation, and writing

Methodology, data analysis, translation, and writing

Abstract

In this study, the aim was to examine the effect of the question preparation training program that measures higher order thinking skills on the self-efficacy of science teachers. The research is in a pre-experimental design model with quantitative origin pre-test/post-test application. The example of the study comprises of 25 science teachers working in public middle schools in Çorum. "Question Developing Self-Efficacy Scale Measuring High Level Learning Level of Science Teachers" was used as data collection tool. Descriptive analysis, difference analysis and effect size analysis were used in the assessment of the data. The t-test for dependent samples were employed to compare the teachers' self-efficacy before the application and their self-efficacy at the end of the application. With the research, it was confirmed that the applied training program significantly increased teachers' self-efficacy in preparing questions measuring higher order skills. Herewith, multifarious proposals were made, like turning the training program into a book, and arranging the content of the in-service training programs by considering the items with low item averages.

To cite this article:

Bolat, A., & Karamustafaoğlu, S. (2023). The effect of question preparation training program that measures higher order thinking skills on the self-efficiency of science teachers. *International e-Journal of Educational Studies*, *7* (15), 568-583. https://doi.org/10.31458/iejes.1314305

Copyright © IEJES

Research Article

The Effect of Question Preparation Training Program that Measures Higher Order Thinking Skills on the Self-Efficiency of Science Teachers*

Ahmet BOLAT [1] **Sevilay KARAMUSTAFAOĞLU** [2]

Abstract

In this study, the aim was to examine the effect of the question preparation training program that measures higher order thinking skills on the self-efficacy of science teachers. The research is in a pre-experimental design model with quantitative origin pre-test/post-test application. The example of the study comprises of 25 science teachers working in public middle schools in Çorum. "Question Developing Self-Efficacy Scale Measuring High Level Learning Level of Science Teachers" was used as data collection tool. Descriptive analysis, difference analysis and effect size analysis were used in the assessment of the data. The t-test for dependent samples were employed to compare the teachers' self-efficacy before the application and their self-efficacy at the end of the application. With the research, it was confirmed that the applied training program significantly increased teachers' self-efficacy in preparing questions measuring higher order skills. Herewith, multifarious proposals were made, like turning the training program into a book, and arranging the content of the in-service training programs by considering the items with low item averages.

Keywords: Science teacher, teacher education, measuring and assessment, question preparation, higher order thinking skills

1. INTRODUCTION

Today's individuals are expected to have high-level thinking skills to satisfy the demands of the modern age. For this reason, more emphasis has been placed on higher-order thinking skills in curricula in recent years. High-level thinking skills are also emphasized in the current science curriculum (Ministry of National Education, [MoNE], 2018). Measuring and evaluating these skills has an important place in the improvement of students' higher-order thinking skills (Risner, 1987).

Higher-order thinking skills require using knowledge rather than having knowledge (Wellman, 1997). In particular, the individual should be able to solve real life problems using the knowledge he has. Individuals with high-level thinking skills identify the problem in real life, develop a method for solving the problem, present a hypothesis, test the hypothesis, and decide the correctness of the hypothesis for the solution. In other words, he tries to figure out the issues he encounters in his everyday life with a scientific approach.

Students' higher-order thinking skills can be developed (Greeno, 1989). One of the significant factors in the development phase is teachers. Teachers make crucial contributions to the improvement of students' (pupils') higher-order thinking skills through education. In order to ensure the

Received Date: 14/06/2023 **Accepted Date:** 26/09/2023 **Publication Date:** 21/10/2023

*__To cite this article:__ Bolat, A., & Karamustafaoğlu, S. (2023). The effect of question preparation training program that measures higher order thinking skills on the self-efficiency of science teachers. *International e-Journal of Educational Studies, 7* (15), 568-583. https://doi.org/10.31458/iejes.1314305

[1] *Lecturer, Hitit University, ahmbolat@yahoo.com, Çorum, Türkiye*

[2] *Prof.Dr., Amasya University, sevilayt2000@yahoo.com, Amasya, Türkiye*

Corresponding Author e-mail adress: ahmbolat@yahoo.com

development of higher-order thinking skills, it is necessary to measure and evaluate the level of these skills. High-level thinking skills are difficult to measure because they include many thinking skills. Teachers should have knowledge and experience specific to this subject (Driana & Ernawati, 2019).

When the national literature on high-level thinking skills is searched, it is understood that there are sworks to determine the cognitive levels of the questions in the textbooks (Çakıcı & Girgin, 2012; Doğan, 2019). In their study, Çakıcı and Girgin (2012) analysed unit assessment questions in secondary school science textbooks. According to the research, the questions mostly belong to traditional question types and are intended to measure low-level thinking skills. Doğan (2009) examined the questions in the 4th grade science textbooks in his research. At the end of the research, the questions mostly belong to traditional question types and are intended to measure low-level thinking skills. Similarly, in a study in the international literature, Risner (1987) examined the epistemic level of the questions in the 5th grade science textbook in his research. She determined that the cognitive level of the questions was below the evaluation level and that they were mostly not intended to measure higher-order thinking skills. According to these studies, textbooks give little or no place to high-level thinking skills (Çakıcı & Girgin, 2012; Doğan, 2019; Risner, 1987). According to the research conducted by Akpınar and Ergin (2006), 1% of the questions asked by science teachers in exams measure high-level thinking skills. In the study conducted by Ayvacı and Türkdoğan (2010), the questions in the exams prepared by the 6th grade science teachers were analysed in regard to the revised Bloom Taxonomy. At the end of the research, 32.1% of the questions consisted of high-level questions. Cansüngü-Koray and Yaman (2002) examined the question preparation skills of science teachers according to Bloom's Taxonomy. 3.83% of the questions arranged by science teachers are aimed at measuring high-level thinking skills. In the study conducted by Özüuygun (2004), 26.9% of the questions prepared by science teachers at the 6th grade are questions that measure high-level thinking skills. According to Dindar and Demir's (2006) study, it was determined that 1.25% of the questions prepared by science teachers in the 5th grade class measured high-level thinking skills. In the study conducted by Mutlu, Uşak and Aydoğdu (2003) they classified and compared the questions prepared by the science teachers working in primary schools and the science questions asked in the High School Entrance Exam (HSEE) according to the Bloom Taxonomy. Accordingly, while 1% of the questions prepared by the teachers' measure their high-level thinking skills, 52% of the 2001 and 2002 LGS questions measure their high-level thinking skills. Güven (2014) examined the questions in the science and technology curriculum published in 2006 in his research. At the end of the research, it was determined that the questions in the program were mostly questions measuring low-level thinking skills. In the study conducted by Umur (2019), it was determined that undergraduate and graduate students of science teaching were insufficient in preparing questions suitable for the outcome. Similar results have emerged in studies conducted in the international literature. The questions prepared by both science teachers and teachers of other courses are insufficient to measure higher-order thinking skills (Driana & Ernawati, 2019; Marso & Pigge, 1988).

In his research, Ar (2019) organized an in-service training program on life-based open-ended question preparation for science teachers. At the end of the study, it was determined that there was a positive change in the thoughts of teachers about preparing life-based open-ended questions. In addition, it has been determined that there is research in the quality of life-based open-ended questions prepared by teachers. Similarly, there are studies that support that applied training programs improve teachers' competence in preparing questions that measure high-level thinking skills (Yip, 2004).

According to the national literature examined, it is seen that studies on questions measuring high-level thinking skills in our country are mostly in the screening model. However, it is understood that there are limited number of experimental studies on teachers' question writing competencies or development (Ar, 2019). It is thought that this study will contribute to the literature at this point. It is thought that the development of teachers' self-efficacy in preparing questions that measure high-level

thinking skills will be beneficial in the development of teacher competencies. The problem of this study is "What is the change in the self-efficacy of science teachers in preparing questions measuring high-level skills in the Education Program for Measuring High-Level Cognitive Skills?" and accordingly, the aim of the study is to investigate the effectiveness of the question preparation training program that measures high-level cognitive skills for science teachers.

2. METHOD

2.1. Research Design

This research was implemented in accordance with a pre-experimental model with a quantitative approach with pretest-posttest application. In this research method, a pre-test is performed to state the level of behaviour of the teacher or students in any subject before the application, and a post-test is applied to determine the level of behaviour at the end of the application. However, there is no control group in the study. The aim here is to reveal the impact of the interference applied to the empirical group on the development of the determined skill (Çepni, 2010). In this study, the change in the self-efficacy of science teachers in preparing questions measuring high-level skills of the "Training Program for Measuring "Higher Order Thinking Skills" applied to the experimental group was examined.

2.2. Universe and Sample

The universe of the research comprises of science teachers working in public middle schools in Çorum in Turkey. The sample of the study is 25 science teachers who worked in Çorum in Turkey and participated in this study. In the selection of the sample, the typical case sampling technique was used from non-probability based sampling techniques. Purposeful typical sampling the sample consists of individuals with an average level of knowledge who are representative of the universe and are considered appropriate for the purpose of the research (Canbazoğlu-Bilici, 2019). Since the thesis study in which this study was obtained was conducted in the field of science education, science teachers were preferred as a sample. The characteristics of these teachers are introduced in Table 1.

Table 1. Statistical distribution of the demographic characteristics of the teachers constituting the sample

Features		f	(%)	Features		f	(%)
Seniority	0-5 Years	1	4	Gender	Male	13	52
	6-10 Years	7	28		Female	12	48
	11-15 Years	8	32		Total	25	100
	16-20 Years	6	24				
	21-25 Years	2	8	Education Level	Licence	20	80
	26 Year and Above	1	4		Master	5	20
	Total	25	100		Total	25	100

While Table 1 is examined, it is understood that the sample is proportionally close to each other with regards to gender, and mostly graduates in terms of education level. In terms of seniority, it is ufound out that the most of teachers have professional experience of 6-20 years. In experimental studies, at least 10-20 sample groups are sufficient for research (cited in Roscoe, 1975; Büyüköztürk, Çakmak, Akgün, Karadeniz & Demirel, 2012). Since the sample size of this study is 25, it can be considered that it meets the minimum sample size requirement for an experimental study.

2.3. Research Process

This study aims to investigate the change in the self-efficacy of science teachers in preparing questions measuring high-level skills in the "Training Program for Measuring High Level Cognitive Skills". For this purpose, 16 weeks (96 hours) online training was given to science teachers working in official secondary schools in Çorum.

The training program was developed by the researcher. The program has been developed in accordance with the Taba Curriculum Development Model. The Taba Model attaches importance to the development of the program by the people who will implement the program. The Taba Model is implemented by following the stages of determining entails, determinant aimss, selecting content, organizer content, choosing learning experiences, organizer learning activities, and evaluating (Erişen, 1998).

While developing the education program, the scale used in the research was applied to the teachers. In addition, interviews with teachers were made to determine the competencies and needs of science teachers regarding high-level thinking skills and measurement. In addition, by scanning the literature, information was obtained about the competencies of teachers in this subject. Thanks to these studies, a needs analysis was carried out. Depending on the data derived from the needs analysis, the objectives of the training program were established. According to the established objectives, 4 acquisitions were prepared. In order to achieve the gains, the main subject headings were created. The basic titles are gathered under 5 titles as “Basic concepts and principles in Measurement and Evaluation in Education”, “Higher Order Thinking Skills”, “Question Development”, “Question Development to Measure Higher Order Thinking Skills”, “Basic Stages of the Test Plan Process”. Then, basic headings were listed and subheadings were prepared. The implementation of the training program was made through distance education due to the global epidemic of covid19. Each field expert completed the training in two stages. In the first stage, theoretical education was given, and in the second stage, an application was made depending on the theoretical education. During the applications, the teachers prepared questions according to the theoretical training given in the first stage. Each prepared question was examined by the field expert and feedback was given to the teachers. After the training program was implemented, the scale was applied again.

The training program started to be implemented in March 2021. 14 different trainings were applied to science teachers. Science education specialists, program development specialists and assessment and evaluation specialists took part in the trainings. The trainings were implemented in two sessions each week. Science teachers who participated in the training prepared questions about the achievements in the determined science curriculum. The prepared questions were examined by expert educators and they gave feedback to the teachers.

In the first week of the training, the assessment and evaluation specialist gave training on the basic concepts in assessment and evaluation. As a part of the training, the points to be considered during the preparation of valid and reliable measurement tools are explained within the framework of the concepts of measurement, evaluation, validity and reliability. In the second week of the training, training on high-level thinking skills was given by the program development specialist. Within the framework of this subject, the concept of high-level skills, the determination of high-level skills and high-level thinking skills according to taxonomies are explained. In detail, it was focused on associating higher-order thinking skills with Bloom's Taxonomy. In the third week of the training, a science education expert gave training on the acquisitions and skills in the science curriculum. As a part of education, the study was conducted to analyze and classify the acquisitions in the science curriculum according to the cognitive levels and life skills in Bloom's taxonomy. In the fourth week of the training, open-ended question preparation training was given to measure high-level thinking skills. The training was implemented in two sessions by the assessment and evaluation specialist. Within the

scope of the training, open-ended questions, preparing rubrics, and preparing questions measuring high-level skills were studied and questions were prepared.

In the fifth week of the training, multiple choice question preparation training was given by the science education specialist. Within the scope of the training, the subject of writing multiple-choice questions suitable for the acquisitions and skills in the curriculum was explained and a question writing study was carried out. The prepared questions were examined together with the teachers and feedback was given to the teachers. In the sixth week of the training, the multiple-choice question preparation training that measures high-level thinking skills was given by an assessment and evaluation specialist. Within the scope of the training, skills, low-level and high-level thinking skills, real-life situations were emphasized. Then, questions measuring higher-order thinking skills were explained through examples. Multiple choice questions measuring high-level thinking skills were prepared by the teachers, and in the second session of the training, these questions were examined and feedback was given to the teachers. In the seventh week of the training, test development training was given by the science education specialist. Within the scope of the training, the steps of the test development process are explained. After the theoretical training, item and test analysis was carried out using the data of a previously applied science test. After the analyzes, the items to be tested were decided and the final version of the test was created. In the second month of the training, the questions prepared by the teachers were examined in two sessions with the participation of assessment and evaluation specialists, language specialists and science education specialists, and the questions were corrected. The questions designed by the teachers were examined separately with regard to scientific accuracy, measurement and evaluation principles, and grammar, and necessary corrections were made in the questions, and thus practical feedback was provided to the teachers. These sessions were held online for two months. The content of the training program is summarized in Table 2.

Table 2. Subject distribution of the training program by week

Week	Basic Subject	Subtopic
1	Basic Concepts and Principles in Measurement and Evaluation in Education	Measuring Evaluation Validity Reliability
2	Higher Order Thinking Skills	Critical Thinking Skill Analytical Thinking Skill Creative Thinking Skill Decision Making Skill Problem Solving Skill
3	Classification of Acquisitions and Skills	Bloom's Taxonomy Scientific Process Skills Life Skills Engineering-Design Skills
4	Open-Ended Question Preparation	Open-Ended Questions Rubric Preparation Preparing a Question to Measure High Level Skills
5	Preparing Multiple Choice Questions	Structure of Multiple Choice Problem Multiple Choice Question Types Writing Multiple Choice Questions Appropriate for Acquisitions and Skills Considerations
6	Preparing Multiple Choice Questions that Measure Higher Order Thinking Skills	Low Level Question Higher Order Thinking Skills Question Context and Features Real Life Situations Question Examples

7	Test Development Stages	Determining the Purpose of the Test Question Preparation Question Editing Pilot Application Main Application Item Analysis Choosing a Question for the Test
8	Redaction	

Continuation of Table 2.

9-16.	Redaction

The study was planned as a single-group empirical design with quantitative approach, pretest-posttest application. The processes of the study are presented in Figure 1.

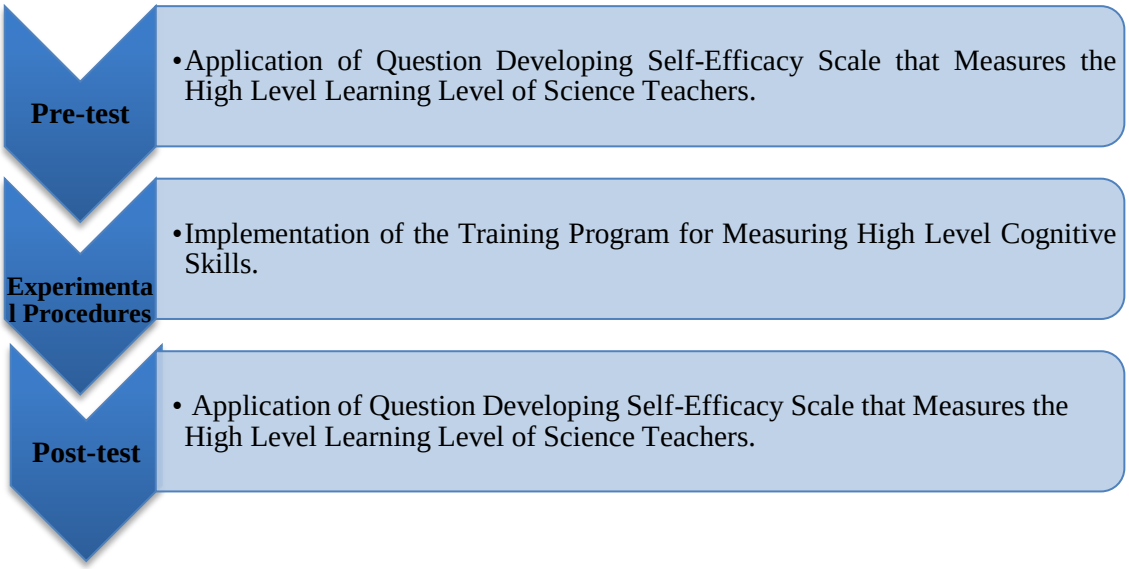


Figure 1. Flow Chart of the Research Process

Before the study, the Question Development Self-Efficacy Scale measuring the High-Level Learning Level of Science Teachers (Bolat, Korkmaz & Karamustafaoğlu, 2021) was applied as a pre-test by the researchers. "Training Program for Measuring High Level Cognitive Skills" was implemented to science teachers for 16 weeks. After the training program, the scale was applied as a post-test.

2.4. Data Collection Tool

In this study, "Question Development Self-Efficacy Scale Measuring Higher Order Thinking Skills of Science Teachers" developed by (Bolat et al., 2021) was used as a data collection tool. The scale is a five-point Likert scale. These options are; strongly disagree (1), disagree (2), undecided (3), agree (4) and strongly agree (5). The scale is a one-dimensional scale comprising of 30 items.

Table 3. Psychometric properties of the scale

Validity				**Reliability**	
Feature	**Value**	**Feature**	**Value**	**Feature**	**Value**
KMO	0,967	x^2/d	2,657	Cronbach Alfa	0,977
x^2	8057,031	RMSEA	0,84	Test Retest r_{min}	0,380 (p< 0,01 and p < 0,05)
sd	435	S-RMR	0,025	Test Retest r_{max}	0,836 (p< 0,01 and p < 0,05)
pp	0	NNFI	0,828	Test Retest r_{top}	0,833 (p < 0,05)
Factor Load (min)	0,713	CFI	0,884		
Factor Load (max)	0,856	GFI	0,768		
Variance Explanation Rate	63,86%	AGFI	0,729		
Continuation of Table 3.					
Common Variance Value	**> 0,50**				
t-test (lower-upper group)	< 0,05				
r_{min}	0,701				
r_{max}	0,826				

When Table 3 is examined, it can be understood that the psychometric properties of the scale are valid and reliable.

2.5. Analysis of Data

For the data analysis of the scale applied in the research, first of all, the arithmetic mean and standard deviation values were calculated. The differences between the individual averages of the items were calculated, and the change in the averages of the items was calculated according to the responses of the participants. According to the pre-test and post-test data obtained from the scale, the t-test was performed to state whether there was a crucial variation between the pre-test and post-test averages. Before the analysis, assumptions were tested for the parametric tests of the data. For this purpose, the dispersion of the data was analysed. When the distribution of the data was analysed, it was understood that the asymmetry and flatness values of both the pre-test and post-test data varied between +1 and -1. Since it was determined that the data were at equal interval scale level and both pre-test and post-test data were normally distributed, t-test was performed for dependent samples (Green & Salkind, 2005). The t-test for dependent samples is performed to state whether there is a crucial variation between the means of the tests if the same test is applied to the same group at certain time intervals (Can, 2019). To calculate the effect size (d) with the t-test for dependent samples, it was calculated by dividing the difference between the means of the measurements by the standard deviation of the series of difference scores [d=Difference between the means of measurement/Standard deviation of the difference scores] (Green & Salkind, 2005). An impact size greater than 1.0 is considered a huge impact, a large effect of 0.8, a medium effect of 0.5, and a minor effect of 0.2 (Morgan, 2004).

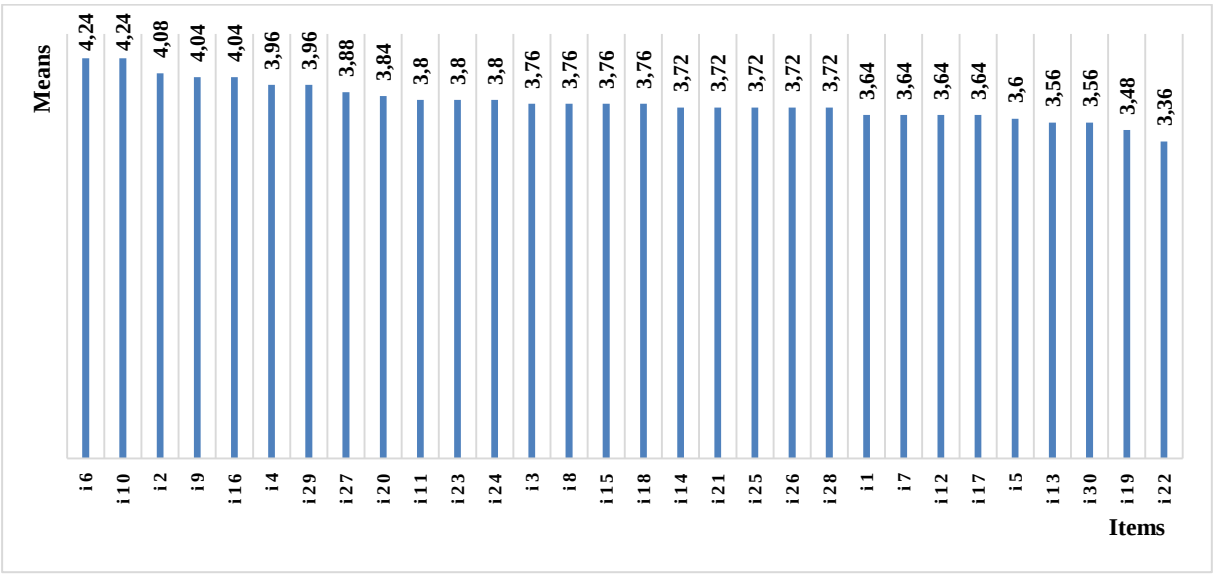
3. FINDINGS

To find a solution to the problem of the research, “Question Development Self-Efficacy Scale Measuring the High Level Learning Level of Science Teachers” was applied as a pre-test before the training program. The item averages calculated from the scale pre-test data are presented in Table 4.

Table 4. Descriptive analysis results of the pre-test data obtained from the scale data.

Item	n	$\overline{X}$	sd	Item	n	$\overline{X}$	sd
i1	25	3,64	0,64	**i16**	25	4,04	0,61
i2	25	4,08	0,49	**i17**	25	3,64	0,76
i3	25	3,76	0,78	**i18**	25	3,76	0,66
i4	25	3,96	0,68	**i19**	25	3,48	0,92
i5	25	3,6	0,87	**i20**	25	3,84	0,85
i6	25	4,24	0,66	**i21**	25	3,72	0,84
i7	25	3,64	0,81	**i22**	25	3,36	0,81
i8	25	3,76	0,72	**i23**	25	3,8	0,91
i9	25	4,04	0,61	**i24**	25	3,8	0,76
i10	25	4,24	0,72	**i25**	25	3,72	0,79
i11	25	3,8	0,71	**i26**	25	3,72	0,74
i12	25	3,64	0,81	**i27**	25	3,88	0,73
i13	25	3,56	0,92	**i28**	25	3,72	0,68
i14	25	3,72	0,79	**i29**	25	3,96	0,74
i15	25	3,76	0,66	**i30**	25	3,56	0,87

The item averages calculated from the scale pre-test data are also presented sequentially in Graph 1.



Graph 1. The item averages calculated from the scale pre-test data

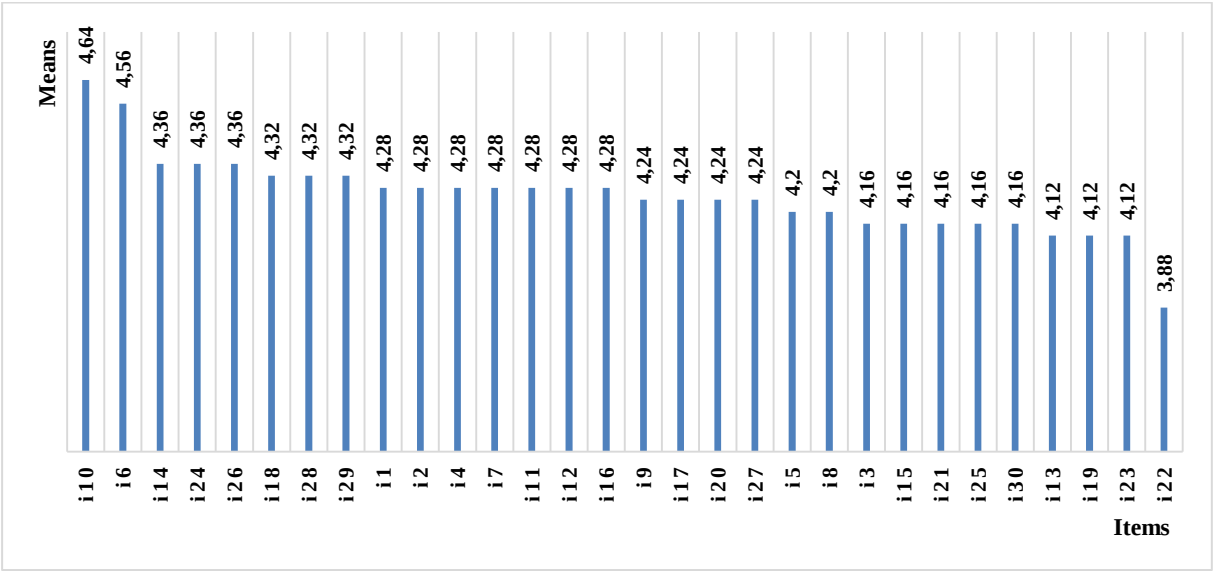
When Table 4 and Graph 1 is examined, it is understood that the item averages vary between 3.36 and 4.24. The 22nd item of the item with the lowest average is "*I can write a question item that measures the ability of students to share the product they have achieved*" is the item. The item with the highest average is item 10, "*I can write a question item that measures the ability of students to determine the independent variable in a given event*" is the item.

After the training program, "*Question Development Self-Efficacy Scale Measuring the High Level Learning Level of Science Teachers*" was applied as a post-test. The item averages calculated from the scale post-test data are presented in Table 5.

Tablo 5. Descriptive analysis results of the post-test data obtained from the scale data.

Item	n	$\overline{X}$	sd	Item	n	$\overline{X}$	sd
i1	25	4,28	0,46	**i16**	25	4,28	0,46
i2	25	4,28	0,46	**i17**	25	4,24	0,52
i3	25	4,16	0,55	**i18**	25	4,32	0,56
i4	25	4,28	0,46	**i19**	25	4,12	0,53
i5	25	4,20	0,58	**i20**	25	4,24	0,52
i6	25	4,56	0,51	**i21**	25	4,16	0,55
i7	25	4,28	0,54	**i22**	25	3,88	0,73
i8	25	4,20	0,65	**i22**	25	4,12	0,67
i9	25	4,24	0,44	**i24**	25	4,36	0,57
i10	25	4,64	0,49	**i25**	25	4,16	0,62
i11	25	4,28	0,54	**i26**	25	4,36	0,57
i12	25	4,28	0,54	**i27**	25	4,24	0,66
i13	25	4,12	0,53	**i28**	25	4,32	0,56
i14	25	4,36	0,49	**i29**	25	4,32	0,63
i15	25	4,16	0,62	**i30**	25	4,16	0,55

The item averages calculated from the scale post-test data are also presented sequentially in Graph 2.



Graph 2. The item averages calculated from the scale post-test data

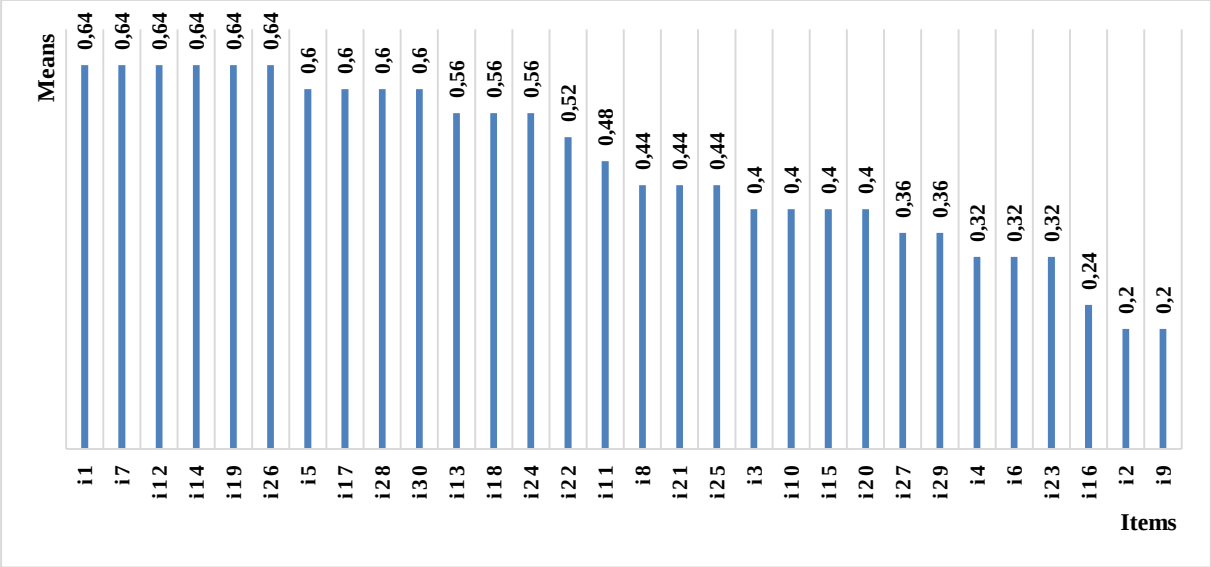
When Table 5 and Graph 2 is examined, it is understood that the item averages vary between 3.88 and 4.64. It is understood from the scale items that both the pretest and the posttest have the same item with the lowest mean and the highest mean. As in the pretest, the item with the lowest average is the 22nd item, "*I can write a question item that measures the ability of students to share the product they have achieved*" is the item. The item with the highest average is item10, "*I can write a question item that measures the ability of students to determine the independent variable in a given event*" is the item. The results of the change between the mean of the pre-test and post-test items are presented in Table 6.

Tablo 6. Descriptive analysis results of the posttest-pretest differences obtained from the scale data.

Item	n	$\overline{X}_{pre\text{-}test}$	$\overline{X}_{post\text{-}test}$	$\overline{X}_{post\text{-}test}$- $\overline{X}_{pre\text{-}test}$	Item	n	$\overline{X}_{pre\text{-}test}$	$\overline{X}_{post\text{-}test}$	$\overline{X}_{post\text{-}test}$- $\overline{X}_{pre\text{-}test}$
i1	25	3,64	4,28	0,64	**i16**	25	4,04	4,28	0,24
i2	25	4,08	4,28	0,20	**i17**	25	3,64	4,24	0,60
i3	25	3,76	4,16	0,40	**i18**	25	3,76	4,32	0,56
i4	25	3,96	4,28	0,32	**i19**	25	3,48	4,12	0,64
i5	25	3,60	4,20	0,60	**i20**	25	3,84	4,24	0,40

i6	25	4,24	4,56	0,32	**i21**	25	3,72	4,16	0,44
i7	25	3,64	4,28	0,64	**i22**	25	3,36	3,88	0,52
i8	25	3,76	4,20	0,44	**i23**	25	3,80	4,12	0,32
i9	25	4,04	4,24	0,20	**i24**	25	3,80	4,36	0,56
i10	25	4,24	4,64	0,40	**i25**	25	3,72	4,16	0,44
i11	25	3,80	4,28	0,48	**i26**	25	3,72	4,36	0,64
i12	25	3,64	4,28	0,64	**i27**	25	3,88	4,24	0,36
i13	25	3,56	4,12	0,56	**i28**	25	3,72	4,32	0,60
i14	25	3,72	4,36	0,64	**i29**	25	3,96	4,32	0,36
i15	25	3,76	4,16	0,40	**i30**	25	3,56	4,16	0,60

The results of the change between the mean of the pre-test and post-test items are also presented sequentially in Graph 3.



Graph 3. The difference between the item averages calculated from the scale pre-test/post-test data

When Table 6 and Graph 3 is examined, it is understood that there is an increase in the averages of all items in the post-test compared to the pre-test. Items with the least increase, item 2, "*I can write a question item that measures students' ability to identify significant similarities and differences between objects or events*" with the 9th item "*I can write a question item that determines the ability of students to measure some kind of magnitude*" are items. The items with the highest increase are the 1st item, "*I can write a question item that measures the ability of the students to record the data obtained as a result of observation and measurement in accordance with the purpose of the problem with various methods such as written expression, picture, table and drawing*" *I can write a question item that measures the ability of students to present their product in appropriate ways using verbal, written or visual materials*", item 12 "*I can write a question item that measures the ability of students to predict the possible solution to a problem they encounter*", item 14 "*I can write a question item that measures the ability to identify one or more of the most prominent variables in an event or relationship*", item 19, "I *can write a question item that measures students' ability to gather information from different sources*" and the 26th item "*I can write a question item that measures the level of being able to establish a relationship between the data collected by the students*" in the articles. Finally, necessary analyzes were made to state whether there was a crucial variation between the pre-test and post-test averages according to the collected data. Before the analysis, it was examined whether the data met the assumptions of the parametric tests. Since it was understood that both pre-test and post-test data were normally distributed, it was analyzed by t-test for dependent samples. Analysis results are presented in Table 4.

Table 4. t-test results for dependent samples of pretest-posttest data obtained from scale data

Measurement	n	$(\overline{X})$	sd	df	t	p*	d
Pre-test	25	3,78	0,57	24	-3,636	0,001	0,72
Post-test	25	4,25	0,41				
Difference Between Means		0,47	0,65				

p*:0,05

When Table 4 is analysed, it is understood that the pre-test mean ($\overline{X}_{pre\text{-}test}$=3,78), and the post-test mean ($\overline{X}_{post\text{-}test}$=4,25). When t-test results for dependent samples are assessed, it is seen that there is a crucial variation between pre-test mean score and post-test mean [$t_{(24)}$=-3,636, p<0,05].

4. DISCUSSION and CONCLUSION

The study was conducted to investigate the question preparation training program that measures high-level thinking skills for science teachers, and to investigate question preparation self-efficacy that measures science teachers' high-level thinking skills. The study was conducted as a single-group empirical design with pre-test/post-test application. 25 science teachers working in public schools in Çorum province attended in the study. As a data collection tool (Bolat et al., 2021), the "*Question Developing Self-Efficacy Scale that Measures the High-Level Learning Level of Science Teachers*" was used. Before the training program, the scale was implemented to the teachers participating in the study as a pre-test. A training program that lasted for 16 weeks was implemented to the teachers. After the training program, the scale was implemented to the teachers again. The study was reported by analysing the pre-test and post-test data.

According to the pre-test results of the scale, it was determined that the item averages ranged between 3.36 and 4.24. The 22nd item of the item with the lowest average is "*I can write a question item that measures the ability of students to share the product they have achieved*" is the item. The item with the highest average is item 10, "*I can write a question item that measures the ability of students to determine the independent variable in a given event*" is the item.

According to the post-test results of the scale, it is understood that the item averages vary between 3.88 and 4.64. It is understood from the scale items that the item with the lowest mean and the highest mean in both the pre-test and post-test are the same. As in the pre-test of the item with the lowest average, the 22nd item "*I can write a question item that measures the ability of students to share the product they have achieved*" is the item. The item with the highest average is "*I can write a question item that measures the ability of students to determine the independent variable in a given event*" is the item. According to the post-test/pre-test differences of the scale, it is understood that there is an increase in the averages of all items in the post-test compared to the pre-test. The 2nd item with the least increase was "*I can write a question item that measures students' ability to identify significant similarities and differences between objects or events*" in the article. The highest increase was item 26, "*I can write a question item that measures the level of being able to establish a relationship between the data collected by the students*" in the article. This situation may have arisen from the types of questions asked in the central exams applied in Turkey. Because teachers tend to use question types in this direction in their own exams (Güleryüz & Erdoğan, 2018). When the central exams applied in the last five years are examined, it is understood that the questions in which the variables are determined are directed to the candidates (2018-2022). It was understood that in the exams of the relevant years examined, there were at least two questions in some years, even five questions for determining the variables. In these exams, the question that the teachers answered at the lowest level according to the findings of the study and that measures the ability to share the product obtained was never included. The mentioned situations may have led to the findings obtained by affecting the question writing practices of the teachers. It may be due to the fact that the least increase

was in the preparation of questions that measure the ability to identify and measure the similarities and differences of objects, the psychomotor aspect of the measurement behavior and the difficulty of measuring psychomotor skills with cognitive diagnostic tools. On the other hand, the low increase in the average of the item about preparing a question that measures the ability to identify the similarities and differences of objects may be due to the fact that this item had a high average in the pretest as well.

When the results of the study are assessed in general, it can be said that the self-efficacy of science teachers in preparing questions measuring high-level thinking skills is high ($\overline{X}_{pre\text{-}test}$=3,78) and ($\overline{X}_{post\text{-}test}$=4,25). In the literature, there are studies that both support and do not support the results of this study. In the study carried out by Kılıç (2020), it was stated that teachers' perceptions of using alternative assessment and evaluation techniques were sufficient and very sufficient. Çakan (2004), Volante and Fazio (2007) determined in their research that teachers' self-efficacy for assessment is low. The voluntary participation of teachers in the training program in this study may indicate that they have a high interest in the subject. Self-efficacy is directly proportional to interest. Because psychological factors such as individuals' interests positively affect self-efficacy (Taylor & Bury, 2007). For these reasons, teachers who are interested in writing questions have high self-efficacy in preparing questions that measure their high-level thinking skills.

When the t-test results for dependent samples are evaluated on the data obtained from the scale, it is seen that there is crucial variation between the pre-test mean score and the post-test mean [$t_{(24)}$=-3,636, p<0,05]. It was observed that the effect of the experimental intervention on the change of science teachers' self-efficacy was moderate (d=0,72). There are studies that show that educational programs increase teachers' self-efficacy are found in the international literature. Gotch, Poppen, Razo, and Modderman (2021), in their research, examined the effect of the professional development training program they implemented on teachers' self-efficacy with different educational tasks, and they found that teachers' self-efficacy was very high and at the end of the training program, teachers' self-efficacy regarding assessment and evaluation was very high. In their study, In their study, Hartell, Gumaelius, and Svärdh (2015) examined the difference between the assessment and evaluation self-efficacy of 60 technology teachers who did not receive any training and 28 technology teachers who had a training program, and compared their assessment-evaluation self-efficacy. The self-efficacy perception of 28 teachers whose training program was applied was found to be significantly higher than the teachers who had never received any training. These results show parallelism with the views of Bandura (1976). According to Bandura, increasing the knowledge and experience of individuals increases self-efficacy.

When the results obtained from the research are compared with the national literature, it is understood that there are look-alike results. Çepni and Şenel-Çoruhlu (2010) determined that the in-service training course prepared for alternative assessment and assessment techniques positively affected the in-class assessment and evaluation competencies of science teachers. Şenel-Çoruhlu, Er Nas and Çepni (2008) determined that the in-service training course they organized for alternative assessment and evaluation techniques had positive changes in the perspectives and skills of science and technology teachers who attended the course. In the study carried out by Ar (2019), there was an increase in the competencies of science teachers in preparing life-based open-ended questions with the applied in-service program. Aslan (2011) determined that at the end of the curriculum he applied for Turkish language and literature teacher candidates, the question-writing skills that measure the higher order thinking skills of the pre-service teachers developed. Bay and Alisinanoğlu (2013) determined that the rate of questions measuring higher order thinking skills in the questions prepared by teachers increased after the training program applied to preschool teachers. Büyükalan-Filiz (2009) determined that the question-answer method training given to classroom teachers improved the skills of preparing

questions that measure teachers' higher order thinking skills. Cumhur (2016) determined that the questioning behaviors of teacher candidates changed positively at the end of the lesson study practice that he applied to the mathematics teachers' candidates. Gürbüz (2014) determined that the mathematics literacy curriculum that he applied to pre-service mathematics teachers increased the question-writing competencies of pre-service teachers that measure their higher order thinking skills. Research in both international, national literature and according to the results of this research, question preparation trainings that measure higher order thinking skills for teachers significantly increase teachers' self-efficacy in preparing questions that measure higher order thinking skills.

In consideration of the results obtained from the research, the undermentioned recommendations are presented:

* Widespread application of the training program applied within the scope of the research can be made.
* The training program can be applied to science teachers in different cities and the results of this research can be supported.
* Training programs for different branch teachers can be organized and their effectiveness can be investigated.
* The training program can be made into a book and made available to teachers.
* Statistical evaluations of teachers' self-efficacy can be examined through interviews and the underlying reasons for the results can be investigated.
* In this study, activities related to skills with low item averages can be designed in an enriched way in in-service training programs to be prepared for teachers.
* Due to the importance of experience in writing questions, the training programs to be prepared should be designed in such a way that teachers practice writing questions during the long process.
* In the training program to be prepared, planning can be made to cover the sub-skills of higher order thinking skills for the questions to be prepared by the teachers.

Acknowledgement

This research was produced from the doctoral thesis of the first author, supervised by the second author.

5. REFERENCES

Akpınar, E. & Ergin, Ö. (2006). Fen bilgisi öğretmenlerinin yazılı sınav sorularının değerlendirilmesi. *Milli Eğitim Dergisi, 35*(172), 225-231.

Ar, M. E. (2019). *Fen bilimleri öğretmenlerine yönelik geliştirilen nitelikli yaşam temelli açık uçlu soru hazırlama kursunun uygulanması ve değerlendirilmesi [Implementation and evaluation of a qualified life-based open-ended question preparation course developed for science teachers*]. (Unpublished master thesis, Uludağ University, Bursa, Turkey).

Aslan, C. (2011). Soru sorma becerilerini geliştirmeye dönük öğretim uygulamalarının öğretmen adaylarının soru oluşturma becerilerine etkisi [The effect of teaching practices aimed at developing questioning skills on teacher candidates' questioning skills]. *Education and Science, 36*(160), 236-249.

Ayvacı, H. Ş. & Türkdoğan, A. (2010). Yeniden yapılandırılan bloom taksonomisine göre fen ve teknoloji dersi yazılı sorularının incelenmesi [Examining the written questions of science and technology course according to the reconstructed bloom taxonomy]. *Journal of Turkish Science Education, 7*(1), 13-25.

Bandura, A. (1997). *Self-efficacy: the exercise of control.* New York: W.H. Freeman and Co.

Bay, D. N. & Alisinanoğlu, F. (2012). Okul öncesi öğretmenlerine verilen soru sorma becerisi öğretiminin öğretmenlerin sorularının bilişsel taksonomisine etkisi [The effect of questioning skills teaching given to preschool teachers on the cognitive taxonomy of teachers' questions]. *Journal of Mersin University Faculty of Education, 8*(3), 80-93.

Bolat, A., Korkmaz, Ö. & Karamustafaoğlu, S. (2021). Question development self-efficacy scale measuring high level learning level of science teachers: Validity and reliability. *Journal of Kirsehir Education Faculty, 22*(1), 372- 416.

Büyükalan-Filiz, S. (2009). Soru cevap yöntemi eğitiminin öğretmenlerin soru sorma bilgisi ve soru sorma tekniklerine etkisi [The effect of question and answer method training on teachers' questioning knowledge and questioning techniques]. *Kafkas University Journal of the Institute of Social Sciences, 1*(3), 167-195.

Can, A. (2019). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi [Quantitative data analysis in scientific research process with SPSS].* Ankara: Pegem Publishing House.

Canbazoğlu-Bilici, S. (2019). Örnekleme yöntemleri [Sampling methods]. In H. Özmen & O. Karamustafaoğlu (Ed.), *Eğitimde araştırma yöntemleri* [*Research methods in education*] (s.56-78), Ankara: Pegem Akademi.

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

Cumhur, F. (2016). *Matematik öğretmeni adaylarının soru sorma davranışlarının gelişiminin incelenmesi: Bir ders imecesi çalışması [Examining the development of pre-service mathematics teachers' questioning behaviors: A lesson study study].* (Unpublished doctoral dissertation, Karadeniz Technical University, Trabzon, Turkey).

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim [Teachers' assessment-evaluation practices and proficiency levels: primary and secondary education]. *Ankara University Journal of Faculty of Educational Sciences, 37*, 99-114.

Çakıcı, Y. & Girgin, E. (2012). İlköğretim II. kademe fen ve teknoloji ders kitaplarındaki ünite sonu değerlendirme sorularının incelenmesi [Primary education II. examining the end-of-unit evaluation questions in the secondary school science and technology textbooks]. *Erzincan University Journal of Faculty of Education, 14*(2), 87-110.

Çepni, S. (2010). *Araştırma ve proje çalışmalarına giriş analizi [Introductory analysis to research and project studies].* Trabzon: Celepler Publishing House.

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 prepared for alternative measurement and evaluation techniques]. *Pamukkale University Faculty of Education Journal, 28*(28), 117-128.

Çoruhlu, T. Ş., 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: The case of Trabzon]. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 2*(2), 1-22.

Dindar, H. & Demir, M. (2006). Beşinci sınıf öğretmenlerinin fen bilgisi dersi sınav sorularının bloom taksonomisine göre değerlendirilmesi [Evaluation of fifth grade teachers' science lesson exam questions according to Bloom's taxonomy]. *Gazi University Journal of Gazi Education Faculty, 26*(3), 87-96.

Doğan, N. B. (2019). *4. sınıf fen bilimleri ders kitaplarındaki değerlendirme sorularının çeşitli açılardan incelenmesi [Examination of evaluation questions in 4th grade science textbooks from various perspectives].* (Master's thesis, Necmettin Erbakan University, Konya, Turkey).

Driana, E. & Ernawati, E. (2019). Teachers' understanding and practices in assessing higher order thinking skills at primary schools. *Acitya: Journal of Teaching and Education, 1*(2), 110-118.

Erişen, Y. (1998). Program geliştirme modelleri üzerine bir inceleme [A review on curriculum development models]. *Educational Management in Theory and Practice, 13*(13), 79-97.

Gilligan, M. E. (2007). *Traditional versus alternative assessments: which type do high school teachers perceive as most effective in the assessment of higher-order thinking skills?* Saint Louis Univ..

Greeno, J. G. (1989). A perspective on thinking. *American Psychological Association, 44*(2), 134-141.

Gotch, C. M., Poppen, M. I., Razo, J. E. & Modderman, S. (2021). Examination of teacher formative assessment self-efficacy development across a professional learning experience. *Teacher Development, 25*(4), 534-548.

Güleryüz, H. & Erdoğan, İ. (2018). Ortaokul fen bilimleri dersi sınav sorularının Bloom'un bilişsel alan taksonomisine göre değerlendirilmesi: Muş İli örneği [Evaluation of middle school science course exam questions according to Bloom's cognitive domain taxonomy: The case of Muş Province]. *Anemon Mus Alparslan University Journal of Social Sciences, 6*(1), 43-49.

Gürbüz, M. Ç. (2014). *PISA matematik okuryazarlık öğretiminin pısa sorusu yazma ve matematik okuryazarlık düzeyleri üzerine etkisi [The effect of PISA mathematics literacy teaching on PISA question writing and mathematical literacy levels].* (Master's thesis, Uludağ University, Bursa, Turkey).

Güven, Ç. (2014). *6, 7, 8. sınıflar fen ve teknoloji dersi öğretim programındaki soruların yenilenmiş Bloom Taksonomisi'ne göre incelenmesi [Examining the questions in the 6th, 7th, 8th grade science and technology curriculum according to the revised Bloom's Taxonomy].* (Master's thesis, Ahi Evran University, Kırşehir, Turkey).

Hartell, E., Gumaelius, L. & Svärdh, J. (2015). Investigating technology teachers' self-efficacy on assessment. *International Journal of Technology and Design Education, 25*(3), 321-337.

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

Ministry of National Education [MoNE]. (2018). İlköğretim kurumları fen bilimleri dersi öğretim programı [Primary education institutions science course curriculum], Talim ve Terbiye Kurulu Başkanlığı, Ankara.

Mutlu, M., Uşak, M. & Aydoğdu, M. (2003). Fen bilgisi sorularının bloom taksonomisine göre değerlendirilmesi [Evaluation of science questions according to Bloom's taxonomy]. *Journal of Kirsehir Education Faculty, 4*(2), 87-95.

Özmen, H. (2019). *Deneysel araştırma yöntemi* [*Experimental research method].* In H. Özmen & O. Karamustafaoğlu (Ed.), *Eğitimde araştırma yöntemleri* [*Research methods in education] (İkinci Baskı)* (s.198-229), Ankara: Pegem Akademi.

Özüuygun, B. (2004). *İlköğretim fen bilgisi öğretmenlerinin soru sorma becerilerinin incelenmesi [Examination of primary school science teachers' questioning skills].* (Master's thesis, Marmara University, İstanbul, Turkey).

Umur, M. (2019). *Fen bilgisi programı lisans ve lisansüstü öğrencilerinin öğretim programı kazanımlarına uygun soru hazırlama yeterliliklerinin incelenmesi* [*Examination of the competence of undergraduate and graduate students of science program to prepare questions appropriate to the curriculum achievements].* (Master's thesis, Kastamonu University, Kastamonu, Turkey).

Risner, G. P. (1987). *Cognitive levels of questioning demonstrated by test items that accompany selected fifth-grade science textbooks* (Doctoral dissertation). Nashville: Vanderbilt University.

Taylor, D. & Bury, M. (2007). Chronic illness, expert patients and care transition. *Sociology of Health & Illness*, *2*9(1), 37-39.

Volante, L. & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749-770.

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

Wellman, L. E. (1997). The use of multiple representations, higher order thinking skills, interactivity, and motivation when designing a cd-rom to teach self-similarity (Master's theses, University of California, Los Angeles, USA).