ACTA DIDACTICA NAPOCENSIA

Volume 15, Number 1, 2022 - DOI: 10.24193/adn.15.1.11

INVESTIGATION OF ACHIEVEMENT TESTS PREPARED BY ELEMENTARY MATHEMATICS TEACHERS AND PRESERVICE TEACHERS

Feride ÖZYILDIRIM-GÜMÜŞ, Gülfem SARPKAYA-AKTAŞ, Hilmi KARACA

Abstract: The only purpose of assessment and evaluation activities should not be to grade students. On the contrary, it should determine the needs, misconceptions and readiness of students with assessment and evaluation activities and selection and placement processes are also carried out. For this reason, the qualifications of the assessment tools used, must be appropriate and sufficient. Although large scale tests give more importance to these qualities, these qualities may sometimes be ignored in teacher-made tests. Based on this point, within the scope of this study, the assessment tools prepared by mathematics teachers and preservice elementary mathematics teachers studying with them were examined in terms of some features. The study was carried out with five mathematics teachers working in five different public elementary schools in a medium-sized city in the Central Anatolia region and five preservice elementary mathematics teachers studying with these teachers. In this context, teachers and preservice teachers were asked to prepare an achievement test and to grade student answers for the mathematics classes they conduct together. In the method of the study carried out, the survey model was used, and qualitative and quantitative data were collected as it was investigated whether there was a difference and relationship between the scores of the students, in addition to the characteristics of the test items prepared. The findings are presented separately for each research question.

Key words: mathematics teacher, preservice mathematics teacher, cognitive demand task level.

1. Introduction

Assessment and evaluation are important activity that helps teachers to make decisions about the education process and determines students' learning road maps beside controlling students' learning at the end of the process (The National Council of Teachers of Mathematics [NCTM], 2000). For this reason, it can be said that, assessment and evaluation should not be regarded as an activity that is made only for grading individuals. Assessment is used for many purposes such as identifying student needs, evaluating teaching, grouping students, grading and motivating students (Stiggins, 1999). In addition, the information obtained during and at the end of the assessment process contributes to teachers' designing the teaching process appropriately for students (Little, 2009). This makes assessment an important component of the teaching process. For this reason, the tests used by teachers during the assessment process in classroom should provide information with high quality, to making decisions about learning, thinking and achievements of students' (DiDonate-Barnes, Fives, & Krause, 2014). Similarly, NCTM (2000) mentions that, teachers make decisions about what students know and what they need to learn based on their assessment results. This situation can make the assessment method preferences and results to be realized in the classroom an important factor in the teaching process. For this reason, it is important that the assessment methods and contents of tests that teachers and preservice teachers who are future teachers will use in the classroom have the necessary competence to reveal students' knowledge and skills.

Received April 2021.

Cite as: Özyıldırım-Gümüş, F., Sarpkaya-Aktaş, G., & Karaca, H. (2022) Investigation of Achievement Tests Prepared by Elementary Mathematics Teachers and Preservice Teachers. *Acta Didactica Napocensia,* 15(1), 124-141, https://doi.org/10.24193/adn.15.1.11

According to NCTM (2000), in order to make assessment more valuable, teachers should go beyond the "right or wrong" analysis and focus on how students think about tasks and try to determine students' understanding. For this reason, it becomes important to prepare the assessment tools used in the process, which serve the purpose and within the framework of certain standards. Because a well-structured assessment tool can improve the quality of the learning and teaching process. For this reason, teachers should consider students' age, experience and special needs when choosing assessment methods. Since using appropriate questions for students in assessment tools will help to show students what their minds can achieve (Ahmed & Pollitt, 2007).

From this point of view, the importance of assessment literacy becomes apparent in the preparation of assessment tools. Assessment literacy can be defined as understanding assessment principles; to be able to select, use and develop appropriate assessment methods and techniques for students; interpretation of the information obtained as a result of the assessment (DeLuca, Chavez, Bellara, & Cao, 2013; Mertler, 2004; Popham, 2004; Stiggins, 1991; Volante & Fazio, 2007). Assessment and assessment literacy have an important place in mathematics education as in all areas of education. One of the 6 principles that NCTM (2000) has determined for school mathematics is the assessment reveals the importance of assessment for teaching mathematics. NCTM (1989) explains the purpose of assessment in mathematics as a process that helps teachers to understand what students know and to make decisions for meaningful teaching. Because assessment is one of the main sources of information that teachers can access about students. Therefore, the quality of teachers' decisions will be as good as the meaningful information they can obtain from this source (NCTM, 2000).

Teachers with higher assessment literacy may be more aware of what they evaluate, why they do it, and how they can assessment the targeted skills. They can detect the situations that may affect the assessment results and the errors that may occur in the process and take precautions (Stiggins, 1991; Stiggins, 1995). In addition, it can be said that they will be successful in preparing good quality and suitable assessment tools. However, in the literature, it is mentioned that the teachers' and preservice teachers' assessment literacy was insufficient and that teachers who had just started the profession did not feel sufficient about assessment literacy (Maclellan, 2004; Mertler, 2004; Plake, 1993; Volante & Fazio, 2007). In other words, it is concluded that teachers and preservice teachers do not consider themselves sufficient in preparing an effective assessment tool. For this reason, assessment for teachers should be one of the focal points in the teacher training process and at every stage of the professional development of teachers (NCTM, 2000). Brookhart (2011) has defined the educational assessment knowledge and skills for teachers that can be used in this process. The main ones can be expressed as content of teaching, concept of assessment, students and class, school and family. In the title of teaching content, it is mentioned that the teacher has knowledge of the content, curriculum and curriculum taught. For the concept of assessment title, the teacher should know what assessment means and what its purpose is and be able to use the assessment tools appropriately. In addition, the teacher should be able to test the level of knowledge in the classroom, analyze the questions and evaluate the performances.

Furthermore, it is considered important that the test items used for assessment should address a certain cognitive level. Various frameworks have been developed to examine the cognitive dimensions of mathematical tasks (Ubuz, Erbaş, Çetinkaya, & Özgeldi, 2010). One of these frameworks is the level of cognitive demands developed as mathematical thinking levels within the scope of the QUASAR Project (Silver & Stein, 1996). Cognitive demands of task levels (CDTL) is defined as the level of thinking and form necessary for the student to successfully answer questions or problems within a task (Silver & Stein, 1996; Stein, Smith, Henningsen, & Silver, 2000). CDTL consists of four categories named as memorization, procedures without connections, procedures with connections and doing mathematics (Stein, Grover, & Henningsen, 1996; Stein & Smith, 1998). The first two of these cognitive demand levels, memorization and procedures without connections, are lower-level demands, and the last two, procedures with connections and doing mathematics, are higher-level demands (Smith & Stein, 1998; Stein et al., 1996; Stein et al., 2000).

In memorization level that is level 1, the tasks are easy to understand, repetitive or have no relationship between them. Tasks at this level are completed simply by knowing the memorized definitions and rules without any action. Activities at the level of procedures without connections

(level 2) are algorithmic and require limited cognitive skills to complete. There is no need to find a connection between the meaning or concepts underlying the processes used. Moreover, there are tasks that require connection with conceptual ideas at the level of procedures with connections (level 3), with direct or indirect paths to follow general operations. Graphics are evaluated at this level in tasks that require connection between multiple representations, such as visual diagrams, symbols. Lastly, there are non-algorithmic and complex tasks that require them to understand the nature of mathematical ideas, process, or connections at the level of doing mathematics (level 4). At this level, students are asked to question tasks, access relevant information, analyze and use them where appropriate (Henningsen & Stein, 1997; Smith & Stein, 1998; Stein et al., 2000).

In this study, CDTL that is containing four levels of cognitive demand, which is put forward by Stein and Smith (1998), is used to determine the cognitive demand levels of the prepared test items. Because this grouping is stated to be a suitable framework for analyzing secondary school mathematics tasks (Ubuz et al., 2010) and used in the analysis of many mathematical tasks (Arbaugh & Brown, 2005; Henningsen & Stein, 1997; Jones & Tarr, 2007; Stein & Smith, 1998).

In studies on assessment literacy, Plake, Impara and Fager (1993) determine that teachers' assessment literacy is higher than preservice teachers. Supporting this situation, Volante and Fazio (2007) express that preservice teachers' assessment literacy is relatively low. They also stated that most of the preservice teachers suggested result-oriented assessments. As an example of studies comparing assessment knowledge and experience, Mertler (1999) asks teachers about the level of assessment knowledge they had before graduation and received answers that it is quite low. When he asked the same group of available assessment knowledge levels, he received the mid-level response. Similarly, Mertler (2003) found that teachers' assessment literacy is higher than preservice teachers' assessment literacy. Mertler concluded that this potentially tends to improve teachers' assessment skills on the job as opposed to structured environments such as courses or workshops. On the other hand, Stiggins (1999), stated that many teachers are not at an adequate level to evaluate student learning, both at the end of the graduation and at the end of the postgraduate education process, and that they have the majority of their assessment information at work. Also, DiDonate-Barnes, Fives and Krause (2014), found that teachers had both theoretical and practical deficiencies in terms of class assessment competencies. They state that the reason for this situation is their teacher training process and lack of suitable strategies for test development. Similar to some study results with teachers, Maclellan (2004) state that preservice teachers' knowledge of assessment methods is not sufficient. In details, she determines that preservice teachers are insufficient in determining the assessment criteria for their tests, explaining what is expected from the test items, scoring and interpreting them.

In the literature, when the appropriateness of assessment and assessment tools in terms of curriculum and objectives of curriculum are analyzed, Özcan and Delil (2018) determine that determined that the exams prepared by the teachers are in high harmony with the achievements in the curriculum. However, they find that some questions are related to more than one objective and some questions are related to objectives of different grade levels. In another study examining the questioning behaviors of teachers and preservice teachers, it is determined that both groups prefer to ask questions with closed and low-level thinking, which have a short and one-answer (Çalık & Aksu, 2018). They also emphasized that the knowledge, skills and attitudes of teachers and preservice teachers should be improved in terms of using question preparation, questioning and questioning techniques.

Marso and Pigge (1991) stated that the cognitive levels of the exam items prepared by teachers are generally examined according to Bloom Taxonomy. In addition, they determine that most of the exam items examined are at the level of knowledge and comprehension and multiple choice, matching and short answer questions are used as the question type.

From all of this knowledge, the aim of this study is determined as examining the test items, prepared by elementary mathematics teachers and the preservice elementary mathematics teachers who they work together, in terms of some features. In addition, it is investigated whether there is a difference and a relationship between the scores of the students in the applied tests. It has been noticed that in in the literature, the studies examining the similarities or differences between the test items prepared by the teachers and preservice teachers are so limited. That's why, this study is expected to make an

important contribution to the literature with this feature. Another feature that distinguishes this study from other studies is that the cognitive levels of the items used in the exams are determined by CDTL. In most studies, the cognitive levels of the test items are determined according to the Bloom Taxonomy (Marso & Pigge, 1991). The reason for using CDTL in this study is that CDTL is a taxonomy developed to determine the cognitive dimensions of mathematical tasks.

Based on this point, three research problems presented below are examined within the scope of this study.

• What are the properties (item type, item number, learning and sub-learning area measured and cognitive demands of task levels) of the test items prepared by teachers and preservice teachers?

• Is there any difference between student scores in the tests prepared by teachers and preservice teachers?

• Is there a relationship between student scores in the tests prepared by teachers and preservice teachers?

2. Method

Research design, study group, data collection tool, data analysis, validity and reliability parts are discussed in this section.

2. 1. Research Design

Within the scope of this study, it is investigated what the properties of the test items prepared by the mathematics teacher and the preservice elementary mathematics teachers they worked with, whether there is a relationship and difference between the grades that the students obtained from these tests. In this context, survey model has been adopted as the research model of the study. Because according to Fraenkel and Wallen (2006), the purpose of survey studies is to define the characteristics of a group as it is. According to the research questions, both quantitative and qualitative data are collected and the data are analyzed separately consistent with the purpose of each research question.

Within the scope of the first research problem, it is aimed to examine the test items primarily in terms of item type, number, learning and sub-learning area measured and CDTL of test items. For this purpose, the data collected through document analysis are analyzed through a case study approach from qualitative research methods. Because in case studies, one or more cases are examined in depth and data can be collected through documents (Christensen, Johnson, & Turner, 2011).

Within the scope of the second and third research problems, since the mean grades taken by the students in the prepared tests and the relationships between these means are examined, a relational study is adopted from the quantitative research methods. Because, according to Christensen, Johnson and Turner (2011), it is aimed to explain the relationships between the variables in the relational studies.

2. 2. Study Group

Within the scope of the study, convenience sampling method has been used as the sampling method. In some studies, it is not possible for researchers to determine a sample randomly or systematically from the population and according to the purpose of the study, the data might be collected from a group that is reached (Fraenkel & Wallen, 2006). Since this research deals with the mathematics teachers working in public elementary schools and preservice elementary mathematics teachers who are studying with those teachers as the part of teaching practice course given at university, the researchers limited the study with the schools that they can reach. In this context, demographic features of the study group are presented in Table 1.

Table 1. *Demographic features of study group*

Group	School	Grade of students took the test	Number of students took the test	Professional experience of the teacher (years)	Sex of the teacher	Academic grade point average (GPA) of preservice teacher (over 4)	Sex of the preservice teacher
1	A	7	30	16	female	2.61	female
2	B	7	46	17	male	2.64	female
3	C	7	28	18	male	2.71	male
4	D	7	67	14	female	2.84	female
5	E	7	67	12	female	3.19	male

The schools involved in the study are five public elementary schools in a medium-sized city. The study is carried out with 5 mathematics teachers who worked in these schools and agreed to participate in the research and 5 preservice elementary mathematics teachers working with these teachers. In the table, each teacher and the preservice teacher with whom he/she works are considered as a group. Teachers and preservice teachers in the group carry out mathematics lessons with the same students in the same classes. In addition, the data obtained from students studying in the classes that teachers and preservice teachers work are also examined for two research problems of the study. All teachers and preservice teachers in the study group have taught seventh grade mathematics lessons. When the related table is analyzed in details, it is seen that the professional experience of mathematics teachers varies between 12 and 18 years, and the average academic grade of preservice teachers varies between 2.61 and 3.19 over 4.00. In other words, it can be said that the professional experience of the teachers in the study group is sufficient, and the preservice teachers' academic achievements are at medium and above the medium levels. In addition, three of the teachers in the study group are female and two are male; three of the preservice teachers are female and two are male.

2.3. Data Collection Tool

The teachers and the preservice teachers in each group carry out math classes in the same class at the same time as they work together. Therefore, the teachers and preservice teachers have grasp of the same process and the same students. Within the scope of the study, firstly, teachers and preservice teachers in each group are asked to prepare an achievement test in accordance with the topics discussed during their classes, in order to apply to the students. In this context, a total of 10 different achievement tests (5 of them are prepared by teachers and 5 of them are prepared by preservice teachers) are obtained from 5 groups and these tests are used as data collection tool. Teachers and preservice teachers are asked to apply the tests prepared separately to the same students on the same day and afterwards, they are asked to grade their own tests for each student.

2.4. Data Analysis

Following the application and grading of the tests, all test items and the students' grades received from these tests are examined during data analysis. In this context, the data collected by document analysis within the framework of the first research problem of the study, firstly item types and frequency of those types are found in the tests prepared by the teachers and preservice teachers, then the learning and sub-learning areas measured by the items are examined within the scope of the Ministry of National Education- MoNE (2018) mathematics lesson curriculum. Each researcher carried out this process separately and it is observed that they agreed on 99.2% of the items. Later, the researchers come together to discuss disagreements on the items and reached the final decision. Finally, the classification made by Smith and Stein (1998) is adopted in order to determine the CDTL of items and each researcher evaluated the test items separately according to that classification. As a result of this process, it is determined that the researchers made 98.9% compatible groupings in terms of CDTL of the items. The researchers came together for the items with disagreement and concluded the process by re-evaluating.

Within the scope of the second and third research problems of the study, the grades obtained by the students from the tests prepared by teachers and preservice teachers are examined. In this section,

where quantitative data are obtained, it is examined whether the data show a normal distribution in determining the statistical methods to be used. Descriptive statistics regarding the mentioned data are presented in Table 2.

Table 2. *Descriptive statistics about the test scores*

Group	Number of students took the test	Average score taken from teacher's test (over 100)	Average score taken from preservice teacher's test (over 100)	Standard deviation of teacher's test	Standard deviation of preservice teacher's test	p value for the normallity test of teacher's test	p value for the normallity test of preservice teacher's test
1	30	55.83	34.50	29.80	31.05	0.07	0.00*
2	46	53.15	35.96	21.01	22.77	0.40	0.17
3	28	91.82	69.96	11.28	19.80	0.00*	0.16
4	67	63.34	56.33	25.08	24.08	0.00*	0.06
5	67	63.34	48.78	25.08	19.79	0.00*	0.47

*$p<0.05$

When the table is examined, it is determined that the test scores for at least one test group (the test prepared by the teacher or the test prepared by the preservice teacher in the same group) in all groups except the second group do not show a normal distribution. Therefore, while parametric methods are conducted in all statistical analyzes for the second group, non-parametric analyzes are conducted in the analyzes performed for other groups.

In addition, the effect sizes of the significant differences between the means obtained as a result of the analyzes carried out within the framework of the second research problem are calculated. While the cohen d value is calculated as the effect size of the differences obtained from parametric analyzes (Cohen, 1998), the r value is calculated for the differences obtained from non-parametric analyzes (Fritz, Morris, & Richler, 2012).

2.5. Validity and Reliability

According to Fraenkel and Wallen (2006), validity defines the meaningfulness, appropriateness, and usefulness of the data. Expert opinion is used for the content validity and appearance validity of the data obtained in this study (Büyüköztürk, 2007). In addition to that, since the description of the study group and the process about obtaining the results of the study are the factors that increase validity, this process has been described in details. In addition, within the scope of reliability studies, researchers firstly grouped the data independently in the context of the related literature. Then, the consistency of the groupings made is examined and necessary arrangements are conducted together by researchers.

3. Results

Results obtained within the scope of the study are presented according to the research questions.

3.1. What Are the Properties of the Prepared Test Items?

In this context, the types and numbers of the test items prepared by teachers and preservice teachers, measured learning areas and sub-learning areas with those items, as well as the CDTL of the items are examined.

3.1.1. Results about item types and numbers. The data regarding the item types and numbers in the tests prepared by the teachers and preservice teachers are given in Table 3.

Table 3. *Item types and numbers in the tests*

Group	Teacher (T)/ Preservice Teacher (PT)	Number of items in the test	Item type /frequency	Frequency of item type
1	T1	21	multiple choice	17
			open ended	4
	PT1	12	multiple choice	4
			open ended	7
			matching	1
2	T2	20	multiple choice	20
	PT2	11	multiple choice	2
			open ended	9
3	T3	15	multiple choice	10
			open ended	5
	PT3	15	multiple choice	10
			open ended	5
4	T4	25	multiple choice	14
			open ended	8
			filling gap	3
	PT4	20	multiple choice	15
			open ended	5
5	T5	25	multiple choice	14
			open ended	8
			filling gap	3
	PT5	26	multiple choice	8
			open ended	3
			true/false	9
			filling gap	6

According to the data obtained and presented in Table 3, the findings related to the item types and numbers in the tests prepared by teachers and preservice teachers are analyzed in three categories. It is observed that the preservice teachers in the first category included more different types of questions than the teachers they worked with. For example, when the test prepared by T1 is examined, it is determined that there are 21 items in totally where 17 of these items are multiple choice, four are open-ended items. It is observed that PT1 who worked with T1 prepared a 12-item test with four multiple choice, seven open-ended and 1 matching type for the same students. According to these findings, it can be said that T1 and PT1 do not give much importance to the variety of item types. However, in the test prepared by PT1 by using fewer items than T1, more open ended item type is included. Similarly, while T2 prepared a 20-item test consisting of multiple choice items, PT2 working with T2 prepared an 11-item test, two multiple choice and nine open-ended items. Finally, T5 prepared a test with a total of 25 items, which are mostly multiple choice with14 items, eight of them are open-ended and three of them are filling the gap type items. On the other hand, PT5, who worked with T5, also included 26 items, that eight of them are multiple choice, three open-ended, nine right / wrong and six filling gap types, by emphasizing the variety of item types.

In the second category, there is a teacher and a preservice teacher whose item types and numbers are similar. In details, T3 and PT3 included 15 items, where 10 of them are multiple choice and five of them are open-ended items, in their tests. According to these findings, it is concluded that T3 and PT3 are not care about the variety of item types, and they prepared a test with a small number of items, mostly using the multiple choice item type.

When the test items prepared by T4 and PT4 in the last category are examined, it is observed that the teacher gave more importance to the variety of item types and number of items than the preservice teacher. T4 included 25 items, including 14 multiple choice, eight open-ended and three filling gap types. On the other hand, PT4 prepared a 20-item test with 15 multiple choice and five open-ended item type.

3.1.2. Results about the distribution of test items in learning and sub-learning areas. The data related to the distribution of the items in the tests prepared by teachers and preservice teachers in the

learning areas and sub-learning are presented in Table 4.

Table 4. *Distribution of the items according to the learning and sub-learning areas*

Group	Teacher (T)/ Preservice Teacher (PT)	Learning area/Sub-learning area	Frequency of objectives	Frequency of items
1	T1	Algebra /equity and equation	3	3
		Numbers and operations / rate and ratio	6	8
		Numbers and operations / percentages	4	4
		Geometry and measurement /lines and angles	2	5
		Not belong to any learning area	0	1
	T1	Numbers and operations / rate and ratio	4	4
		Numbers and operations / percentages	3	2
		Geometry and measurement /lines and angles	1	2
		Geometry and measurement /polygons	3	4
2	T2	Numbers and operations / rate and ratio	6	6
		Numbers and operations / percentages	4	4
		Algebra /Algebraic expressions	1	1
		Algebra / equity and equation	2	2
		Geometry and measurement /lines and angles	1	3
		Not belong to any learning area	0	4
	PT2	Numbers and operations / rate and ratio	4	4
		Numbers and operations / percentages	3	2
		Geometry and measurement /lines and angles	1	2
		Geometry and measurement /polygons	2	3
3	T3	Numbers and operations / rate and ratio	4	5
		Numbers and operations / percentages	3	4
		Algebra / equity and equation	1	2
		Geometry and measurement /lines and angles	2	4
	PT3	Numbers and operations / rate and ratio	3	5
		Numbers and operations / percentages	3	4
		Algebra / equity and equation	3	3
		Geometry and measurement /lines and angles	2	2
		Geometry and measurement /polygons	1	1
4	T4	Numbers and operations /operations in rational numbers	1	1
		Numbers and operations / rate and ratio	7	9
		Numbers and operations / percentages	3	4
		Algebra / equity and equation	3	3
		Geometry and measurement /lines and angles	1	6
		Not belong to any learning area	0	2
	PT4	Numbers and operations / rate and ratio	5	9
		Numbers and operations / percentages	4	5
		Geometry and measurement /lines and angles	1	2
		Not belong to any learning area	0	4
5	T5	Numbers and operations /operations in rational numbers	1	1
		Numbers and operations / rate and ratio	7	9
		Numbers and operations / percentages	3	4
		Algebra / equity and equation	3	3
		Geometry and measurement /lines and angles	1	6
		Not belong to any learning area	0	2
	PT5	Numbers and operations / rate and ratio	6	9
		Numbers and operations / percentages	3	7
		Algebra / equity and equation	1	1
		Geometry and measurement /lines and angles	2	3
		Not belong to any learning area	0	6

Findings obtained according to Table 4 are again examined in three categories. It is determined that the items in the first category, T1 and PT1, measured the gains in different learning and sub-learning areas. In other words, T1 and PT1, who worked in the same class during the same period, are focused on measuring the success of students in that class in different sub-learning areas. For example, while T1 measures three objectives with three items in the equity and equation sub-learning area of algebra learning area, PT1 does not include any items and objectives from this sub-learning area. On the other hand, while PT1 uses four items for three objectives in the polygons sub-learning area of geometry and measurement learning area, it is determined that T1 does not include this sub-learning area in her test. Besides, it is determined that both T1 and PT1 uses different number of items with different objectives related to the lines and angles sub-learning area of geometry and measurement learning area, the rate and ratio sub- learning area of numbers and operations, and the percentages sub-learning area of same learning area in their tests. In addition, when the curriculum is examined, it is determined that T1 deals with the majority of the objectives in the mentioned sub-learning areas, but PT1 covers the objectives in a narrower scope in their tests. In addition, one of the items prepared by T1 does not address any objective in the curriculum. Another example that can be included in this category is T2 and PT2. The test prepared by T2 includes two sub-learning areas in the algebra learning area by considering half of the objectives in the curriculum, whereas PT2 does not include any objective related to the algebra learning area in her test. On the other hand, while the polygons sub-learning area of measurement learning area is included with two objectives out of five by PT2, T2 does not include the polygons sub-learning area in his/her learning area. In addition to that, it is determined that the 4 items included in T2' test do not measure any objective in the curriculum. Finally, T2 and PT2 use one objective out of the two in the lines and angles sub-learning area of geometry and measurement learning area. It is observed that the objectives in rate and ratio sub-learning area and percentages sub-learning area of numbers and operations learning area are included in T2's test more comprehensive than PT2's test.

In the second category, it is observed that the preservice teacher deals with the objectives and sub-learning areas more comprehensively than the teacher. When the tests prepared by T3 and PT3 in this category are analyzed, it is determined that the extra geometry and measurement learning area of PT3 includes one objective out of five in the polygons sub-learning area. Similarly, while T3 measures one out of four objectives in the equity and equation sub-learning area of algebra learning area with two items, PT3 measures three objectives of the same sub-dimension with three items. On the other hand, it is determined that both T3 and PT3 use similar number of objectives and items in their tests for the lines and angles sub-learning area of geometry and measurement learning area and for the percentages sub-learning rate and the ratio sub-learning area of the numbers and operations learning area.

In the last category, there are T4, PT4, T5 and PT5, where teachers' tests include more sub-learning areas than preservice teachers' tests. For example, when T4 measures the operations on rational numbers sub-learning of the numbers and operations learning area with one item for one objective out of five objectives given in the curriculum PT4, does not address this sub-learning area in her test. Similarly, while T4 includes three items for three objectives from the equity and equation sub-learning area of the algebra learning area, which is included in the curriculum with four objectives, PT4 does not include this sub-learning area in her test. On the other hand, both T4 and PT4 include the objectives for the lines and angles sub-learning area of geometry and measurement learning area, rate and ratio sub-learning area with the percentages sub-learning area of the numbers and operations learning area with small differences in their tests. Finally, it is determined that three items prepared by T4 and four items prepared by PT4 do not measure any objectives in the relevant curriculum. Similarly, when the tests prepared by T5 and PT5 are examined, it is observed that operations on rational numbers sub-learning area of the numbers and operations learning area that is presented with five objectives in the curriculum, T5 include one item for one objective, where PT5 do not include any objective related to that sub-learning area. Moreover, it is determined that two items prepared by T5 and six items prepared by PT5 do not measure any objectives in the relevant curriculum. Apart from that, it is determined that T5 and PT5 deal with many objectives with similar number of items in the rate and ratio sub-learning area and the percentages sub-learning area of numbers and operations learning area, the equity and equation sub-learning area of the algebra learning area, and the lines and angles sub-learning area of geometry and measurement learning area in their tests.

3.1.3. Results about the CDTL of items. The data regarding the distribution of the items in the tests prepared by teachers and preservice teachers according to CDTL are presented in Table 5.

Table 5. *Distribution of test items according to CDTL*

Group	Teacher (T)/ Preservice Teacher (PT)	Number of items in the test	Level of CDTL	Number of items in CDTL	Percentage of items in CDTL
1	T1	21	level 1- memorization	0	0
			level 2- procedures without connections	8	38.09
			level 3- procedures with connections	13	61.90
			level 4- doing mathematics	0	0
	PT1	12	level 1- memorization	1	8.33
			level 2- procedures without connections	1	8.33
			level 3- procedures with connections	10	83.33
			level 4- doing mathematics	0	0
2	T2	20	level 1- memorization	2	10
			level 2- procedures without connections	10	50
			level 3- procedures with connections	8	40
			level 4- doing mathematics	0	0
	PT2	11	level 1- memorization	1	9.09
			level 2- procedures without connections	4	36.36
			level 3- procedures with connections	6	54.55
			level 4- doing mathematics	0	0
3	T3	15	level 1- memorization	0	0
			level 2- procedures without connections	6	40
			level 3- procedures with connections	9	60
			level 4- doing mathematics	0	0
	PT3	15	level 1- memorization	0	0
			level 2- procedures without connections	4	26.67
			level 3- procedures with connections	11	73.33
			level 4- doing mathematics	0	0
4	T4	25	level 1- memorization	4	16
			level 2- procedures without connections	10	40
			level 3- procedures with connections	11	44
			level 4- doing mathematics	0	0
	PT4	20	level 1- memorization	4	20
			level 2- procedures without connections	9	45
			level 3- procedures with connections	5	25
			level 4- doing mathematics	2	10
5	T5	25	level 1- memorization	4	16
			level 2- procedures without connections	10	40
			level 3- procedures with connections	11	44
			level 4- doing mathematics	0	0
	PT5	26	level 1- memorization	7	26.92
			level 2- procedures without connections	5	19.23
			level 3- procedures with connections	14	53.85
			level 4- doing mathematics	0	0

When the data presented in Table 5 are examined, the CDTL of the items used by the teachers and preservice teachers in the tests are analyzed in three categories. The first category includes teachers and preservice teachers in the 1st, 2nd and 3rd groups. In this category, preservice teachers prepare items with higher CDTL than the teachers they worked together. For example, while 61% of the items prepared by T1 are at the third level, that is procedures with connections, and 39 % of them are at the second level, procedures without connections. On the hand 83% of the items prepared by PT1 are at the third level, which is procedures with connections. Similarly, while T2 prepared his test as 40% from the third level items and 50% from the second level, it is determined that the PT2 prepared 55% from the third level and 36% from the second level. Lastly, it seen that 60% of the test items prepared

by T3 are at the third level and 40% of them are at the second level where PT3 prepares the items as 73% of them at third level and 27% of them at second level.

In the second category, the teacher prepares items with higher CDTL than the preservice teachers. For instance, it is observed that T4, which is in this category, prepares items as 44% of them are at third level 40% of them are at the second level, where 16% of them at the first level which is the lowest one. On the other hand, in the test prepared by PT4, 20% of the items at the first level, 45% of them are at the second level and 25% of them are at the third level.

It is determined that the CDTL of the test items prepared by T5 and PT5 in the last category differ. For example, it is determined that T5 prepares 16% of the items at the lowest level, of 40 %of them are at the second level and 44% of them are at third level. On the other hand, it is seen that 27% of the items prepared by PT5 are at the first level, 19% of them are at the second level and 54% of them are at the third level.

In general, it is determined that the CDTL of the items that preservice teachers used are at the second and third level and only one pre-service teacher (PT4) prepared a question at the highest level of the CDTL.

3.2. Is There Any Difference Between the Students' Grades Obtained from The Teachers' and Preservice Teachers' Tests?

Another sub-problem examined within the scope of the study is whether there is a significant difference between the grades of the students obtained from the teachers' tests and preservice teachers' tests. The results of the analysis carried out in this context are presented in Table 6.

Table 6. *Comparison of the students' grades from the tests*

Group	Number of student took the tests	Mean of teacher's test	Mean of preservice teacher's test	t	z	Mean rank (for z value)	Sum of ranks (for z value)	p	Effect size
1	30	55.83	34.50	-	-4.63*	1.50	3.00	0.00**	-0.84
2	46	53.15	35.96	7.18	-	-	-	0.00**	1.05
3	28	91.82	69.96	-	-4.65*	0	0	0.00**	-0.87
4	67	63.34	56.33	-	-3.76*	26.13	496.50	0.00**	-0.45
5	67	63.34	48.78	-	-5.70*	20.55	226.00	0.00**	-0.69

*based on positive ranks **p<0.05

When the table above is analyzed, it is seen that the students in all groups have a lower average in the tests prepared by preservice teachers than the tests prepared by teachers. In the analyzes performed to determine whether these differences are significant, it is seen that the differences are significant for group 1 ($z = -4.63$; $p = 0.00 < 0.05$); group 2 ($t = 1.86$; $p = 0.00 < 0.05$) group 3 ($z = -4.65$; $p = 0.00 < 0.05$), group 4 ($z = -3.76$; $p = 0.00 < 0.05$) and group 5 ($z = -5.70$; $p = 0.00 < 0.05$). In other words, students in all groups have been more successful in the tests prepared by teachers than the tests prepared by preservice teachers. In addition, the effect size of these statistical differences is calculated. It has been determined that the effect size in group 4 is at medium level and at high level in other groups (Coolican, 2009).

3.3. Is There a Relationship Between the Grades That the Students Obtained From the Tests Prepared by Teachers and Preservice Teachers?

Finally, within the scope of this study, it is examined whether there is a relationship between the grades of the students obtained from the tests prepared by teachers and preservice teachers. Related analysis results are presented in Table 7.

Table 7. *The relationship between students' grades from the tests*

Group	Number of student took the tests	Mean of teacher's test	Mean of preservice teacher's test	Correlation value	p value
1	30	83.77	88.37	0.69	0.00*
2	30	55.83	34.50	0.86	0.00*
3	46	53.15	35.96	0.73	0.00*
4	28	91.82	69.96	0.63	0.00*
5	67	63.34	56.33	0.85	0.00*
6	23	64.13	65.74	0.86	0.00*
7	32	63.75	67.91	0.85	0.00*
8	67	63.34	48.78	0.78	0.00*

*$p<0.01$

4. Discussion and Conclusions

4.1. Conclusion and Discussion About Item Types and Numbers

Three different results are reached by examining the tests prepared by teachers and preservice teachers according to the number and type variety of items. Firstly, the variety of item types used by preservice teachers in tests is better than the teachers they work with. Second result is that the number and variety of items used by teachers and preservice teachers in the tests they prepared are similar. Lastly, the teacher gives more place to the item variety and number than the preservice teacher. According to the results of this study, generally the types of test items prepared by teachers and preservice teachers are multiple choice, open ended, filling gap, matching and true / false. In the tests prepared by mathematics teachers and preservice teachers, it is seen that the most used item type is multiple choice. The variety of item types used by teachers and preservice teachers in test is important. Since understanding the characteristics of different item types f is an important step for the correct interpretation and use of assessment results (Brassil & Couch, 2019). However, in the literature, there are results regarding the different item types that measure different characteristics of student learning. For example, Ozuru, Briner, Kurby and McNamara (2013) express that performance in open-ended questions is related to the quality of students' self-explanations, but performance in multiple-choice questions is related to the previous level of knowledge of students. These results show that open-ended and multiple choice questions measure different aspects of comprehension processes. Also, in their study, Chen, Gotwals, Anderson and Reckase (2016) found that a combination of different item types can measure all aspects of the structure, that is, item types in different formats can evaluate slightly different aspects of learning progress. A combination of different formats can provide more comprehensive information about students' abilities, given that multiple choice questions focus on students' ability to determine the best choice, but open-ended questions assess students' ability to organize and synthesize their knowledge to solve problems (Chen, Gotwals, Anderson, & Reckase, 2016; Yao & Boughton, 2009). Considering that students prefer multiple choice test items based on memorization rather than open-ended exams (Ben-Chaim, & Zoller, 1997), it is understandable that teacher and preservice teachers prefer multiple choice item types. Similarly, Birgin (2007) find that most of the preservice teachers tend to use tests with multiple choice items, short answer items, and filling gap item types. In addition to that, although the teachers and preservice teachers use different types of items to measure the different thinking structures of their students, the multiple choice items which encourage students to answer questions through a limited timeframe without thinking deeply are used too often (Krulick, Rudnick, & Milou, 2003; Romberg & Lange, 2005; Sheffield & Cruikshank, 2000) and all these results are compatible with the findings of this study.

4.2. Conclusion and Discussion About Learning Area of Items

Again, three different results are obtained within the scope of the study regarding the learning areas included in the tests prepared by teachers and preservice teachers. Firstly, it is found that the teacher and preservice teacher considered different sub-learning areas in the tests they prepared. For example, T1 and PT1, who work in the same class during the same period, focused on measuring the success of students in that class with different sub-learning areas. Secondly, it is concluded that the preservice

teachers examined the objectives and sub-learning areas more comprehensively in the tests they prepared. Thirdly, the fact that some teachers included more sub-learning areas than preservice teachers is among the results of the study. Another result of this research is that four teachers and two preservice teachers include test items related to objectives that are not valid for the related grade level. These results may be interpreted that preservice teachers and especially teachers do not consider the content validity in the tests they prepared. Similar to this situation, Quansah, Amoako and Ankomah (2019) examine the tests prepared by mathematics teachers and state that teachers cannot prepare a test that includes all of their learning areas. It cannot be expected to reflect the skills and learning outcomes for an assessment that is not valid, since it cannot possibly measure the content of learning areas in the curriculum (Nitko, 2001). Because if the objectives in the learning areas at the related grade level are not included in the tests prepared, it reduces the validity of the assessment. The aim of teachers to include the items related to the objectives in the previous classes may be an effort to prepare test with easier items where they see that the student level is low in their classes

4.3. Conclusion and Discussion About CDTL

According to the evaluations of the questions prepared by the teachers and preservice teachers according to their cognitive demand levels, three basic results are obtained. First of them is that preservice teachers prepare higher CDTL items than teachers they worked together. The second is that teachers prepare items at a higher CTDL than preservice teachers. Finally, it is the situation that the test items prepared by the teacher and the preservice teacher that they work together do not show a pattern at BIS levels of test items. In general, the test items prepared by teachers and preservice teachers are at level 2 and level 3, and only one preservice teacher (PT4) prepared a test item at the highest level (level4). In previous studies, it has been observed that mathematics teachers prepare low cognitive level items in their tests (Güler, Özdemir, & Dikici, 2012; Köğce & Baki, 2009; Marso & Pigge, 1991; Quansah, Amoako, & Ankomah, 2019). In another study, the CDTL levels of the questions used in the teaching environment are examined and it is determined that the teachers asked questions at the level of level 2 and level 3 generally (Ubuz & Sarpkaya, 2014). Similarly, in this research also, it is seen that the items are prepared mainly at the level level2 and level 3. However, Quansah, Amoako and Ankomah (2019), express that 30% of the test items must be at low cognitive level and 70% at high cognitive level, and the items prepared by mathematics teachers who participated in their research do not meet this criterion.The National Council of Teachers of Mathematics (NCTM, 2000) in the United States proposes the use of questions covering cognitive processes so that the results can be solved in different ways, that is, students who have more than one solution can use their higher level thinking skills. Therefore, it may be thought that the items requiring higher CDTL should be included in the tests in classes. In addition, in their study, Köğce and Baki (2009) examined the cognitive levels of the items asked by mathematics teachers when assessing their students according to Bloom's taxonomy and as a result, they find that teachers prefer the items at the level of understanding and application. Since understanding and application level items may be considered as level 2, level 3 according to CDTL. Therefore, the results of this study carried out are similar to the results of Köğce and Baki (2009).

On the other hand, Bal (2012) conclude that preservice teachers prefer the use of tests that will reveal their cognitive levels. Although preservice teachers want to be evaluated with tests that will reveal cognitive processes for themselves, it is determined that they do not pay attention to cognitive thinking levels in the tests they will prepare for their students. It is thought that this situation arises because the preservice teachers do not have sufficient knowledge in preparing a test.

4.4. Conclusion and Discussion About the Difference and Relationship Between the Grades of the Students Obtained from the Tests

According to the findings obtained within the scope of this study, the students are more successful in the test prepared by teachers than the tests prepared by the preservice teachers. In his study, Mertler (2003) investigate the assessment literacy differences between teachers and preservice teachers and mentioned that the preservice teachers got the lowest grade from the sub-competency to prepare valid grading procedures. One result of this study conducted, that the students got higher grades from the

tests prepared by the teachers, may have been caused by the lack of preservice teachers' grading procedures like in the study Mertler (2003) expressed.

In another study by Cizek, Fitzgerald and Richard (1996), teachers are examined in terms of their experiences and found that there is no difference in their assessment and evaluation knowledge. Considering that teachers are more experienced than preservice teachers this situation is not parallel to the results of this study. In addition, students getting lower grades from the tests prepared by preservice teachers may be due to the result of the teachers asking test items about the objectives at lower grade levels. In teachers' tests, the inclusion of objectives from previous grade level's learning areas may have lowered the level of items so that students may have higher grades than the tests prepared by preservice teachers.

Finally, as a result of the study, it is obtained from the tests prepared by the teachers and preservice teachers that there is a significant and positive relationship between the grades of the students. This result indicates that as the grades obtained from the tests prepared by the preservice teachers increase, the grades obtained by the tests prepared by the teachers increase. Therefore, it can be interpreted that the tests prepared by the teachers and preservice teachers can measure the same features of students.

5. Suggestions

In this study, in which the tests prepared by teachers and preservice teachers to examine students' learning are examined, it is observed that teacher and preservice teachers mostly used multiple choice item type. Also, it is among the results that teachers generally include items from previous grade level objectives. When analyzed the items in terms of CDTL, it is found that teachers and preservice teachers generally prepare items at level 2 and level 3. This shows the relationship between the tests prepared by teachers and preservice teachers who they worked together have similar characteristics. According to the results found in this study, the following suggestions can be given.

It is important for teachers to be able to choose, develop and implement appropriate tests for their classes by realizing the objectives they want to achieve with their assessment and evaluation activities (Stanford & Reeves, 2005). For this reason, it can be suggested that mathematics teachers and preservice teachers may use table of specification in order not to ignore the objectives stated in the curriculum. Also, in their lessons, they should be encouraged to use questions that require higher thinking, such as doing mathematics, as well as questions that are based on memorization, that requiring lower-level thinking. Such questions should be helped to raise awareness of students about improving their cognitive competence. awareness can be increased through in-service training courses for teachers and practical studies in preservice teachers for assessment and evaluation courses.

References

Ahmed, A., & Pollitt, A. (2007). Improving the quality of contextualized questions: An experimental investigation of focus. Assessment in Education: *Principles, Policy & Practice, 14*(2), 201–232. https://www.tandfonline.com/doi/full/10.1080/09695940701478909?casa_token=WLeoNp0xx18AAAAA%3AxtOc3A_C_y1xesG8jNX8yX0kTaLOXvPknXG57VblY9vzIxuffBX74iGrgGcXAsQF5fEIIE1XVgvS8gg

Arbaugh, F., & Brown, C. (2005). Analyzing mathematical task: A catalyst for change? *Journal of Mathematics Teacher Education, 8*(6), 499-536.

Bal, A. P. (2012). Assessment preferences of pre-service teachers about mathematics course. *The Journal of Turkish Educational Sciences, 10*(3), 477-479. http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=6f56545d-3add-409a-ab00-27a9fd5fc104%40sdc-v-sessmgr01

Ben-Chaim, D., & Zoller, U. (1997). Examination-type preferences of secondary school students and their teachers in the science disciplines. *Instructional* Science, *25* (5), 347-367. https://link.springer.com/article/10.1023/A:1002919422429

Birgin O. (2007). Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki okur-yazarlık düzeylerinin incelenmesi [Examining the literacy levels of classroom teacher candidates on measurement and evaluation] E. Erginer (Ed.), XVI. Ulusal Eğitim Kongresi (Cilt 3, 498-503). Ankara: Detay Yayıncılık.

Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education, 6*(1), 1-17

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2010.00195.x?casa_token=tPu2vPAWrloAAAAA:hLcroD1umFizQVcSiyeJ3LhXwDCkN00kyUP4aKyL2YfpyVzfR6tzj1KfDsgKFDZy8lvdHEvtXlhI-IO6

Büyüköztürk, Ş. (2007). *Sosyal Bilimler İçin Veri Analizi El Kitabı* [Manual of data analysis for social sciences], Ankara, Pegem A Yayıncılık.

Chen, J., Gotwals, A. W., Anderson, C., & Reckase, M. (2016). The influence of item formats when locating a student on a learning progression in science. *International Journal of Assessment Tools in Education, 3*(2), 101-122. https://dergipark.org.tr/en/pub/ijate/article/245196

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2011). *Research methods, design, and analysis* (11th ed.). Boston, MA: Pearson.

Cizek, G. J., Fitzgerald, S. M. & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179. https://www.tandfonline.com/doi/abs/10.1207/s15326977ea0302_3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coolican, H. (2009). *Research methods and statistics in psychology.* London, United Kingdom: Hodder.

Çalık, B., & Aksu, M. (2018). A systematic review of teachers' questioning in Turkey between 2000-2018. *Elementary Education Online, 17*(3), 1548-1565. http://ilkogretim-online.org.tr/index.php/io/article/view/2801

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator, 48*(2), 128-142. https://www.tandfonline.com/doi/full/10.1080/08878730.2012.760024?casa_token=WnyHdszb9GYAAAAA%3ADKK6nT7VjJ1a3WYxwWhEF61Zh_EoHXXNncKty19f3-t7lnzL16zc0HuE68sEIdSOry09KpJVVQTu4kc

DiDonato-Barnes, N., Fives, H., & Krause, E. S. (2014). Using a table of specifications to improve teacher-constructed traditional tests: An experimental design. *Assessment in Education: Principles, Policy & Practice, 21*(1), 90-108. https://www.tandfonline.com/doi/full/10.1080/0969594X.2013.808173?casa_token=QXn9kOvwQowAAAAA%3Aj1UFRbXP1ZHezjcvf5LwZEY1C5BgjIjH0aQYHuiHi1y1xeHeVnnafFhuwVSEL-7B-SUPkx-hKD-J5Is

Güler, G., Özdemir, E., & Dikici, R. (2012). A comparative analysis of elementary mathematics teachers' examination questions and SBS mathematics questions according to Bloom's taxonomy. *Erzincan University Journal of Education Faculty, 14*(1), 41-60.

Fraenkel, J. R. & Wallen, N. E. (2006). *How to design and evaluate research in education.* New York: McGraw Hill.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fa0024338

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for research in mathematics education, 28*(5), 524-549.

Jones, D.L., & Tarr J. E. (2007). An examination of the levels of cognitive demand required by probability tasks in middle grades mathematics textbooks. *Statistics education research journal, 6*(2), 4-27. https://d1wqtxts1xzle7.cloudfront.net/46485826/SERJ62_Jones_Tarr.pdf?1465934196=&response-content-disposition=inline%3B+filename%3DAn_Examination_of_the_Levels_of_Cogitive.pdf&Expires=1613832153&Signature=FNmVfF6sC74OFG1wmbFjEIs0zv9HW2zLerjNl29vqClpOAVlmWccnT4hF~1V~PZE94hFvyvM4UXx3H5O8XftkXGUjJand3rSsS25hCqw2yuY8iFeggoFJPDF3T9KvlDaG~tNFN4wq2VmZqZmbwsIGKxlVc7XBnlyzxuEwcgxB26OKAEBZPMHklZHI2BtNlV~3~bOwu5xf3YnmrRO-4tiDuRTOMWR1Dj-XTaUQ5lJHywnGVQaRiUOZov3CSRF9U0uCK~Mpxx~qrlDqqKR~Oa-yzEXHpPGPiriCGHTzlMrKhMrsqgRZ7wXdE3uHA5mZIEBy~1j7ntLqDusqeZeNDvadg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Köğce, D., & Baki, A. (2009). A Comparision of high-school mathematics teachers' examination questions and mathematics questions in the university entrance exams according to bloom's taxonomy. *Pamukkale University Journal Of Education, 26*(26), 70-80.

Krulick, S., Rudnick, J., & Milou, E. (2003). *Teaching Mathematics in the Middle School.* Newyork: Pearson Education.

Little, M.E. (2009). Teaching mathematics: Issues and solutions. *Teaching Exceptional Children Plus, 6*(1), 1-15. https://eric.ed.gov/?id=EJ875420

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education, 20*(5), 525–535. https://www.sciencedirect.com/science/article/pii/S0742051X04000575?casa_token=ZJHVPBljY94AAAAA:lRO_O8ZOYx_MVum4-20cI94YSEd7LuWnzgCIsuWP1YpslXWOdPMShMrPyxM0HYFGm23EtEXbb5w

Marso, R. N., & Pigge, F. L. (1991). An analysis of teacher-made tests: Item types, cognitive demands, and item construction errors. *Contemporary Educational Psychology, 16*(3), 279–286. https://www.sciencedirect.com/science/article/abs/pii/0361476X9190027I

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education, 120*(2), 285-296. https://go.gale.com/ps/anonymous?id=GALE%7CA59644154&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00131172&p=AONE&sw=w

Mertler, C. A. (2003). Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference? Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, Ohio.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 31*(1), 49–64. https://www.jstor.org/stable/41064623?casa_token=HzUggcreTN4AAAAA%3AHNVReMrTxc4nA_yyKUb_sHa76h2noiadKVx1ndpS5ZdYT9h9lCxMG0hiRhKQhxiVj3onVODSPz73B_HCGsKcf5yY7c_O9MLYHxUK6mk2ZCFT7ZVKkWd7&seq=1#metadata_info_tab_contents

MoNE (2018). *Mathematics Course Curriculum (1st-8th Grades)*. Ankara: MEB.

National Council for Teachers of Mathematics [NCTM] (1989), *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM

National Council for Teachers of Mathematics [NCTM] (2000). *Principles and Standards for School Mathematics.* Reston, VA: NCTM

Nitko, J. A. (2001). *Educational Assessment of Students*. New Jersey: Prentice Hall.

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67*(3), 215.

Özcan, B.N., & Delil, A. (2018). An analysis of the tests prepared by elementary mathematics teachers in terms of curriculum achievements. *Kastamonu Education Journal*, 26(6), 1909-1917.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6*(1), 21-27.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10–12. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.1993.tb00548.x

Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership, 62,* 82-83. https://eric.ed.gov/?id=EJ716762

Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, *6*(1), 1-8. https://dergipark.org.tr/en/pub/ijate/issue/40373/481164

Romberg, T. A., & Lange, J. (2005). Research in Assessment Practices. In T. A. Romberg, T. P. Carpenter ve F. Dremock (Ed.), Understanding Mathematics and Science Matters (279-307). Mahway, New Jersey: Lawrence Erlbaum Assosiates Publishers.

Sheffield, L. J., & Cruikshank, D. E. (2000). *Teaching and learning elementary and middle school Mathematics.* New York: John Wiley & Sons.

Silver, E. A., & Stein, M. K. (1996). The QUASAR Project: The "Revolution of the Possible" in Mathematics Instructional Reform in Urban Middle Schools. *Urban Education, 30*(4), 476–521. https://journals.sagepub.com/doi/abs/10.1177/0042085996030004006?casa_token=nKg__CtPKF4AAAAA:Yyfp3ey2FnVPL5jHnWGpiX3E99mFoQjTpcK-geHANkeuYfjsXKtk92CGWzzQ8k9OgoPZDpgcble9jQ

Smith, M. S., & Stein, M. K. (1998). Reflections on Practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3(*5), 344–350.

Stanford, P., & Reeves, S. (2005). Assessment that drives instruction. *Teaching Exceptional Children, 37*(4), 18-22. https://journals.sagepub.com/doi/pdf/10.1177/004005990503700402?casa_token=C1S2PffXzUoAAAAA:GjiKkGGDT5Wxn4oW-zT2Wztv5fxGmjxfZe94_yV6aI-a1KsFJO3LTBFNAMetTS4AYZJ3iK7pJVcgmw

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American educational research journal, 33*(2), 455-488. https://journals.sagepub.com/doi/abs/10.3102/00028312033002455?casa_token=8xX93-rf828AAAAA:d14Y8TGkoVNyq-8nfdBmNIlyhfxzzoGKtW_JY6Eeu_PO44AiYpkzS1nhqw8yXt1_7DOnhNIQz8QR-w

Stein, M.K., & Smith, M.S. (1998). Mathematical task as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School, 3*(4), 268-277.

Stein, M. K., Smith, M.S., Henningsen, M., & Silver, E.A. (2000). *Implementing standarts based mathematics instruction: A casebook for Professional development*. New York: Teachers College Press. https://pdfs.semanticscholar.org/deac/6a43706ea11bdd9fb07a73b54a08f0c114fb.pdf

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*(7), 534-539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238-245. https://search.proquest.com/openview/420d00f4b01136f8d23c90af1e8936fd/1/advanced

Stiggins, R. J. (1999). Are you assessment literate? *The High School Journal, 6*(5), 20-23. https://eric.ed.gov/?id=EJ585543

Ubuz, B., Erbaş, A.K., Çetinkaya,B., & Özgeldi, M.(2010). Exploring the Quality of the Mathematical Tasks in the New Turkish Elementary School Mathematics Curriculum Guidebook: The Case of Algebra. *ZDM Mathematics Education, 42,* 483-491.

Ubuz, B., & Sarpkaya, G. (2014). The investigation of algebraic tasks in sixth grades in terms of cognitive demands: mathematics texbook and classroom implementations. *Elementary Education, 13*(2), 594-606. http://ilkogretim-online.org/index.php/ioo/article/view/850

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. https://eric.ed.gov/?id=EJ780818

Yao, L., & Boughton, K.A. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement, 46* (2), 177–197. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3984.2009.00076.x?casa_token=e3s5CDxCOgUAAAAA:ULKSWOf3oBzInUYMd4sm-aQWnrsHOXRVo7M1rVV8BOcCH67Cbs17OQEtc6w4v4wZw_LH3RXPdLxg5Orc

Authors

Feride ÖZYILDIRIM-GÜMÜŞ, *corresponding author*, Department of Elementary Mathematics Education, Aksaray University, Aksaray, Turkey. E-mail: ferideozyildirimgumus@gmail.com

Gülfem SARPKAYA-AKTAŞ, Department of Elementary Mathematics Education, Çukurova University, Adana, Turkey. E-mail: gulfemsarpkaya@yahoo.com

Hilmi KARACA, Department of Elementary Mathematics Education, Aksaray University, Aksaray, Turkey. E-mail: hkrc47@gmail.com