

An Investigation of the Effect of a Teaching Experiment on the Assessment Literacy of Pre-service Mathematics Teachers According to Teacher Competence Standards

Gülnur Özdil^a and Ahmet Kaçar^b

^aKastamonu University, Institute of Science and Technology, Kastamonu, Turkey (ORCID: 0000-0002-6355-2027)

^bKastamonu University, Faculty of Education, Kastamonu, Turkey (ORCID: 0000-0002-0072-2033)

Article History: Received: 31 March 2023; Accepted: 22 June 2023; Published online: 31 August 2023

Abstract: Literacy is a concept that contains the functional use of knowledge and skills in a specific field. In this sense, teachers who are assessment literate are expected to know assessment methods, choose and develop the appropriate tools, evaluate results and give the feedbacks. In Turkey, teacher training programs include assessment courses. It is intended in this study to examine the assessment literacy of pre-service mathematics teachers who took this course before and after a teaching experiment conducted within the scope of assessment in mathematics education according to the Standards for Teacher Competence in Educational Assessment of Student (AFT, NCME & NEA, 1990) and to reveal the changes that occurred. The participants are eight pre-service teachers in the last year of the mathematics teaching undergraduate program in a state university in Turkey. The research was carried out with a qualitative approach and descriptive analysis was used. The research data were collected in two stages using Classroom Assessment Literacy Inventory in Mathematics Education consisting of open-ended questions. According to the results obtained, it is revealed that the pre-service teachers consider assessment professionally important, but they do not feel sufficient in this field. In addition, it was observed that pre-service teachers did not have sufficient knowledge about alternative assessment approaches. The results of the research show that the participants have misconceptions about the main concepts such as validity and reliability.

Keywords: Assessment literacy, Pre-service teacher education, Classroom assessment, Pre-service mathematics teacher

Öz: Okuryazarlık, belirli bir alandaki bilgi ve becerilerin işlevsel kullanımını içeren bir kavramdır. Bu anlamda değerlendirme okuryazarı olan öğretmenlerden değerlendirme yöntemlerini bilmeleri, uygun araçları seçip geliştirmeleri, sonuçları değerlendirmeleri ve geri bildirimde bulunmaları beklenmektedir. Türkiye’de öğretmen yetiştirme programları eğitimde ölçme değerlendirme derslerini içermektedir. Bu çalışmada, bu dersi alan matematik öğretmen adaylarının değerlendirme okuryazarlığının matematik eğitiminde ölçme değerlendirme konusu kapsamında gerçekleştirilen bir öğretim deneyi öncesi ve sonrasında Öğrencinin Eğitimsel Değerlendirmesinde Öğretmen Yeterliklerine Yönelik Standartlara (AFT, NCME & NEA, 1990) göre incelenmesi ve meydana gelen değişimlerin ortaya konulması amaçlanmıştır. Katılımcılar, Türkiye’de bir devlet üniversitesinde matematik öğretmenliği lisans programının son sınıfında öğrenim gören sekiz öğretmen adaydır. Araştırma nitel bir yaklaşımla gerçekleştirilmiş ve betimsel analiz kullanılmıştır. Araştırma verileri açık uçlu sorulardan oluşan Sınıfta Değerlendirme Okuryazarlığı Envanteri ile iki aşamada toplanmıştır. Elde edilen sonuçlara göre öğretmen adaylarının değerlendirmeyi mesleki açıdan önemli gördükleri ancak bu alanda kendilerini yeterli hissetmedikleri ortaya çıkmıştır. Ayrıca öğretmen adaylarının alternatif değerlendirme yaklaşımları hakkında yeterli bilgiye sahip olmadıkları görülmüştür. Araştırma sonuçları, katılımcıların geçerlik ve güvenirlik gibi temel kavramlar konusunda kavram yanlışlarına sahip olduğunu göstermektedir.

Anahtar Kelimeler: Ölçme değerlendirme okuryazarlığı, Öğretmen yetiştirme eğitimi, Sınıf içi ölçme değerlendirme, Matematik öğretmeni adayı

[Türkçe sürüm için tıklayınız](#)

1. Introduction

Teachers are the people who check first-hand whether the teaching carry out in accordance with the aims and make decisions about teaching and examinations (Yantim & Wongwanich, 2014). One of the most principal tasks of teachers is to evaluate student performance (Fan, Wang & Wang, 2011). The knowledge and skills of teachers on educational assessment are expressed with the concept of assessment literacy (Alkharusi, 2011). Stiggins (1995) described assessment literates as ‘they know the difference between sound and unsound assessment: knowing what they are assessing, why they are doing it, how to assess the skill/knowledge of interest, how to generate sound examples of student performance, what can potentially go wrong with the assessment, and how to prevent those problems before occurring’.

Corresponding Author: Gülnur Özdil  **email:** gulnurozdil@outlook.com

Citation Information: Özdil, G. & Kaçar, A. (2023). An investigation of the effect of a teaching experiment on the assessment literacy of pre-service mathematics teachers according to teacher competence standards. *Turkish Journal of Mathematics Education*, 4(2), 27-58.

Individual learning differences are an indication that a one-size-fits-all assessment model is not possible or feasible (Lee & Son, 2015). Considering these differences, the selection and application of methods that can help teaching and detect learning deficiencies depends on teachers' expertise in this subject. Teachers with assessment capacity can improve students' learning, reasoning and beliefs. For this reason, assessment knowledge skills are seen as an important professional competence (Fan, Wang & Wang, 2011).

In recent years, teachers' assessment expertise has gotten more importance (Mertler, 2004). In studies on teachers' assessment knowledge and skills, most of the teachers had classroom assessment literacy at the low level (Yamtim & Wongwanich, 2014; Ergül, 2019). Also, teachers stated that they could not get adequate assessment education in undergraduate or in-service training (Gelbal & Kelecioğlu, 2007; Kilmen & Çıkırıkcı Demirtaşlı, 2009).

In Canada, the United Kingdom, and the United States, it is seen that the emphasis given to transparent assessment practices in the classroom has increased as well as the increasing large-scale assessment within the scope of public accountability. Teachers' competency in the field of student assessment and evaluation has become a necessity. Therefore, pre-service teacher education should be revised for a qualified and confident beginning to profession. Given the extensive significance of both large-scale and classroom assessment, assessment literacy should be an explicit component in pre-service teacher education programs (DeLuca & Klinger, 2010).

It is necessary to identify the deficiencies in teacher education in order for pre-service teachers to start their profession well-trained and take steps to eliminate these deficiencies. Determining the assessment literacy of pre-service teachers is very important in terms of eliminating educational deficiencies before they graduate (Roslan, Nisho & Jawawi, 2022). In order to improve the assessment knowledge and skills of pre-service teachers in Turkey, there is a two-credit course in one semester in their undergraduate teacher education program. The course is taught two hours per week. The content of this course includes the importance of assessment, basic theoretical knowledge specific to the field (reliability and validity), statistical operations (test and item analysis), types and properties of assessment tools, evaluation, feedback and scoring. However, it is known that the credits of this course are low, the course content is quite inadequate for classroom assessment practices, and the opportunity to practice assessment experience is limited. It is not mandatory for teachers to learn about educational assessment after completing their undergraduate education (Fan, Wang & Wang, 2011). This situation increases the importance of assessment courses in teacher training education.

Many studies indicate that pre-service teachers feel insufficient for assessing learning, preparing assessment tools, grading and evaluating (Roslan, Nisho & Jawawi, 2022; Çelik & Arslan, 2012; Kilmen, Akın Kösterelioğlu, & Kösterelioğlu, 2007). In the studies conducted to determine the literacy levels of pre-service teachers in different branches, it was seen that the literacy levels of pre-service teachers were low (Gül, 2011; Koç, 2019; Berk, 2021).

Along with these, numerous studies have also been conducted on knowledge, skills and views of in-service and pre-service mathematics teachers in the field of assessment. Some of these studies are on assessment literacy (Ergül, 2019), some on attitudes, views or perceptions towards assessment (Göktaş & Şad, 2021; Karakuş, 2010), some on perceptions of assessment competence (Dokumacı Sütçü & Bulut, 2016) and some on their assessment applications (Salmon, 1996; Suziki, 1998; Önel et al., 2020; Tüknüklü, 2003; Güven Akdeniz, 2021). Many math teacher educators agree that it is not possible to include all the needful knowledge and skills in a short-term methods course (Lee & Son, 2015). However, when the course content is based on the main principles or standards of the field, it becomes possible to identify pre-service teachers' strengths and weaknesses.

1.1. Standards for Teacher Competence in Educational Assessment of Students

The American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) have constructed the standards for teacher competence in educational assessment of students in 1990 (Plake, 1993). The standards guide teachers to describe professional roles and responsibilities, and to follow appropriate assessment process. The standards can be summarized as follows (AFT, NCME & NEA, 1990):

Standard 1. *Teachers should be qualified in choosing assessment methods appropriate for instructional decisions.* Teachers need to know the kinds of assessment methods in details with their strengths and weaknesses. It is essential to choose appropriate, beneficial, technically adequate, and fair assessment methods at first step. They should be familiar with the procedures for selecting and evaluating assessment methods in light of lesson plans and lesson contents. Teachers with the knowledge and skills indicated by this standard know that appropriate assessment methods will provide data that helps achieve instructional goals.

Standard 2. *Teachers should be skilled in developing assessment methods appropriate for instructional decisions.* Teachers often use their own implement or design approaches rather than published or external assessment tools when assessing in the classroom. Teachers who meet this standard have the designing according to theoretical information, application, collection of information and decision-making skills.

Standard 3. *The teacher should be skilled in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.* It is important that teachers are able to apply, score, administer, interpret properly as much as select and develop good assessment methods. Teachers should develop scoring key and rating scales to produce consistent results. Teachers should be able to understand and interpret conceptionally the commonly reported scores and statistical terms: percentile ranks, percentile band scores, standard scores, grade equivalents, measures of central tendency, dispersion, relationships, reliability, and errors of measurement.

Standard 4. *Teachers should be skilled in using assessment results when making individual decisions about students, planning teaching, developing curriculum, and improving school.* Teachers should have the conceptual and application skills to make educational decisions using results. They must interpret the results according to curriculum validity and avoid mistakes that may lead to misinterpretations. The assessment information obtained should be used to organize instructional plan in order to help students' educational development.

Standard 5. *Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.* Valid grading is an important evidence of teachers' assessment skills. Grading shows both a student's level of performance and a teacher's valuing of that performance. The principles to obtain valid grades should be known and applied rationally and fairly by teachers. Teachers must avoid incorrect grading procedures such as using grades as punishment and enrich their interpretations about students' attainments based on evidence.

Standard 6. *Teachers should be skilled in communicating assessment results to students, parents, other relevant people, and other educators.* Teachers not only use assessment results for own educational decisions but also report to students, parents, other educators, and different audiences. If assessment results cannot be communicated effectively, they may not be used useful or may not be used at all. To communicate effectively with others about student assessment, teachers must be able to use assessment terms understandably and must be able to explain the meaning, limitations, and implications of assessment results. Teachers may have to confirm their assessment procedures with evidence. Sometimes, they can also help the public to interpret assessment results properly.

Standard 7. *Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.* Teachers should be well-versed about ethical and legal responsibilities in assessment. If they encounter inappropriate assessment practices of others, they should try to discourage them. Teachers should have a fair attitude and behavior that protects the rights of the relevant people in line with the ethical rules of the profession during the application of assessment practices, interpretation, use and communication of the results. They should also know the decisions and laws that affect their students, schools, districts. Teachers should be aware that various assessment procedures can be misused or overused with harmful consequences such as violating students' right to privacy, embarrassing them, or using their scores inappropriately.

1.2. Assessment Literacy Scales Organized According to Standards

Considering the seven standards mentioned above, Plake and Impara developed an instrument named the "Teacher Assessment Literacy Questionnaire (TALQ)" consisting of 35 items to survey teachers' assessment literacy. Questionnaire consists of 35 multiple choice items with four possible answers and one correct answer. The instrument was applied to 555 inservice teachers in the United States. The correct answers average of the test is 23 out of 35 items. It shows that the teachers might not be well competent to assess student learning (Plake et al., 1993). Also, Campbell, Murphy, and Holt (2002) applied the TALQ to 220 pre-service teachers who completed an assessment course. Similarly, it showed that correct answers average 21 out of 35 items. Mertler (2004) used TALQ with a limited amount of reforming to suit the demographic situation of a group of inservice and pre-service teachers. As a result of this research, average 21 of 35 items were found to be correct.

TALQ, which can be defined as the original scale of assessment literacy, is the first scale organized according to AFT&NCME&NAE (1990) standards (Gürsoy, 2017). Scales such as the Classroom Assessment Literacy Inventory (CALI) (Mertler, 2003) and the Assessment Literacy Inventory (ALI) (Mertler and Campbell, 2005), which are based on the 1990 Standards and consist of 35 multiple-choice questions, were created by revising or being inspired by TALQ. The tools that measure assessment literacy created after TALQ have generally started to be standards-based (Ball, 2017).

In Turkey, Gül (2011) applied the new version of TALQ after making some changes as one of the data collection tools in his master's thesis. As a result of the inventory applied to define the assessment literacy of 180 teacher candidates studying in Science, Social Studies, Mathematics and Turkish Language Teaching, it was seen that an average of 18 items were answered correctly out of 34 questions. As a result of the research, it has been determined that teacher candidates have great deficiencies in assessment and evaluation.

In most of the studies examining the assessment and evaluation literacy of pre-service or in-service teachers, it is seen that the scales are prepared as multiple-choice and the data are analyzed quantitatively (Azrak & Yalçinkaya, 2019; Ergül, 2019; Koç, 2019; Yamtim & Wongwanich, 2014; Karaman & Şahin, 2014, Mertler, 2004). These studies are of great importance in terms of revealing literacy levels and identifying deficiencies. However, due to the nature of quantitative research, it is not possible to make in-depth interpretations beyond these results with the data obtained. In this study, it is aimed to reach rich data content and contribute to the studies from a different perspective by making the Turkish version of TALQ, a widely used scale in the field, open-ended by taking expert opinion and adapting it to mathematics education.

1.3. Teaching Experiment

This study was revealed by analyzing the data obtained from the teaching experiment conducted within the scope of the researchers' doctoral thesis. Teaching experiment is a research method that proceeds in the form of redesigning the teaching by analyzing the data obtained while teaching in the classroom with clinical methodology, and cyclically collecting and analyzing data on this basis (Czarnocha & Prabhu, 2006). The teaching experiment begins with the researcher designing and planning a teaching process based on his assumptions. In the framework of the constructivist approach, teaching is conducted by making in-class applications called teaching episodes (Akin & Kaba, 2016). Students are provided with the opportunity to improve their knowledge and skills through educational tools.

In this context, the literature was searched and a curriculum was created for this research that includes the knowledge and skills that pre-service mathematics teachers will need in the field of assessment. The open-ended inventory prepared to examine the effect of the teaching experiment on the assessment literacy of pre-service mathematics teachers was applied as pre-test and post-test. A study of the pre-service mathematics teachers' assessment literacy based on standards for teacher competence would present important data that could be utilized to determine developmental approaches to provide pre-service teachers.

1.4. Research Problem

In this study, the concept of "assessment literacy" was investigated and it was purposed to determine the assessment literacy levels of pre-service mathematics teachers according to the Standards for Teacher Competence in the Educational Assessment of Students. The aim of this study is to reveal the change in the assessment literacy levels of pre-service mathematics teachers according to the standards after the teaching experiment.

2. Method

2.1. Study Group

The participants are eight pre-service teachers in the last year of the mathematics teaching undergraduate program in the 2019-2020 academic year in a state university in north of Turkey. The participants were determined by the criterion sampling method which is one of the purposeful sampling methods. The criteria were determined as having successfully passed the assessment course in the previous year and continuing their internship during the research. The study group consists of eight senior students studying in mathematics teaching who meet these criteria.

These pre-service teachers took the assessment course in the previous year and were successful. During the period when the research was carried out in the internship experience, they had the opportunity to observe experienced teachers and practice own knowledge in the classroom. It is expected that pre-service teachers' internship experiences increase their awareness of the field of assessment, thus strengthening the possibility of this study to reach a deeper and richer content.

2.2. Process

The researchers informed the pre-service teachers that the study was about the assessment literacy. It has been emphasized that the data to be collected will not affect the course grades in any way and has been stated it is hoped that the results of the research will enable better assessment teaching for pre-service teachers in the future.

They were informed that their individual findings would report in no way with own identities. All participants were informed about the research protocol and consent forms were sent. The pre-service teachers signed the consent forms containing information about the research and mailed them to the researcher.

COVID-19 pandemic changed the educational situation from face-to-face to online. For this reason, the research that started face-to-face had to continue online. In this sense, lesson plans have been made suitable for online education. The course was held and recorded through a live online video communication program. Teaching experiment training continued for 16 weeks. Meanwhile, the researchers confirmed that no adverse events were encountered in the ordinary course of the traineeship processes of the pre-service teachers during the research. It was thought that with the internship experience, it would enable the pre-service teachers to examine the practices of an experienced mathematics teacher with what they learned in the teaching experiment.

Classroom Assessment Literacy Inventory in Mathematics Education, which consists of open-ended questions used as a data collection tool in this research, was applied in the form of a live online exam with paper and pencil. At the end of the period, photos of the answer sheets were requested from the pre-service teachers.

2.3. Data Collection Tools

Measurement and Evaluation Literacy Inventory was adapted to the context of Turkey from the 'Teacher Assessment Literacy Questionnaire (TALQ)' prepared by Barbara S. Plake and James C. Impara (1993) to be used in a master's thesis by Gül (2011). The inventory contains five multiple-choice items for each standard of teacher competence in educational assessment of students. After validity, reliability, and pilot studies, Gül (2011) applied the inventory to 180 pre-service teachers.

Classroom Assessment Literacy Inventory in Mathematics Education is an open-ended data collection tool adapted to mathematics education from the multiple-choice Measurement and Evaluation Literacy Inventory edited by Gül (2011). As some questions could not be converted to open-ended format, six out of 35 questions were eliminated. The inventory was reviewed for validity by two academics expert in mathematics education and assessment. Necessary revisions were made based on the review results.

Table-1 specifies which competence standards conform to the open-ended questions adapted to mathematics education, considering the classification of the Classroom Assessment Literacy Inventory questions according to the standards and the elimination of questions.

Yamtim and Wongwanich (2014), who used the Teacher Evaluation Literacy Questionnaire in their research, divided the results into three categories in order to understand the classroom assessment literacy levels of pre-service teachers: low (less than 60%), medium (60-79%), and high (80% and higher). Therefore, in order to evaluate the inventory over a total of 100 points, it was deemed appropriate to distribute the points equally to seven standards. As the number of questions in the sixth and seventh standards is low, the additional 2 points are distributed to these standards as well. An attempt was made to distribute the items equally among the standards in terms of the total score of the standards, and the additional points were shared among the items within the standards that the researchers found appropriate. Accordingly, the pre-test and post-test scores of mathematics pre-service teachers were determined.

Table 1. Classification of Inventory Questions According to Standards

Standards for Teacher Competence in Educational Assessment of Student	Item Number and Scoring	Total Score
1. Choosing appropriate assessment methods	1 (3p), 2 (3p), 3 (4p), 4 (4p)	14
2. Developing appropriate assessment methods	5 (3p), 6 (2p), 7 (3p), 8 (3p), 9 (3p)	14
3. Administering, scoring and interpreting learning outcomes	10 (3p), 11 (3p), 12 (3p), 13 (3p), 14 (2p)	14
4. Using assessment outcomes to make decisions	15 (2p), 16 (3p), 17 (3p), 18 (3p), 19 (3p)	14
5. Developing valid grading procedures	20 (4p), 21 (3p), 22 (3p), 23 (4p)	14
6. Communicating assessment results	24 (5p), 25 (5p), 26 (5p)	15
7. Knowing unethical practices	27 (5p), 28 (5p), 29 (5p)	15

2.4. Data Analysis

The qualitative and quantitative data obtained from the research were evaluated separately. To obtain quantitative data, a graded scoring key was created for the responses to the inventory consisting of open-ended questions. The responses of each pre-service teachers were scored separately by two researchers according to the graded scoring key. Inter-rater reliability computation was performed to determine the consistency between the scores given by multiple raters (Bikmaz Bilgen & Doğan, 2017). The calculation of the Krippendorff Alpha coefficient is done with the formula $= 1 - \frac{D_o}{D_e}$. D_o (observed discrepancy) and D_e (expected discrepancy) are calculated with formulas using the data matrix created from the scoring of the raters (Kanik et.al, 2010).

According to the Krippendorff Alpha coefficient calculation, inter-rater reliability was calculated as 0.81. An alpha value of 0.80 and above indicates high agreement (Arslan Mancar, 2019). Pre-test and post-test scores were determined by reaching a consensus by discussing the differences in scoring. The obtained quantitative data were analyzed through descriptive statistics in terms of frequency distribution and mean.

Qualitative research aims to interpret and describe people's qualities such as thoughts, opinions, and attitudes in detail (Kıral, 2020). Document analysis is one of the data collection methods used in qualitative research. In educational research, the impact of a study can produce documents for researchers to obtain detailed information according to specific purposes, and participants can be asked to create the content of these documents (Sak et al. 2021). Therefore, the pre-service teachers' inventory responses were categorized into general themes, and detailed analyses were started for each theme. In these analyses, data were processed not by quantitative analysis by converting them into numbers, but in a way that could draw inferences related to the topic being investigated.

3. Results and Discussion

3.1. Quantitative Analysis of Inventory Results

Through the analysis of pre-test responses to the inventory, it was revealed that mathematics pre-service teachers scored between 47 and 62 out of 100 points on assessment literacy, with an average total score of 55.75. When pre-test scores were examined based on seven standards for teacher competence in educational assessment, as shown in Table 2, the average scores ranged from 4.125 to 13.625. Standard 7, Knowledge of Unethical Practices, received the highest average score (13.625), while Standard 2, Developing Appropriate Assessment Methods, received the lowest average score (4.125).

In the scoring of post-test responses to the inventory, mathematics teacher candidates scored between 49 and 73 out of 100 points, with an average total score of 61.375. When scores were examined based on each standard, as shown in Table 2, the average scores ranged from 3.625 to 13. Standard 7, Knowledge of Unethical Practices, received the highest average score (13), while Standard 6, Communicating Assessment Results, received the lowest average score (3.625).

In this study, it was observed that most of the pre-service teachers had low classroom assessment literacy scores before the teaching experiment. This result is similar to the research results of Berk (2021), Koç and Bulut (2020), Koç (2019), İzci et al. (2018), Gül (2011) and Mertler (2004), who conducted studies on the assessment knowledge and skills of pre-service teachers.

Ergül and Çetin (2021) conducted a study with 189 prospective secondary school teachers to determine their assessment literacy levels according to the competencies standards and found that the average number of correct answers was 13 out of 30 questions (43%). Karaman and Şahin (2014) applied the Assessment Literacy Inventory, which was inspired by the original inventory applied in this study, to 289 pre-service teachers and found that the average scores of pre-service teachers were (53%). The pre-test mean score (55.75%) and post-test mean score (61.375%) of this study were slightly higher than the averages in Ergül and Çetin (2021) and Karaman and Şahin (2014). It can be said that the education provided for assessment literacy contributed to the development of pre-service teachers.

Table 2. Comparison of Pre-test and Post-test Inventory Scores

Sample size n=8	Minimum score		Maximum score		Average		Standard Deviation		Total Score
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	
1.Choosing appropriate assessment methods	4	4	11	11	7.625	8.125	2.774	2.357	14
2.Developing appropriate assessment methods	1	4	7	9	4.125	8	2.1	3.251	14
3.Administering, scoring and interpreting learning outcomes	5	5	7	11	6.125	8.375	0.991	2.722	14
4.Using assessment outcomes to make decisions	8	8	11	14	8.875	11	1.246	1.69	14
5.Developing valid grading procedures	5	6	11	13	8.375	9.25	1.847	2.659	14
6.Communicating assessment results	5	0	8	8	7	3.625	1.414	2.722	15
7.Knowing unethical practices	9	8	15	15	13.625	13	2.2	2.673	15

With the scale inspired by the original inventory used in this study, in studies conducted abroad, pre-service teachers' correct answer means was found by Mertler and Campbell (2005) 68%, by Mertler (2003) 54%, by Campbell et al. (2002) 60%. In studies conducted with pre-service teachers both in Turkey and abroad on assessment literacy, it is seen that pre-service teachers are at low and medium levels. It is thought that this situation is mostly due to the fact that this field is not given enough importance in teacher education (Siegel & Wissehr, 2011).

The scores of TALQ which is the source of our inventory is divided into three categories by Yamtim and Wongwanich (2014) to evaluate the results of the scale applied to pre-service teachers: low (less than 60%), medium (60-79%), and high (80% and higher). Similarly, Table 3 shows the levels of eight pre-service teachers according to each standard based on the evaluation of pre-test and post-test results on a 100-point scale.

Table 3. Score Level of Pre-service Teachers According to Standards (n=8)

Standards	Low		Medium		High	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
1.Choosing appropriate assessment methods	5	5	3	3	-	-
2.Developing appropriate assessment methods	8	5	-	1	-	2
3.Administering, scoring and interpreting learning outcomes	8	5	-	2	-	1
4.Using assessment outcomes to make decisions	5	1	3	6	-	1
5.Developing valid grading procedures	4	4	4	2	-	2
6.Communicating assessment results	8	8	-	-	-	-
7.Knowing unethical practices	-	1	1	1	7	6
Overall Situation	5	3	3	5	-	-

Accordingly, it is seen that most of the pre-service teachers received poor scores for five standards in the pre-test results: Standard 1: choosing appropriate assessment methods (5/8), Standard 2: developing appropriate assessment methods (8/8), Standard 3: administering, scoring and interpreting learning outcomes (8/8), Standard 4: using assessment outcomes to make decisions (5/8), and Standard 6: communicating assessment results (8/8). According to the pre-test scores, pre-service teachers are only at the fair level for one standard: Standard 5: developing valid grading procedures (4/8). When the pre-test scores of pre-service teachers are examined, it is seen that they are at an upper level regarding Standard 7: knowing unethical practices (7/8). Teachers need to make accurate assessments in order to develop students in accordance with the educational program (Yamtim & Wongwanich, 2014). According to the pre-test results, pre-service teachers' scores were mostly low. In order to ensure that students can achieve the teaching aims, pre-service teachers need to improve their assessment literacy knowledge and skills.

According to the post-test scores obtained at the end of the teaching experiment designed for the assessment field in mathematics education and applied to pre-service teachers, the level of pre-service teachers in each standard can be seen in Table 3. It is seen that most of the pre-service teachers are at a low level in four of the seven standards. These standards are Standard 1: choosing appropriate assessment methods (5/8), Standard 2: developing appropriate assessment methods (5/8), Standard 3: administering, scoring and interpreting learning outcomes (5/8), and Standard 6: communicating assessment results (8/8). According to the post-test scores, pre-service teachers are at a fair level in Standard 4: using assessment outcomes to make decisions (6/8) and Standard 5: developing valid grading procedures (2/8). As in the pre-test scores, it is seen that pre-service teachers are at an upper level regarding Standard 7: Knowing unethical practices (6/8).

When Table 3 is examined, it can be seen that pre-service teachers partially transition from low to medium and high levels in Standards 2, 3, 4, and 5. To the inventory results, the teaching experiment cannot be considered effective for Standards 1, 6, and 7. Similarly, In the study of Beziat and Coleman (2015), no significant difference was found in the results of the TALQ applied as a pre-test and post-test to determine the effect of the assessment training given during teacher education.

3.2. Qualitative Analysis of Inventory Results

The distribution of 29 open-ended questions in the inventory to seven standards is given in Table 1. The responses of eight pre-service teachers to the questions related to the standards were coded and subcategories and categories were formed. The answers given by the pre-service teachers were expressed with frequencies in the relevant categories. No frequency was given for the questions left unanswered. In addition, the researchers present some of the answers given to the readers with direct quotations.

Standard 1. *Teachers should be qualified in choosing assessment methods appropriate for instructional decisions.* In order to determine the competencies in this standard, questions directed to pre-service teachers were about ensuring reliability, selecting valid assessment procedures, and determining academic achievement and motivation.

The frequency distributions of the answers to the questions asked to the pre-service teachers about the most important issues for selecting classroom assessment practices are given in Table 4. As can be seen in Table 4, pre-service teachers' opinions about what needs to be done to achieve valid assessment results differed in the pre-test and post-test. Three out of eight pre-service teachers in the pre-test and six out of eight in the post-test stated that the assessment tool should be appropriate to the instructional purposes. Classification of items according to cognitive taxonomies was mentioned only by 3 pre-service teachers in the pre-test. The selection of items to identify learning deficiencies was stated by 5 in the pre-test and 4 in the post-test. The pedagogical suitability of the assessment tool was stated by 2 in the pre-test and 3 in the post-test. The selection of a assessment tool suitable for the subject was stated by one pre-service teacher in the pre-test and one in the post-test. Creating items that attract students' interest was stated by 3 pre-service teachers in the pre-test and 1 pre-service teacher in the post-test. Only 2 pre-service teachers in the pre-test stated that giving importance to grammar and visual factors while preparing the assessment tool would increase the validity of the assessment results.

Table 4. Pre-service Teachers' Answers to the Questions Prepared for Standard 1 in the Pre-test Application of the Inventory

Categories	Subcategories	Pre-test (f)	Post-test (f)
Preparations to be made to reach valid assessment results	Appropriateness to the instructional purposes	3 (S1, S3, S6)	6 (S1, S3, S4, S5, S6, S8)
	Classification according to cognitive taxonomies	3 (S1, S2, S4)	-
	Selecting items to identify learning deficiencies	5 (S1, S3, S5, S7, S8)	4 (S3, S4, S5, S7)
	Pedagogical appropriateness	2 (S6, S7)	3 (S2, S4, S7)
	Appropriateness to the structure of the subject	1 (S4)	1 (S7)
	Attracting students' interest	3 (S3, S4, S6)	1 (S3)
	Grammar and visual factors	2 (S5, S7)	-
Theoretical Properties	Reliability	1 (S4)	6 (S1, S3, S4, S5, S6, S8)
	Validity	1 (S2)	2 (S5, S6)
	Consistency	1 (S5)	2 (S1, S5)
	Practicality	2 (S4, S7)	1 (S4)
	Objectivity	2 (S1, S8)	2 (S1, S5)
	Item difficulty	4 (S1, S3, S5, S7)	1 (S6)
	Item discrimination	-	2 (S2, S6)
Type of assessment	Diagnostic	2 (S2, S6)	-
	Formative	2 (S3, S4)	3 (S6, S7, S8)
	Summative	2 (S3, S4)	-
Type of assessment tool	Cognitive	7 (S1, S2, S4, S5, S6, S7, S8)	5 (S1, S2, S3, S5, S7)
	Affective	3 (S4, S5, S8)	7 (S1, S2, S3, S4, S5, S7, S8)

It is a more accurate approach to discuss the validity of interpretations based on results obtained from a tool that assesses students' achievements regarding a taught subject than to talk about the validity of the test itself. Therefore, validity can vary depending on instructional purposes to assess, the applied group, and the conditions (Demircioğlu, 2015). Validity depends largely on some points taken into consideration while preparing the test, such as having a sufficient number of items covering the topics and learning levels that the test aims to assess, selecting an appropriate assessment tool, and writing the items in a clear and understandable way. When asked about what they pay attention to the most when determining the assessment method related to a subject taught to pre-service teachers, only three out of eight pre-service teachers mentioned the appropriateness of the assessment tool to instructional purposes in the pre-test. Ensuring the content validity of the test is one of the most important issues to be considered while preparing a test. Below are the answers given by two different pre-service teachers in the pre-test regarding the aspects they pay attention to the most in the assessment tool.

S3: "First, I prepare a test to check whether instructional purposes have been achieved. While preparing this, I take care to include various items."

S7: "To assess the subject I taught, I would choose activities that are easy to visualize in children's minds. I would choose a grid and add visuals related to the subject because I would have firstly explained with pictures. I would like to receive feedback in the way I ask."

In the post-test results, it is seen that the answers regarding the relevance of the assessment tool to the instructional purposes were included in the responses of six out of eight pre-service teachers. The responses of the pre-service teachers include qualities such as being appropriate to the purposes and content, being suitable for the level and characteristics of the students, being economical and practical, using alternative assessment methods that are compatible with the constructivist approach, ensuring content validity and avoiding errors in assessment.

S4: "I pay attention to its suitability to the purposes and content, to be suitable for the level and characteristics of the students, and to be economical and practical."

S8: "I pay attention to whether the test serves the purpose I want to assess. I try to prepare the questions in a way that does not cause chance success."

As can be seen in Table 4, the answers given by the pre-service teachers in the category of theoretical properties about selecting a assessment tool can be seen. In the pre-test, one pre-service teacher mentioned the concepts of validity and reliability, which are the most basic theoretical features that should be present in the assessment tool. In the post-test, six pre-service teachers mentioned the concept of reliability and two mentioned the concept of validity.

When pre-service teachers were asked about the most valid evidence that can be obtained through assessment in understanding whether students are motivated by their courses and academic achievement, it is seen that the answers were given as 'attitude tests' and 'surveys' in the post-test. As seen in Table 4, in the pre-test, 7 of the 8 pre-service teachers stated that they could obtain their knowledge from cognitive and 3 from affective assessment tools. In the post-test, 5 of 8 stated that they could obtain their knowledge from cognitive and 7 of 8 from affective assessment tools. The most valid evidence showing whether students gain the knowledge and skills that teachers want to impart through teaching practices are data obtained from different assessment tools (Bahar et al.,2015). A strong motivation is required to increase students' interest in mathematics (Şimşir Gökalp, 2021). Considering the affective characteristics of students is an element that will affect the success of the course. Cognitive and affective tests should be evaluated together as an indicator of motivation related to their academic achievement.

Some answers from the pre-test data on pre-service teachers' emphasis on cognitive tests;

S1: "Which of the cognitive level steps do the questions he solves cover? How is participation in process and outcome evaluations?"

S2: "Getting high scores in multiple-choice tests can be given as an example. Having done his duty with active participation in the project and performance assessment"

The answers showing the importance given to affective tests by the pre-service teachers in the post-test data are given below.

S3: "Attitude scales can be applied to students. The results obtained from these scales can be evaluated and a conclusion can be reached. Thus, we have both a proof, and we can learn from the students whether they are motivated or not."

S8: "The carefully prepared attitude scales can be applied as pre-test and post-test. The results obtained from here inform the teacher."

Standard 2. *Teachers should be skilled in developing assessment methods appropriate for instructional decisions.* Regarding the competencies of this standard, questions prepared within the scope of determining an assessment strategy, increasing the reliability of the test, assessing high-level thinking skills, preparing items, and reaching correct assessment decisions were asked to the pre-service teachers.

For the validity of the results obtained from a assessment tool, it is not possible to make comments only by statistical procedures. Some studies are needed to make a decision about validity. Examination of the tests by experts and comparison of the results obtained from other tests with statistical procedures are examples of these studies (Demircioğlu, 2015). It is seen that the pre-service teachers' answer of comparing with different assessment results in order to decide on the validity of the scores obtained increased in the post-test. In both the pre-test and post-test, only one pre-service teacher stated that the appropriateness of the test to the instructional purposes should be examined in terms of the validity of the scores. The decrease in the number of pre-service

teachers who answered "utilizing statistical procedures" in the post-test may indicate that validity in classroom assessment is understood conceptually rather than statistically.

Table 5. Pre-service Teachers' Answers to the Questions Prepared for Standard 2 in the Pre-test Application of the Inventory

Categories	Subcategories	Pre-test (f)	Post-test (f)
Deciding on the validity of the scores	Checking the conformity of the test to the instructional purposes	1 (S1)	1(S4)
	Comparison with different assessment results	2 (S6, S8)	6 (S1, S2, S3, S5, S7, S8)
	Evaluating student answers	2 (S1, S7)	-
	Utilizing statistical procedures	3 (S1, S2, S4)	1 (S4)
Factors that increase the reliability of multiple-choice tests	Appropriateness to the instructional purposes	1 (S3)	-
	Getting expert opinion	1 (S5)	-
	Number of students participating in the test	1 (S7)	-
	Examination environment	1 (S4)	2 (S5, S7)
	Exam duration	1 (S3)	2 (S3, S7)
	Item writing	1 (S3)	3 (S3, S4, S5)
	Item difficulty	1 (S4)	-
	Number of questions	2 (S3, S4)	1 (S4)
	Suitability to student level	-	1 (S7)
	Arrangement of distractors	-	4 (S1, S2, S4 S6)
Writing a sample item suitable for the instructional purpose	Correction formula	-	2 (S2, S5)
	Appropriate	3 (S3, S4, S6)	6 (S2, S4, S5, S6, S7, S8)
	Inappropriate	5 (S1, S2, S5, S7, S8)	2 (S1, S3)

As seen in Table 5, when asked how a teacher should decide on the validity of the scores according to the assessment strategy determined by the teacher, 1 out of 8 pre-service teachers in the pre-test and post-test gave the answer of checking the conformity of the test to the instructional purposes. The answer of comparing the test results with different assessment results was given by 2 pre-service teachers in the pre-test and 6 pre-service teachers in the post-test. Only in the pre-test, 2 pre-service teachers stated that the validity of the scores should be decided by evaluating the student answers. 3 pre-service teachers in the pre-test and 1 pre-service teacher in the post-test stated that statistical procedures should be used in the validity of the scores.

The answers given in the pre-test are mostly that they will make this decision with statistical processes.

S2: "By doing standard deviation and item analysis"

S4: "He calculates the validity and reliability of the assessment tool and interprets according to the results."

In the post-test data, the sample answers given by pre-service teachers to the comparison with other assessment results are presented below.

S3: "He can determine the students' success in class by comparing the results of their assessments."

S8: "It is difficult to obtain information about students with a single assessment tool. At the very least, if more than one test is administered as a diagnostic, formative and summative, and students' responses are consistent, it will better determine the validity of the results."

Table 5 shows the answers given by the pre-service teachers when asked about the factors that increase the reliability of the multiple-choice test. Two of the pre-service teachers stated that the number of questions in the pre-test affected the reliability. In Table 5, it is seen that the frequencies of the answers of the pre-service teachers in the sub-categories are appropriateness to the instructional purposes (1), getting expert opinion (1), number of students participating in the test (1), examination environment (1), exam duration (1), item writing (1), item difficulty (1), number of questions (2). When the subcategories in the post-test are analyzed in Table 5,

it is seen that the subcategories are examination environment (2), exam duration (2), item writing (3), number of questions (1), suitability to the student level (1), arrangement of directions (4) and correction formula (2).

Reliability is related to the sensitivity, consistency and stability of measurement results. Therefore, homogeneity in terms of subject and behavior, the large number of items, the clear and understandable writing of the items, the application of the test in good conditions and the objective scoring are the factors that increase reliability (Demircioğlu, 2015). While preparing multiple-choice tests, the creation of choices is as important as item writing (Yılmaz, 2015). It is noteworthy that pre-service teachers mentioned these factors more in the post-test.

Examples of pre-service teachers' answers to increase the reliability of the multiple-choice test in the pre-test data;

S5: "The reliability of the test can be increased by being evaluated by more than one expert"

S7: "I think it should be evaluated over more participants..."

From the post-test data, it can be said that pre-service teachers have more comprehensive knowledge about increasing the reliability of the multiple-choice test.

S3: "The number of questions can increase to a certain extent. Questions can be asked simply and clearly. Distractors can be prepared to serve their purpose."

S4: "Preparing the questions homogeneously, using the correction formula for applying the test to a heterogeneous group, and applying the test in standard environments will increase reliability."

S7: "It must have been prepared in accordance with the level and applied in a certain environment, and also for a certain period of time. That way it can be increased."

Standard 3. *The teacher should be skilled in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.* Regarding this competence standard, the questions prepared within the scope of time management in interpreting the test scores, choosing the appropriate scoring method, evaluating the student scores and applying the test were directed to the pre-service teachers.

Table 6. Pre-service Teachers' Answers to the Questions Prepared for Standard 3 in the Pre-test Application of the Inventory

Categories	Subcategories	Pre-test (f)	Post-test (f)
Evaluation of a high score in a test	Does not mean absolute success	1 (S2)	6 (S2, S3, S4, S5, S6, S8)
	The subject was learned	7 (S1, S3, S4, S5, S6, S7, S8)	2 (S1, S7)
Project/product evaluation strategy	Using question-answer method	1 (S5)	-
	Preparing a rating scale	1 (S4)	4 (S3, S4, S7, S8)
	Process observation	1 (S3)	-
	Preparing a scoring key	2 (S1, S8)	-
	Creating a performance task	2 (S2, S6)	4 (S1, S2, S5, S6)
	Giving full points to the correct answer	1 (S7)	-

Pre-service teachers were asked about their comments about a student who got 90 points out of 100 points. The answers given by the pre-service teachers are divided into two sub-categories as seen in Table 6. In the pre-test, 7 pre-service teachers stated that students with high scores learned the subject, 1 pre-service teacher stated that this did not mean absolute success. In the post-test, 6 pre-service teachers stated that a high score in a test does not mean absolute success, while 2 pre-service teachers stated that the subject was learned to a great extent.

Examples of the answers given to this question from the pre-test data are presented below.

S3: "If a teacher makes the scoring process out of 100 and there is a student who gets 100 in this exam, it means that there is no deficiency in the instructional process that are intended to be achieved. From this point of view, it can be said that a student who gets 90 has a few shortcomings in the way of gaining the aims. It is necessary to discover where this deficiency is, why it is and be corrected."

S4: "The student may have acquired the aimed gains and stuck with a question that requires high-level thinking skills. There may have been one of the types of error in the assessment."

S5: "The student who got 90 must have understood the subject. The reason why he did not get 100 may be carelessness, confusion."

In the post-test data, most of the pre-service teachers stated that different assessment results should be used in interpreting student achievement. Examples of their responses to this question are presented below.

S1: "It is not possible to assess the success of a student with a single exam. Because chance success can be effective depending on the content of the exam. Or there is a subject that the student knows. The content validity of the test is low, so it may have been high scores. Of course, apart from these negativities, the student may have really learned and done."

S8: "The level of the questions and the observed success of the student in the class are important. Correct answers may have been given to the questions by luck. But in general, if the questions are not suitable for this, if the student is successful in the lessons, if he is active in the class, if his scores are consistent in different assessments, it can be evaluated that he has understood the subject."

Table 6 shows that the frequencies of the answers given by the pre-service teachers to the question of the teacher's evaluation of a model development assignment on the subject are using the question-answer method (1), preparing a rating scale (1), process observation (1), giving full points to the correct answer (1), preparing a scoring key (2) and creating a performance task (2).

Performance assessments are a type of assessment based on a problem that allows for different solutions, where various approaches and materials are tried and higher level skills are used. Performance assessment should be spread over a long period of time, cover many skills, include group and individual assessments, focus on the product and process, and use many data collection techniques (Çepni, 2015). In performance assessments, first of all, the teacher should create contextual tasks and determine the assessment criteria and stages (Palm, 2008). When the pre-test answers of the pre-service teachers were analyzed, they partially included the stages of performance assessment. Most of the answers given in the post-test addressed the planning and evaluation processes of the performance task together.

When the pre-service teachers were asked about the assessment strategy to be followed when students were asked to develop a model for measuring the area of a triangle, examples of the answers given in the pre-test are presented below.

S5: "An open-ended assessment may be more appropriate here. Questions such as 'Why did you design this model? What does this model show us?' can be a good method for assessment."

S7: "Full points can be given to students who succeed in creating materials that can determine the height and base. Such a strategy can be followed."

The effect of including performance assessment methods in the course content of the teaching experiment and practicing them is felt in all the answers given to this question in the post-test.

S2: "Performance task can be given. It can make a process-oriented evaluation, not only result-oriented."

S3: "Analytical rubric should be used. There should be a report at the end of the process."

S5: "It would be correct to evaluate this study as a performance task. While students are developing the model, keeping the process under control with short interviews will make grading more accurate."

Standard 4: *Teachers should be skilled in using assessment results when making individual decisions about students, planning teaching, developing curriculum, and improving school.* Regarding this standard's competencies, the questions prepared within the scope of determining the readiness of the students, predicting the test score, using and criticizing the assessment results were asked to the pre-service teachers.

Mathematics is a science consisting of theories based on system and order (Nasibov & Kaçar, 2005). Mathematics education, unlike other branches of science, comprises certain mathematical processes that include numbers, space, objects and relationships between events (Baki, 2008). For this reason, the assessment methods that mathematics teachers choose as a reference in the evaluation processes of their students are important.

Table 7. Pre-service Teachers' Answers to the Question Prepared for Standard 4 in the Pre-test Application of the Inventory

Category	Subcategories	Pre-test (f)	Post-test (f)
Type of knowledge preferred in mathematics lesson assessments	Norm-based knowledge	2 (S1, S4)	3 (S1, S4, S5)
	Criterion-based knowledge	6 (S2, S3, S5, S6, S7, S8)	5 (S2, S3, S6, S7, S8)

While assessing the arithmetic operations in the fractions subject in the mathematics lesson, the pre-service teachers were asked whether they prefer norm-based knowledge (evaluation of each student's performance according to the performance of other students) or criterion-based information (evaluating each student by referring to the individual success revealed according to learning outcomes). As seen in Table 7, in the pre-test, 2 pre-service teachers said that norm-based knowledge and 6 pre-service teachers said that evaluation should be based on criterion-based knowledge. In the post-test, 3 pre-service teachers stated that norm-based knowledge and 5 pre-service teachers stated that criterion-based knowledge should be preferred.

The answers regarding the preference of norm-knowledge in the assessment of mathematics lesson from the data obtained in the pre-test are as follows.

S1: "Norm-based knowledge should be used. Because he cannot prepare a separate lesson plan for each student. It cannot wait for each student to reach a certain level in achievements. It should also save time."

S4: "Norm-based knowledge is more useful. While planning the lesson, it is necessary to see the situation of the students collectively and to follow the path accordingly."

Similarly, from the post-test data, it is seen that there is a conceptual confusion about the individual assessment criteria of the student and the average achievement of the group. In terms of mathematics education, it should be emphasized that individual feedback should be given to students and that student success should be associated with learning contents. Correct and effective feedback is an important step in good communication between students and teachers (Türnüklü, 2003).

S1: "I recommend using norm-based knowledge. Topics should be adjusted according to the general situation of the society. There is not enough time in your education system to cater to every student."

S4: "He can have information about the general situation of the classroom by using norm-based information. In this way, he makes a planning by considering the whole class."

S5: "I think norm-based information would be more useful for this subject. The interval in which the students gather in general will give us more accurate results in this regard."

Standard 5. *Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.* Pre-service teachers were asked questions about designing assessments, ranking student achievements, and grading in relation to this standard's competencies.

Pre-service teachers were asked to state an assessment design that best reveals student success and performance and allows for a reliable assessment.

S7: "They are activities. It can be a concept map and a branched tree."

S8: "The questions in the prepared test should be well thought out, the margin of error is low, and items that question information at all levels should be prepared."

In the post-test, it is seen that pre-service teachers mentioned diagnostic, formative and summative assessment designs. At the beginning of the lesson a diagnostic assessment that assesses the information that forms the basis of the subject, a formative assessment that keeps the student active during the course of the lesson, and a summative assessment that determines whether the learning takes place at the end of the lesson should be done (Uğurlu & Akkoç, 2011).

S2: "When alternative and traditional assessment tools are used together"

S7: "I think using formative assessment tests can make it reliable."

S8: "I think it is more beneficial to apply more than one test in the formative and summative assessments and to assess the consistency of the students' scores, as we do."

After the teaching experiment, pre-service teachers stated that traditional and complementary assessment methods should be used together. In their study, Baştürk and Dönmez (2011) found that pre-service mathematics teachers shared ideas suitable for the complementary assessment and evaluation approach, but they could not reflect them in their lessons.

Standard 6. *Teachers should be skilled in communicating assessment results to students, parents, other relevant people, and other educators.* Regarding this competence standard, questions about evaluating test results, explaining success differences, and making valid comments on student achievements were directed to pre-service teachers.

Pre-service teachers were told that a student got a score out of 100 in Science: 75, Social Studies: 75 and Mathematics: 75. Looking at the scores of this student, they were asked to make a valid comment on their success in the lessons.

In the pre-test, examples of answers given by the pre-service teachers regarding that the student in question has learned adequately are given below.

S5: "He understood the subject in both verbal and numerical lessons. But in-depth learning could not be achieved."

S6: "The student has an average success."

S7: "The student knows the subject in general terms, but he has some deficiencies. It means that some information is not formed enough."

It is seen that the post-test data are compatible with the pre-test data.

S1: "In this case, he seems to be similarly competent in all three courses. I don't think 75 points is low, I think he is a good student."

S4: "The student has achieved most of the goals desired to be achieved. The parts that are missing or have difficulties should be identified and an appropriate path should be followed for their development."

S8: "I think that he may have understood the essence of the subjects, but that he did not look deeply into the subjects and was limited to what the teacher told them."

Standard 7. *Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.* Related to this standard competencies, pre-service teachers were asked questions about protecting the confidentiality of test data, fair practice and violation of ethical rules.

All of the pre-service teachers gave correct answers to the questions asked about the violation of the ethical rules of assessment. A teacher tells his students that they can look at their exam papers after the lesson at the teachers' room so that they can learn the results of the exam. The students were also asked what is wrong with looking at their own marks by seeing the papers of their other friends.

The pre-test responses that are not suitable for this standard, in which pre-service teachers get the highest scores, are presented below. It is noteworthy that a pre-service teacher interpreted an unfair practice as if it were right while explaining the reason why it was wrong for students to see the answer sheets.

S1: "Due to the student's effort in the lesson, teachers sometimes correct minor mistakes. Students can see this and react to their teachers or friends. Parent, school administration, this may take longer. Instead, it would be ideal for him to solve the questions in class or show his own answer key."

4. Conclusion and Recommendations

Assessment literacy focuses on one's understanding of assessment concepts and procedures. That is, an assessment-literate educator needs to understand what really constitutes a test's reliability and how effects its differences (Popham, 2011). It is necessary to identify the deficiencies by reaching the pre-service teachers' understanding of the assessment field and by obtaining indicators related to their assessment literacy.

The findings of the study show that pre-service mathematics teachers' assessment literacy is at a low level and needs to be improved. The knowledge and skills that pre-service teachers have acquired with a two-credit theoretical course in only one semester of their undergraduate education are not sufficient to master assessment approaches and techniques. Despite the increasing emphasis on assessment knowledge and skills in the education community, the number of assessment courses in pre-service education programs remains low. It is also known that the contents of these courses are not sufficient for a teaching career (DeLuca & Klinger, 2010; Koç, 2019). Also, pre-service teachers need longer hours to become more familiar with assessment practices during internships. Pre-service teachers need to be exposed to various assessment practices so that they can think critically, learn to cope with some difficulties, and improve themselves (Xu & He, 2019).

Both quantitative and qualitative data findings of the study show that the teaching experiment positively affected pre-service teachers' assessment literacy levels. Similarly, it has been observed that some of the trainings conducted to assess pre-service mathematics teachers had positive results. Uğurlu and Akkoç (2011) found that after the workshop they held, pre-service teachers gained their positive effects such as giving grades, detecting problems in assessment process, and improving the lesson. In order to achieve better results, the deficiencies in the course content of the teaching experiment applied in this study should be eliminated and made in accordance with the standards.

Pre-service teachers may find it difficult to transfer what they learned in assessment courses in undergraduate education to the real classroom environment. For this reason, it is necessary to support them on how to transfer what they have learned and how to evaluate them better (Yan & Pastore, 2022). Pre-service teachers need to understand assessment literacy as the ability to facilitate and meaningfully assess student learning rather than

assess student knowledge (Clark, 2015). Internship experiences can inspire pre-service teachers about how an experienced teacher does it. When the pre-service teachers were asked which assessment methods they observed during their internship experience, all of them gave the names of the traditional methods (Özdil & Kaçar, 2022). The fact that methods whose benefits are emphasized in theory are not preferred in practice may reduce the value and interest of pre-service teachers in these methods. For this reason, faculty-school cooperation and exchange of ideas should be active throughout the course experiences during the internship process.

The findings and results of this study are intended to contribute to the literature and serve to improve pre-service teachers' assessment education. Due to the low levels for assessment literacy of pre-service mathematics teachers, changes can be made in undergraduate education programs. Assessment courses can be run with practices as well as theoretical knowledge. More studies are needed to examine pre-service teachers' assessment literacy. While the important findings of this work, there are some limitations. A similar study can be done with pre-service mathematics teachers from different faculties. Although this study has important findings, it has some limitations such as being conducted online, in one faculty and with eight pre-service teachers. If a similar study is conducted face-to-face with pre-service mathematics teachers in different faculties, it may strengthen our understanding of the phenomenon of pre-service teachers' assessment literacy.

Author contributions: All authors contributed to the design, collection of data, analysis and interpretation of the results. They all approve of the final work.

Ethics declaration: Author declared that the Ethics Committee of Kastamonu University approved the study, on 12 October 2020 with approval code: 3/17.

Funding: No funding was reported for this study.

Declaration of interest: The authors declare no competing interest.

Bir Öğretim Deneyinin Matematik Öğretmen Adaylarının Değerlendirme Okuryazarlıklarına Etkisinin Öğretmen Yeterlik Standartlarına Göre İncelenmesi

1. Giriş

Öğretmenler, öğretimin amaçlara uygun olarak yürütülüp yürütülmediğini birinci elden kontrol eden, öğretim ve sınavlarla ilgili kararlar veren kişilerdir (Yamtim & Wongwanich, 2014). Öğretmenlerin en temel görevlerinden biri öğrenci performansını değerlendirmektir (Fan, Wang & Wang, 2011). Öğretmenlerin eğitsel değerlendirme konusundaki bilgi ve becerileri, değerlendirme okuryazarlığı kavramı ile ifade edilmektedir (Alkharusi, 2011). Stiggins (1995) değerlendirme okuryazarlığını şöyle tanımlamıştır: 'Sağlam ve sağlam olmayan değerlendirme arasındaki farkı, neyi değerlendirdiklerini, neden yaptıklarını, ilgilendikleri beceriyi/bilgiyi nasıl değerlendireceklerini, öğrenci performansının sağlam örneklerini nasıl üreteceklerini, değerlendirmede potansiyel olarak neyin yanlış gidebileceği ve bu sorunların ortaya çıkmadan önce nasıl önleneceği bilirler'.

Bireysel öğrenme farklılıkları, herkese uyan tek bir değerlendirme modelinin mümkün veya uygulanabilir olmadığını bir göstergesidir (Lee & Son, 2015). Bu farklılıklar dikkate alınarak öğretime yardımcı olabilecek yöntemlerin seçimi ve uygulanması, öğrenme eksikliklerinin tespiti öğretmenlerin bu konudaki uzmanlığına bağlıdır. Değerlendirme kapasitesine sahip öğretmenler, öğrencilerin öğrenmelerini, akıl yürütmelerini ve inançlarını geliştirebilir. Bu nedenle değerlendirme bilgisi becerileri önemli bir mesleki yeterlilik olarak görülmektedir (Fan, Wang & Wang, 2011).

Son yıllarda öğretmenlerin değerlendirme uzmanlığı daha fazla önem kazanmıştır (Mertler, 2004). Öğretmenlerin değerlendirme bilgi ve becerileri üzerine yapılan araştırmalarda öğretmenlerin çoğu sınıf değerlendirme okuryazarlığına düşük düzeyde sahiptir (Yamtim & Wongwanich, 2014; Ergül, 2019). Ayrıca öğretmenler lisans ve hizmet içi eğitimde yeterli değerlendirme eğitimi alamadıklarını belirtmişlerdir (Gelbal & Kelecioğlu, 2007; Kilmen & Çıkrıkçı Demirtaşlı, 2009).

Kanada, Birleşik Krallık ve Amerika Birleşik Devletleri'nde kamuya hesap verebilirlik kapsamında artan geniş ölçekli değerlendirmenin yanı sıra sınıflarda şeffaf değerlendirme uygulamalarına verilen önemin arttığı görülmektedir. Öğretmenlerin öğrenci ölçme ve değerlendirme alanında yetkinliği bir zorunluluk haline gelmiştir. Bu nedenle mesleğe nitelikli ve özgüvenli bir başlangıç için hizmet öncesi öğretmen eğitimi gözden geçirilmelidir. Hem büyük ölçekli hem de sınıf içi değerlendirmenin kapsamlı önemi göz önüne alındığında, değerlendirme okuryazarlığı, hizmet öncesi öğretmen eğitimi programlarında açık bir bileşen olmalıdır (DeLuca & Klinger, 2010).

Öğretmen adaylarının mesleğe iyi yetişmiş olarak başlamaları için öğretmen eğitimindeki eksikliklerin tespit edilmesi ve bu eksikliklerin giderilmesine yönelik adımlar atılması gerekmektedir. Öğretmen adaylarının değerlendirme okuryazarlıklarının belirlenmesi, eğitim eksikliklerinin daha mezun olmadan giderilmesi açısından oldukça önemlidir (Roslan, Nisho & Jawawi, 2022). Türkiye'de öğretmen adaylarının değerlendirme bilgi ve becerilerini geliştirmek amacıyla lisans öğretmenliği programlarında bir yarıyılta iki kredilik ders bulunmaktadır. Kurs haftada iki saat olarak verilmektedir. Bu dersin içeriğinde değerlendirmenin önemi, alana özgü temel teorik bilgiler (güvenilirlik ve geçerlilik), istatistiksel işlemler (test ve madde analizi), değerlendirme araçlarının türleri ve özellikleri, değerlendirme, geri bildirim ve puanlama yer alır. Ancak bu dersin kredisinin düşük olduğu, ders içeriğinin sınıf içi değerlendirme uygulamaları için oldukça yetersiz olduğu ve değerlendirme deneyimini uygulama imkânının sınırlı olduğu bilinmektedir. Öğretmenlerin lisans eğitimlerini tamamladıktan sonra eğitimsel değerlendirmeyi öğrenmeleri zorunlu değildir (Fan, Wang & Wang, 2011). Bu durum, öğretmen yetiştirme eğitiminde ölçme-değerlendirme derslerinin önemini artırmaktadır.

Birçok araştırma, öğretmen adaylarının öğrenmeyi değerlendirme, ölçme araçları hazırlama, not verme ve değerlendirme konusunda kendilerini yetersiz hissettiklerini göstermektedir (Roslan, Nisho & Jawawi, 2022; Çelik & Arslan, 2012; Kilmen vd., 2007). Farklı branşlardaki öğretmen adaylarının okuryazarlık düzeylerini belirlemek için yapılan araştırmalarda öğretmen adaylarının okuryazarlık düzeylerinin düşük olduğu görülmüştür (Gül, 2011; Koç, 2019; Berk, 2021).

Bunların yanı sıra matematik öğretmen adaylarının ve hizmet içi matematik öğretmenlerinin değerlendirme alanındaki bilgi, beceri ve görüşlerine yönelik çok sayıda araştırma yapılmıştır. Bu çalışmalar değerlendirme okuryazarlığı (Ergül, 2019), değerlendirmeye yönelik tutum, görüş veya algılar (Göktaş & Şad, 2021; Karakuş, 2010), değerlendirme yeterliğine yönelik algılar (Dokumacı Sütçü & Bulut, 2016) ve değerlendirme uygulamaları üzerinedir (Salmon, 1996; Suziki, 1998; Önel vd., 2020; Türnüklü, 2003; Güven Akdeniz, 2021). Birçok matematik öğretmeni eğitimcisi, kısa süreli bir yöntem dersinde gerekli tüm bilgi ve becerilerin bulunmasının mümkün olmadığı konusunda hemfikirlerdir (Lee & Son, 2015). Ancak ders içeriği, alanın temel

ilkelerine veya standartlarına göre işlendiğinde, öğretmen adaylarının güçlü ve zayıf yönlerini belirlemek mümkün hale gelmektedir.

1.1. Öğrencilerin Eğitsel Değerlendirmesinde Öğretmen Yeterliliği İçin Standartlar

Amerikan Öğretmenler Federasyonu (AFT), Ulusal Eğitimde Ölçme Konseyi (NCME) ve Ulusal Eğitim Derneği (NEA), 1990'da öğrencilerin eğitsel değerlendirmesinde öğretmen yeterliliği için standartlar oluşturmuştur (Plake, 1993). Standartlar, öğretmenlere profesyonel rolleri ve sorumlulukları tanımlama ve uygun değerlendirme sürecini takip etme konusunda rehberlik eder. Standartlar şu şekilde özetlenebilir (AFT, NCME & NEA,1990):

Standart 1. *Öğretmenler, öğretim kararlarına uygun değerlendirme yöntemlerini seçme konusunda yetkin olmalıdır.* Öğretmenlerin, değerlendirme yöntemlerinin çeşitlerini güçlü ve zayıf yönleriyle ayrıntılı olarak bilmeleri gerekir. İlk adımda uygun, faydalı, teknik olarak yeterli ve adil değerlendirme yöntemlerinin seçilmesi esastır. Ders planları ve ders içerikleri ışığında değerlendirme yöntemlerini seçme ve değerlendirme prosedürlerine aşina olmalıdırlar. Bu standardın belirttiği bilgi ve becerilere sahip öğretmenler, uygun değerlendirme yöntemlerinin öğretim hedeflerine ulaşılmasına yardımcı olacak veriler sağlayacağını bilirler.

Standart 2. *Öğretmenler, öğretim kararları için uygun değerlendirme yöntemleri geliştirme konusunda yetenekli olmalıdır.* Öğretmenler sınıfta değerlendirme yaparken genellikle yayımlanmış veya harici değerlendirme araçları yerine kendi uygulama veya tasarım yaklaşımlarını kullanırlar. Bu standardı sağlayan öğretmenler, teorik bilgilere göre tasarlama, uygulama, bilgi toplama ve karar verme becerilerine sahiptir.

Standart 3. *Öğretmen hem harici hem de kendi hazırladığı değerlendirme yöntemlerinin uygulama, puanlama ve sonuçlarını yorumlama konusunda yetenekli olmalıdır.* Öğretmenlerin iyi değerlendirme yöntemlerini seçip geliştirebilmeleri kadar, doğru uygulayabilmeleri, puanlayabilmeleri, yönetebilmeleri, yorumlayabilmeleri de önemlidir. Öğretmenler yaygın olarak bildirilen puanları ve istatistiksel terimleri kavramsal olarak anlamalı ve yorumlayabilmelidir: yüzdelik sıralar, yüzdelik aralık puanları, standart puanlar, sınıf eşdeğerleri, merkezi eğilim ölçüleri, dağılım, ilişkiler, güvenilirlik ve ölçme hataları.

Standart 4. *Öğretmenler, öğrenciler hakkında bireysel kararlar verirken, öğretimi planlarken, müfredat geliştirirken ve okul iyileştirmesi yaparken değerlendirme sonuçlarını kullanma becerisine sahip olmalıdır.* Öğretmenler, sonuçları kullanarak eğitimsel kararları vermek için kavramsal ve uygulama becerilerine sahip olmalıdır. Sonuçları müfredat geçerliliğine göre yorumlamalı ve yanlış yorumlamalara yol açabilecek hatalardan kaçınmalıdır. Elde edilen değerlendirme bilgileri, öğrencilerin eğitimsel gelişimine yardımcı olmak için öğretim planını düzenlemek için kullanılmalıdır.

Standart 5. *Öğretmenler, öğrenci değerlendirmelerini kullanan geçerli öğrenci derecelendirme prosedürleri geliştirme konusunda yetenekli olmalıdır.* Geçerli not verme, öğretmenlerin değerlendirme becerilerinin önemli bir kanıtıdır. Not verme hem öğrencinin performans düzeyini hem de öğretmenin bu performansa verdiği değeri gösterir. Geçerli not almanın ilkeleri, öğretmenler tarafından akılcı ve adil bir şekilde bilinmeli ve uygulanmalıdır. Öğretmenler, notları ceza olarak kullanmak gibi yanlış not verme prosedürlerinden kaçınmalı ve öğrencilerin kazanımları hakkındaki yorumlarını kanıtlara dayalı olarak zenginleştirmelidir.

Standart 6. *Öğretmenler, değerlendirme sonuçlarını öğrencilere, velilere, ilgili kişilere ve diğer eğitimcilere iletme konusunda yetenekli olmalıdır.* Öğretmenler, değerlendirme sonuçlarını yalnızca kendi eğitim kararları için kullanmakla kalmaz, aynı zamanda öğrencilere, velilere, diğer eğitimcilere ve farklı izleyicilere rapor verir. Değerlendirme sonuçları etkili bir şekilde iletilmezse, faydalı olarak kullanılamaz veya hiç kullanılamaz. Öğretmenler, öğrenci değerlendirmesi hakkında başkalarıyla etkili bir şekilde iletişim kurmak için, değerlendirme terimlerini anlaşılır bir şekilde kullanabilmeli ve değerlendirme sonuçlarının anlamını, sınırlıklarını ve sonuçlarını açıklayabilmelidir. Öğretmenler, değerlendirme prosedürlerini kanıtlarla doğrulamak zorunda kalabilirler. Bazen, halkın değerlendirme sonuçlarını doğru bir şekilde yorumlamasına da yardımcı olabilirler.

Standart 7. *Öğretmenler, etik dışı, yasa dışı ve diğer açılardan uygun olmayan değerlendirme yöntemlerini ve değerlendirme bilgilerinin kullanımını tanıma konusunda yetenekli olmalıdır.* Öğretmenler, değerlendirmede etik ve yasal sorumluluklar konusunda bilgili olmalıdır. Başkalarının uygun olmayan değerlendirme uygulamalarıyla karşılaşılırsa, onların vazgeçirmeye çalışmalıdırlar. Öğretmenler, ölçme-değerlendirme yöntemlerinin uygulanması, sonuçların yorumlanması, kullanılması ve iletilmesi sırasında mesleğin etik kuralları

doğrultusunda ilgili kişilerin haklarını koruyan adil tutum ve davranışlarda bulunmalıdır. Öğrencilerini, okullarını, ilçelerini etkileyen kararları ve kanunları da bilmelidirler. Öğretmenler, çeşitli değerlendirme prosedürlerinin, öğrencilerin mahremiyet hakkını ihlal etmek, onları utandırmak veya puanlarını uygunsuz şekilde kullanmak gibi zararlı sonuçlarla kötüye kullanılabileceğinin farkında olmalıdır.

1.2. Standartlara Göre Düzenlenen Değerlendirme Okuryazarlığı Ölçekleri

Yukarıda belirtilen yedi standardı göz önünde bulunduran Plake ve Impara, öğretmenlerin değerlendirme okuryazarlığını ölçmek için 35 maddeden oluşan “Öğretmen Değerlendirme Okuryazarlığı Anketi (ÖDOA)” adlı bir araç geliştirdi. Anket, dört olası cevap ve bir doğru cevap içeren çoktan seçmeli 35 maddeden oluşmaktadır. Anket, Amerika Birleşik Devletleri’ndeki 555 hizmet içi öğretmene uygulandı. Testin doğru cevap ortalaması 35 sorudan 23’ tür. Öğretmenlerin, öğrencilerin öğrenmesini değerlendirmek için yeterince yetkin olmayabileceğini göstermektedir (Plake vd., 1993). Ayrıca Campbell, Murphy ve Holt (2002) bir değerlendirme kursunu tamamlayan 220 öğretmen adayına ÖDOA uygulamıştır. Benzer şekilde, doğru cevapların 35 maddeden ortalama 21 olduğunu göstermiştir. Mertler (2004), ÖDOA’ yı bir grup hizmet içi ve hizmet öncesi öğretmenin demografik durumuna uyacak şekilde sınırlı miktarda düzenleyerek kullandı. Diğer araştırmalara benzer şekilde 35 maddeden ortalama 22 doğru bulunmuştur.

Değerlendirme okuryazarlığının orijinal ölçeği olarak tanımlanabilecek ÖDOA, AFT&NCME&NAE (1990) standartlarına göre düzenlenen ilk ölçektir (Gürsoy, 2017). Bu ölçekten yenilenen veya esinlenen Sınıf Değerlendirme Okuryazarlığı Envanteri (SDOE) (Mertler, 2003) ve Değerlendirme Okuryazarlığı Envanteri (DOE) (Mertler ve Campbell, 2005) 35 çoktan seçmeli madde içerir ve 1990 Standartlarını temel alır. ÖDOA’ dan sonra oluşturulan değerlendirme okuryazarlığı ölçen araçlar genellikle standartlara dayalı olarak tasarlanmaya başlamıştır (Ball, 2017).

Türkiye’de Gül (2011), bazı değişiklikler yaptıktan sonra ÖDOA’ nın yeni sürümünü yüksek lisans tezinde veri toplama araçlarından biri olarak uygulamıştır. Fen Bilimleri, Sosyal Bilgiler, Matematik ve Türkçe Öğretmenliği bölümlerinde öğrenim gören 180 öğretmen adayının değerlendirme okuryazarlığını belirlemek amacıyla uygulanan envanter sonucunda 34 soruda ortalama 18 maddenin doğru cevaplandığı görülmüştür. Araştırma sonucunda öğretmen adaylarının ölçeğe ve değerlendirme konusunda büyük eksiklikleri olduğu tespit edilmiştir.

Hizmet öncesi ya da içi öğretmenlerin değerlendirme okuryazarlıklarını inceleyen çalışmaların çoğunda ölçeklerin çoktan seçmeli olarak hazırlandığı ve verilerin nicel olarak incelendiği görülmektedir (Azrak & Yalçınkaya, 2019; Ergül, 2019; Koç, 2019; Yamtim & Wongwanich, 2014; Karaman & Şahin, 2014, Mertler, 2004). Yapılan bu çalışmaların okuryazarlık düzeylerinin ortaya konulması ve eksikliklerin tespit edilmesi açısından önemi büyüktür. Ancak nicel araştırmaların doğası gereği elde edilen verilerle bu sonuçların ötesinde derinlemesine yorum yapabilmek mümkün değildir. Bu çalışmada alanda çok kullanılmış bir ölçek olan ÖDOA’ nın Türkçe versiyonunun uzman görüşü olarak açık uçlu hale getirilmesi ve matematik eğitimine uyarlanmasıyla zengin veri içeriğine ulaşmak ve yapılan çalışmalara farklı bir yönden katkı sağlamak amaçlanmıştır.

1.4. Öğretim Deneyi

Bu çalışma, araştırmacıların doktora tezi kapsamında yapılan öğretim deneyinden elde edilen verilerin analiz edilmesiyle ortaya çıkarılmıştır. Öğretim deneyi sınıfta klinik metodoloji ile öğretim gerçekleştirirken elde edilen verilerin analiziyle öğretimin yeniden tasarlanması ve döngüsel olarak bu temelde verilerin toplanması ve analiz edilmesi şeklinde ilerleyen bir araştırma yöntemidir (Czarnocha & Prabhu, 2006). Öğretim deneyi araştırmacının varsayımlarına dayalı bir öğretim sürecini tasarlaması ve planlanması ile başlar. Öğretim bölümleri olarak adlandırılan yapılandırmacı yaklaşım çerçevesinde sınıf içi uygulamalar yaparak öğretim gerçekleştirilir (Akın & Kabael, 2016). Hazırlanan eğitimsel araçlar vasıtasıyla öğrencilere bilgi ve becerilerini geliştirme imkanı sunulmaktadır.

Bu bağlamda, literatür araştırılarak bu araştırma için matematik öğretmen adaylarının ölçeğe değerlendirme alanında ihtiyaç duyacağı bilgi ve becerileri içeren bir ders programı oluşturulmuştur. Öğretim deneyinin matematik öğretmen adaylarının değerlendirme okuryazarlığına etkisini incelemek için hazırlanan açık uçlu envanter ön-test ve son-test olarak uygulanmıştır. Matematik öğretmeni adaylarının, öğretmen yeterliliği standartlarına dayalı değerlendirme okuryazarlığı üzerine bir çalışma, öğretmen adaylarına sağlanacak gelişimsel yaklaşımları belirlemek için kullanılabilecek önemli veriler sunacaktır.

1.5. Araştırma Problemi

Bu çalışmada “değerlendirme okuryazarlığı” kavramı incelenmiş ve öğrencilerin Eğitimsel Değerlendirmesinde Öğretmen Yeterliliği Standartlarına göre matematik öğretmeni adaylarının değerlendirme okuryazarlık düzeylerinin belirlenmesi hedeflenmiştir. Bu çalışmanın amacı, matematik öğretmeni adaylarının

değerlendirme okuryazarlık düzeylerinin standartlara göre öğretim deneyi sonrasındaki değişimini ortaya koymaktır.

2. Yöntem

2.1. Çalışma Grubu

Katılımcılar, Türkiye'nin kuzeyindeki bir devlet üniversitesinde 2019-2020 eğitim-öğretim yılında matematik öğretmenliği lisans programının son sınıfında öğrenim gören sekiz öğretmen adaydır. Katılımcılar, amaçlı örnekleme yöntemlerinden biri olan ölçüt örnekleme yöntemi ile belirlenmiştir. Ölçütler önceki yıl değerlendirme dersini başarıyla geçmiş olmaları ve araştırma esnasında stajlarına devam ediyor olmaları olarak belirlenmiştir. Çalışma grubu, bu ölçütleri sağlayan matematik öğretmenliğinde okuyan sekiz son sınıf öğrencisinden oluşmaktadır.

Bu öğretmen adayları geçen yıl değerlendirme dersini almış ve başarılı olmuşlardır. Araştırmanın staj deneyiminde yürütüldüğü dönemde, deneyimli öğretmenleri gözlemleme ve sınıfta öğrendiklerini uygulama fırsatı buldular. Öğretmen adaylarının staj deneyimlerinin ölçme alanına ilişkin farkındalıklarını artırması, böylece çalışmamızın daha derin ve zengin bir içeriğe ulaşma olasılığını güçlendirmesi beklenmektedir.

2.2. Süreç

Araştırmacılar öğretmen adaylarına değerlendirme okuryazarlığı ile ilgili bir çalışma olduğu bilgisini vermiştir. Toplanacak verilerin ders notlarını hiçbir şekilde etkilemeyeceği vurgulanmış ve araştırma sonuçlarının gelecekte öğretmen adayları için daha iyi değerlendirme öğretimine olanak sağlayacağını umulduğu belirtilmiştir.

Tüm katılımcılara araştırma protokolü hakkında bilgi verilmiş ve onam formları gönderilmiştir. Bireysel bulgularının hiçbir şekilde kendi kimlikleri ile raporlanmayacağı konusunda bilgilendirilmişlerdir. Öğretmen adayları araştırma ile ilgili bilgileri içeren onam formlarını imzalayarak araştırmacıya postalamışlardır.

Öğretim deneyinin ders içeriği matematik eğitiminde ölçme değerlendirme bilgi ve becerilerini içeren şekilde oluşturulmuştur. Matematik eğitimi ve ölçme değerlendirme alan uzmanlarının görüşleri alınarak içeriklerde gerekli düzenlemeler yapılmıştır. 15 öğretim bölümünden oluşan programın öğretim deneyi doğasına uygun şekilde yapılandırmacı bir yaklaşıma uygun tasarlanmıştır.

COVID-19 salgını, eğitim durumunu yüz yüze eğitimden çevrimiçi eğitime dönüştürmüştür. Bu nedenle yüz yüze başlayan araştırma çevrimiçi olarak devam etmek durumunda kalınmıştır. Bu anlamda ders planları çevrimiçi eğitime uygun hale getirilmiştir. Dersler canlı çevrimiçi görüntülü bir iletişim programı üzerinden yapılmış ve kayıt altına alınmıştır. Öğretim deneyi eğitimleri 16 hafta devam etmiştir. Bu esnada araştırmacılar, araştırma süresince öğretmen adaylarının staj süreçlerinin olağan seyrinde herhangi bir olumsuzlukla karşılaşmadığını teyit etmişlerdir. Staj deneyimiyle öğretmen adaylarının öğretim deneyinde öğrendikleri ile deneyimli bir matematik öğretmenin uygulamalarını incelemelerine imkan sağlayacağı düşünülmüştür.

Bu araştırmada veri toplama aracı olarak kullanılan açık uçlu sorulardan oluşan Matematik Eğitiminde Değerlendirme Okuryazarlığı Envanteri canlı çevrimiçi sınav şeklinde kağıt kalemle uygulanmış ve süre sonunda öğretmen adaylarından cevap kağıtlarının fotoğrafları istenmiştir.

2.3. Veri Toplama Aracı

Sınıfta Değerlendirme Okuryazarlığı Envanteri, Gül (2011) tarafından yüksek lisans tezinde kullanılmak üzere Barbara S. Plake ve James C. Impara (1993) tarafından hazırlanan benzer bir araç olan 'Öğretmen Değerlendirme Okuryazarlığı Anketi'nden Türkiye bağlamına uyacak şekilde değiştirilmiştir. Envanter, öğrencilerin eğitimsel değerlendirmesinde öğretmen yeterliliği için standartların her biri için çoktan seçmeli beş madde içermektedir. Geçerlik, güvenirlik ve pilot uygulama çalışmaları yapıldıktan sonra Gül (2011) envanteri 180 öğretmen adayına uygulamıştır.

Matematik Eğitiminde Değerlendirme Okuryazarlığı Envanteri, Gül (2011) tarafından düzenlenen çoktan seçmeli Sınıfta Değerlendirme Okuryazarlığı Envanteri'nden matematik eğitime uyarlanmış açık uçlu bir veri toplama aracıdır. Bazı sorular açık uçlu hale getirilemediği için 35 sorudan 6 tanesi elenmiştir. Envanter, matematik eğitimi ve değerlendirmesi konusunda uzman iki akademisyen tarafından geçerlik açısından incelenmiştir. İnceleme sonuçlarına göre gerekli düzeltmeler yapılmıştır.

Sınıfta değerlendirme okuryazarlığı envanteri (Gül, 2011) sorularının standartlara göre sınıflandırılması göz önünde bulundurularak elenen soruların çıkarılmasıyla birlikte, matematik eğitime uyarlanan açık uçlu soruların hangi yeterlik standartlarına uygun olduğu Tablo 1' de belirtilmiştir.

Öğretmen Değerlendirme Okuryazarlığı Anketini araştırmasında kullanan Yamtim ve Wongwanich (2014), elde edilen sonuçları öğretmen adaylarının sınıf değerlendirme okuryazarlık düzeylerinin anlaşılabilmesi için

düşük (%60' tan az), orta düzeyde (%60-%79) ve iyi düzeyde (%80 ve üzeri) olarak üç bölüme ayırmışlardır. Bu nedenle envanterin toplam 100 puan üzerinden değerlendirilebilmesi için yedi standarda eşit puan dağılımı yapılması uygun bulunmuştur. Altı ve yedinci standartlarda soru sayısının az olmasından dolayı artan 2 puan bu standartlara dağıtılmıştır. Standartların toplam puanı içindeki maddelere eşit olarak dağılımı yapılmaya çalışılmış, artan puan standart içindeki maddelerden araştırmacılar tarafından uygun bulunanlar arasında paylaştırılmıştır. Buna göre matematik öğretmen adaylarının ön-test ve son-test puanları belirlenmiştir.

Tablo 1. Envanter Sorularının Standartlara Göre Sınıflandırılması

Öğrencinin Eğitimsel Değerlendirmesinde Öğretmen Yeterliliği Standartları	Madde No ve Puanlamaları	Toplam Puan
1. Uygun değerlendirme yöntemlerini seçme	1 (3p), 2 (3p), 3 (4p), 4 (4p)	14
2. Uygun değerlendirme yöntemleri geliştirme	5 (3p), 6 (2p), 7 (3p), 8 (3p), 9 (3p)	14
3. Öğrenme çıktılarını yönetme, puanlama ve yorumlama	10 (3p), 11 (3p), 12 (3p), 13 (3p), 14 (2p)	14
4. Karar vermek için değerlendirme sonuçlarını kullanma	15 (2p), 16 (3p), 17 (3p), 18 (3p), 19 (3p)	14
5. Geçerli derecelendirme prosedürlerinin geliştirme	20 (4p), 21 (3p), 22 (3p), 23 (4p)	14
6. Değerlendirme sonuçlarını iletme	24 (5p), 25 (5p), 26 (5p)	15
7. Etik olmayan uygulamaları bilme	27 (5p), 28 (5p), 29 (5p)	15

2.4. Veri Analizi

Araştırmadan elde edilen nitel ve nicel veriler ayrı ayrı değerlendirilmiştir. Nicel verileri elde etmek için açık uçlu sorulardan oluşan envanterin cevapları için dereceli puanlama anahtarı oluşturulmuştur. Her bir öğretmen adayının cevapları dereceli puanlama anahtarına göre iki araştırmacı tarafından ayrı ayrı puanlanmıştır. Birden çok puanlayıcının verdiği puanlar arasındaki uyumun belirlenmesi için puanlayıcılar arası güvenilirlik hesaplaması yapılmaktadır (Bıkmaz Bilgen & Doğan, 2017). Krippendorff Alfa katsayısı hesaplaması $\alpha = 1 - \frac{D_o}{D_e}$ formülü ile yapılmaktadır. D_o (gözlenen uyumsuzluk) ve D_e (beklenen uyumsuzluk) değerlendiricilerin puanlamalarından oluşturulan veri matrisinden yararlanılarak formüllerle hesaplanmaktadır (Kanık vd., 2010). Bu hesaplamalara göre puanlayıcılar arası güvenilirlik 0,81 olarak hesaplanmıştır. Alfa değerinin 0,80 ve üzerinde olması uyumun yüksek olduğunu belirtmektedir (Arslan Mancar, 2019). Puanlamadaki farklılıklar üzerinde tartışılarak görüş birliğine varılarak ön-test ve son-test puanları belirlenmiştir. Elde edilen nicel veriler, frekans dağılımı ve ortalama açısından betimsel istatistikler aracılığıyla analiz edilmiştir.

Nitel araştırmalar, insanların düşünce, görüş ve tutum gibi niteliklerini ayrıntılı bir şekilde yorumlamayı ve betimlemeyi amaçlamaktadır (Kıral, 2020). Nitel araştırmalarda kullanılan veri toplama yöntemlerinden birisi de doküman analizidir. Eğitim alanında yapılan nitel araştırmalarda yapılan bir çalışmanın etkisi araştırmacının belirli amaçlar doğrultusunda detaylı bilgiye ulaşmak için dokümanlar üretilebilmekte ve katılımcılardan doküman içeriğini oluşturmaları istenebilir (Sak vd., 2021). Bu nedenle, öğretmen adayının envanter cevapları genel kategorilere ayrılmış ve her bir tema ile ilgili ayrıntılı analizlere başlanmıştır. Bu analizlerde verilerin nicelleştirilerek sayısal analiz yapmaktan öte araştırılan konuyla ilgili çıkarımlarda bulunabilecek biçimde verilerin işlenmesi tercih edilmiştir.

3. Bulgular ve Tartışma

3.1. Envanter Sonuçlarının Nicel Olarak İncelenmesi

Envanterin ön-test cevaplarının incelenmesiyle matematik öğretmeni adaylarının değerlendirme okuryazarlığı toplam puanlarının 100 puan üzerinden en az 47 puan, en yüksek 62 puan olduğu ve toplam puanların ortalamasının 55,75 olduğunu ortaya çıkmıştır. Öğrencilerin eğitimsel değerlendirmesinde öğretmen yeterliliği için yedi standarda göre ön-test puanları incelendiğinde, Tablo 2' de görüldüğü gibi ortalama puanların 4.125 ile 13.625 puan arasında değiştiği görülmüştür. En yüksek ortalama puanı alan Standart 7: Etik olmayan uygulamaları bilme (Ortalama = 13.625) ve en düşük puan ortalamasına sahip Standart 2: Uygun değerlendirme yöntemlerinin geliştirme (Ortalama = 4.125) olarak görülmektedir.

Envanterin son-test cevaplarının puanlamasında matematik öğretmen adayları 100 puan üzerinden en az 49 puan, en yüksek 73 puan almışlardır. Toplam puanının ortalama 61,375 puan olduğunu hesaplanmıştır. Her bir standarda göre son-test puanları incelendiğinde, Tablo 2' de görüldüğü gibi ortalama puanların 3.625 ile 13 puan arasında değiştiği görülmüştür. En yüksek ortalama puanı alan Standart 7: Etik olmayan uygulamaları bilme (Ortalama = 13) ve en düşük puan ortalamasına sahip Standart 6: Değerlendirme sonuçlarını iletme (Ortalama = 3.625) olarak görülmektedir.

Bu araştırmada öğretmen adaylarının çoğunun öğretim deneyi öncesinde düşük düzeyde sınıf değerlendirme okuryazarlığı puanlarına sahip oldukları görülmüştür. Bu sonuç öğretmen adaylarının ölçme değerlendirme bilgi

ve becerileri üzerine çalışmalar yapan Berk (2021), Koç ve Bulut (2020), Koç (2019), İzci ve ark. (2018), Gül (2011) ve Mertler (2004)' in araştırma sonuçları ile benzerlik göstermektedir.

Ergül ve Çetin' in (2021) 189 ortaokul öğretmen adayıyla değerlendirme okuryazarlığı seviyelerini yeterlilikler standartlarına göre belirlemek için yaptıkları çalışmada sordukları 30 sorunun ortalama doğru yanıt sayısını 13 (%43) olarak bulmuşlardır. Karaman ve Şahin' in (2014) bu çalışmada uygulanan envanterin orijinalinden esinlenilerek oluşturulan Ölçme-değerlendirme Okuryazarlık Envanterinin 289 öğretmen adayına uygulanması sonucunda öğretmen adaylarının aldıkları ortalama puanlar (%53) olarak bulunmuştur. Bu çalışmanın ön-test puan ortalaması (%55,75) ve son-test puan ortalaması (%61,375) Ergül ve Çetin' in (2021) ve Karaman ve Şahin (2014) çalışmasındaki ortalamalardan biraz yüksek bulunmuştur. Değerlendirme okuryazarlığına yönelik verilen eğitimin öğretmen adaylarının gelişimine katkıda bulunduğu söylenebilir.

Tablo 2. Ön-test ve Son-test Envanter Puanlarının Karşılaştırması

Örneklem sayısı n=8	Minimum puan		Maksimum puan		Ortalama		Standart sapma		Toplam Puan
	Ön- test	Son- test	Ön- test	Son- test	Ön- test	Son- test	Ön- test	Son- test	
1.Uygun değerlendirme yöntemlerini seçme	4	4	11	11	7,625	8,125	2,774	2,357	14
2.Uygun değerlendirme yöntemleri geliştirme	1	4	7	9	4,125	8	2,1	3,251	14
3.Öğrenme çıktılarını yönetme, puanlama ve yorumlama	5	5	7	11	6,125	8,375	0,991	2,722	14
4.Karar vermek için değerlendirme sonuçlarını kullanma	8	8	11	14	8,875	11	1,246	1,69	14
5.Geçerli derecelendirme prosedürlerinin geliştirme	5	6	11	13	8,375	9,25	1,847	2,659	14
6.Değerlendirme sonuçlarını iletme	5	0	8	8	7	3,625	1,414	2,722	15
7.Etik olmayan uygulamaları bilme	9	8	15	15	13,625	13	2,2	2,673	15

Bu çalışmada kullanılan envanterin orijinalinden esinlenilerek yapılan ölçek ile yurt dışında gerçekleştirilen çalışmalarda öğretmen adaylarının doğru cevap ortalamalarını Mertler ve Campbell (2005) %68, Mertler (2003) %54, Campbell ve arkadaşları (2002) %60 olarak bulmuşlardır. Gerek yurt içi gerekse yurt dışında öğretmen adayları ile değerlendirme okuryazarlığı üzerine gerçekleştirilen çalışmalarda öğretmen adaylarının düşük ve orta seviyede bulundukları görülmektedir. Bu durumun çoğunlukla öğretmen yetiştirme eğitimlerinde bu alana yeteri kadar önem verilmemesinden kaynaklandığı düşünülmektedir (Siegel& Wissehr, 2011).

Çalışmamızda kullandığımız envanterin kaynağı olan Öğretmen Değerlendirme Okuryazarlığı Anketini çalışmasında kullanan Yamtim ve Wongwanich (2014), öğretmen adaylarına uyguladıkları ölçeğin sonuçlarını değerlendirebilmek için elde edilen puanları düşük (%60' tan az), orta düzeyde (%60-%79) ve yüksek düzeyde (%80 ve üzeri) olarak üç bölüme ayırmışlardır. Benzer şekilde ön-test ve son-test sonuçlarının 100 puan üzerinden değerlendirilmesi sonucunda her bir standarda göre sekiz öğretmen adayının hangi düzeyde oldukları Tablo 3' te belirtilmektedir.

Tablo 3. Öğretmen Adaylarının Standartlara Göre Puan Seviyesi (n=8)

Standartlar	Düşük		Orta		Yüksek	
	Ön-test	Son- Test	Ön-test	Son -Test	Ön-test	Son -Test
1. Uygun değerlendirme yöntemlerini seçme	5	5	3	3	-	-
2. Uygun değerlendirme yöntemleri geliştirme	8	5	-	1	-	2
3. Öğrenme çıktılarını yönetme, puanlama ve yorumlama	8	5	-	2	-	1
4. Karar vermek için değerlendirme sonuçlarını kullanma	5	1	3	6	-	1
5. Geçerli derecelendirme prosedürlerinin geliştirme	4	4	4	2	-	2
6. Değerlendirme sonuçlarını iletme	8	8	-	-	-	-
7. Etik olmayan uygulamaları bilme	-	1	1	1	7	6
Genel Değerlendirme	5	3	3	5	-	-

Matematik eğitiminde ölçme değerlendirme alanı için tasarlanan ve öğretmen adaylarıyla uygulanan öğretim deneyi sonunda yapılan son-test puanlarına göre her bir standarda göre öğretmen adaylarının bulundukları düzey Tablo 3' te görülmektedir. Elde edilen verilere göre öğretmen adaylarının çoğunun yedi standarttan dördünde düşük düzeyde olduğu görülmektedir. Bu standartlar Standart 1: uygun değerlendirme yöntemlerini seçme (5/8), Standart 2: uygun değerlendirme yöntemleri geliştirme (5/8), Standart 3: öğrenme çıktıları uygulamaya, puanlama ve yorumlama (5/8), Standart 6: Değerlendirme sonuçlarını iletme (8/8) olarak görülmektedir. Son-test puanlarına göre öğretmen adaylarının Standart 4: Karar vermek için değerlendirme sonuçlarını kullanma (6/8) ve Standart 5: Geçerli derecelendirme prosedürlerinin geliştirme (2/8) konularında orta düzeyde oldukları görülmektedir. Öğretmen adayları ön-test puanlarında olduğu gibi Standart 7: Etik olmayan uygulamaları bilme (6/8) konusunda üst düzeyde bulundukları görülmektedir.

Tablo 3 incelendiğinde 2., 3., 4. ve 5. Standartlarda öğretmen adaylarının düşük seviyeden orta ve yüksek seviyelere kısmen geçiş yaptığı görülebilir. Envanter sonuçlarına göre öğretim deneyinin 1., 6. ve 7. standartlar için etkisinden söz edilememektedir. Benzer şekilde Beziat ve Coleman'ın (2015) öğretmen eğitimi sırasında verilen değerlendirme eğitiminin yönünü belirlemek için ön test ve son test olarak kullanılan ÖDOA sonuçlarında bir fark bulunmamıştır.

3.2. Envanter Sonuçlarının Nitel Olarak İncelenmesi

Envanterdeki 29 açık uçlu sorunun yedi standarda dağılımı Tablo 1' de verilmiştir. Sekiz öğretmen adayının standartlara ilişkin sorulara verdiği yanıtlar kodlanarak alt kategoriler ve kategoriler oluşturulmuştur. Öğretmen adaylarının verdiği cevaplar ilgili kategorilerde frekanslarla ifade edilmiştir. Yanıtsız bırakılan sorulara frekans verilmemiştir. Bunun yanı sıra, araştırmacılar verilen bazı yanıtları doğrudan alıntılar yaparak okuyuculara sunmaktadır.

Standart 1. *Öğretmenler, öğretim kararlarına uygun değerlendirme yöntemlerini seçme konusunda yetkin olmalıdır.* Bu standarttaki yeterlikleri belirlemek açısından öğretmen adaylarına yöneltilen güvenirliliğin sağlanması, geçerli ölçme değerlendirme seçimi, akademik başarının ve motivasyonun belirlenmesi kapsamında sorular yöneltilmiştir.

Öğretmen adaylarına sınıf içi ölçme değerlendirme uygulamalarını seçmeye yönelik en çok önem vereceği hususlar hakkında sorulan sorulara verdikleri yanıtların frekans dağılımları Tablo 4' te verilmiştir. Tablo 4' te görüldüğü gibi öğretmen adaylarının geçerli ölçme sonuçlarına ulaşmak için yapılması gerekenler hakkındaki düşünceleri ön-testte ve son-testte farklılıklar içermektedir. Ölçme aracının kazanımlara uygun olması gerektiğini ön-testte sekiz öğretmen adayından 3' ü, son-testte 6' sı ifade etmiştir. Maddelerin bilişsel taksonomilere göre sınıflandırılmasını sadece ön-testte 3 öğretmen adayı belirtmiştir. Öğrenme eksikliklerini belirlemek üzere madde seçimini ön-testte 5, son-testte 4 öğretmen adayı ifade etmiştir. Ölçme aracının pedagojik uygunluğu ön-testte 2, son-testte 3 öğretmen adayı tarafından belirtilmiştir. Konuya uygun ölçme aracı seçimini ön-testte ve son-testte birer öğretmen adayı ifade etmiştir. Öğrencinin ilgisini çeken maddeler oluşturmayı ön-testte 3 öğretmen adayı, son-testte 1 öğretmen adayı ifade etmiştir. Ölçme aracı hazırlanırken dil bilgisine ve görsel unsurlara önem verilmesinin ölçme sonuçlarının geçerliğini artıracaklarını sadece ön-testte 2 öğretmen adayı belirtmiştir.

Öğrencilerin öğretilen bir konuyla ilgili başarılarını ölçen bir araçtan elde edilen sonuçlara dayalı yorumların geçerliğinden bahsetmek, testin geçerliğinden bahsetmekten daha doğru bir yaklaşımdır. Bu nedenle geçerlik başta testin ölçmek istediği davranışı hedeflemesine, uygulanan gruba ve şartlara göre değişebilir (Demircioğlu, 2015). Geçerlik büyük oranda test hazırlanırken dikkat edilen bazı hususlara bağlıdır. Bunlar testin ölçmek istediği konuları ve öğrenme basamaklarını kapsayan yeterli madde sayısının olması, uygun ölçme aracının seçimi, maddelerin açık ve anlaşılır yazılması gibi hususlardır. Öğretmen adaylarına öğretilen bir konuyla ilgili ölçme ve değerlendirme yöntemini belirlerken en çok neye veya nelere dikkat ettikleri sorulduğunda ön-testte ölçme aracının kazanımlara uygunluğundan sekiz öğretmen adayından üçünün değindiği görülmüştür. Testin kapsam geçerliğinin sağlanması test hazırlarken dikkat edilecek en önemli hususlardan birisidir. Aşağıda iki farklı öğretmen adayının ölçme aracında en çok dikkat ettiği hususlara ön-testte verdiği yanıtlar verilmiştir.

Tablo 4. Öğretmen Adaylarının Envanterin Ön-test Uygulamasında Standart 1'e Yönelik Hazırlanan Sorulara Verdikleri Yanıtlar

Kategoriler	Alt Kategoriler	Ön-test (f)	Son-test (f)
Geçerli ölçme sonuçlarına ulaşmak için yapılması gereken hazırlıklar	Kazanıma uygunluk	3 (Ö1, Ö3, Ö6)	6 (Ö1, Ö3, Ö4, Ö5, Ö6, Ö8)
	Bilişsel taksonomilere göre sınıflandırma	3 (Ö1, Ö2, Ö4)	-
	Öğrenme eksikliklerini belirleyecek maddeler seçme	5 (Ö1, Ö3, Ö5, Ö7, Ö8)	4 (Ö3, Ö4, Ö5, Ö7)
	Pedagojik uygunluk	2 (Ö6, Ö7)	3 (Ö2, Ö4, Ö7)
	Konunun yapısına uygunluk	1 (Ö4)	1 (Ö7)
	Öğrencilerin ilgisini çekmek	3 (Ö3, Ö4, Ö6)	1 (Ö3)
	Dil bilgisi ve görsel unsurlar	2 (Ö5, Ö7)	-
Teorik özellikler	Güvenirlilik	1 (Ö4)	6 (Ö1, Ö3, Ö4, Ö5, Ö6, Ö8)
	Geçerlik	1 (Ö2)	2 (Ö5, Ö6)
	Tutarlık	1 (Ö5)	2 (Ö1, Ö5)
	Kullanışlılık	2 (Ö4, Ö7)	1 (Ö4)
	Objektiflik	2 (Ö1, Ö8)	2 (Ö1, Ö5)
	Madde gücü	4 (Ö1, Ö3, Ö5, Ö7)	1 (Ö6)
	Madde ayırt ediciliği	-	2 (Ö2, Ö6)
Değerlendirme türü	Hazırbulunusluk	2 (Ö2, Ö6)	-
	Süreç	2 (Ö3, Ö4)	3 (Ö6, Ö7, Ö8)
	Sonuç	2 (Ö3, Ö4)	-
Ölçme aracı türü	Bilişsel	7 (Ö1, Ö2, Ö4, Ö5, Ö6, Ö7, Ö8)	5 (Ö1, Ö2, Ö3, Ö5, Ö7)
	Duyuşsal	3 (Ö4, Ö5, Ö8)	7 (Ö1, Ö2, Ö3, Ö4, Ö5, Ö7, Ö8)

Ö3: “Öncelikle hedef davranışların kazanılıp kazanılmadığını yoklamak amacıyla bir test hazırlarım. Bunu hazırlarken çeşitli maddelerin olmasına özen gösteririm.”

Ö7: “Öğrettiğim konuyu değerlendirmek için çocukların zihinlerinde canlandırılması kolay etkinlikler seçerdim. Yapılandırılmış grid seçerdim. Görüntüler eklerdim konu ile ilgili. Çünkü daha önceden resimlerle anlatmış olurum. Sorduğum şekilde geri dönütler almak isterdim.”

Son-test verilerinde verilen yanıtlarda ölçme aracının hedeflenen niteliklere uygunluğu sekiz öğretmen adayından altısının yanıtlarının içinde olduğu görülmektedir. Öğretmen adaylarının yanıtlarında ölçme aracının kazanımlara, öğrencilerin seviyesine uygun, ekonomik ve kullanışlı olması, yapılandırmacı yaklaşıma uygun tamamlayıcı ölçme yöntemlerinin kullanılması, kapsam geçerliğini sağlaması, ölçmeye hata karışmaması gibi nitelikler bulunmaktadır.

Ö4: “Hedef ve içeriklere uygun olmasına, öğrencilerin düzeylerine ve özelliklerine uygun olmasına, ekonomik ve kullanışlı olmasına dikkat ederim.”

Ö8: “Testin ölçmek istediğim amacına hizmet edip etmemesine dikkat ederim. Soruların şans başarısına sebep olmayacak şekilde hazırlamaya özen gösteririm.”

Tablo 4’ te görüldüğü gibi teorik özellikler kategorisinde öğretmen adaylarının ölçme değerlendirme aracı seçme konusunda verdiği yanıtlar görülmektedir. Ölçme aracında bulunması gereken en temel teorik özellikler olan geçerlik ve güvenirlik kavramlarına ön-testte birer öğretmen adayı değinmiştir. Son-testte altı öğretmen adayının güvenirlik kavramına, iki öğretmen adayının geçerlik kavramına değindiği tespit edilmiştir.

Öğretmen adaylarına öğrencilerin dersleri ve akademik başarıları ile ilgili motive edip etmediğini anlamada ölçme değerlendirme yoluyla edinilebilecek en geçerli kanıtlar sorulduğunda, son-testte verilen cevapların çoğunlukla ‘tutum testleri’ ve ‘anketleri’ içerdiği görülmektedir. Tablo 4’ te görüldüğü gibi ön-testte 8 öğretmen adayından 7’si bilişsel ve 3’ü duyuşsal ölçme araçlarından, son-testte 5’i bilişsel ve 7’ si duyuşsal ölçme araçlarından elde edebileceğini belirtmiştir. Öğretmenlerin öğretim uygulamalarıyla verilmek istenilen bilgi ve becerilerin öğrencilerin kazanıp kazanmadığını gösteren en geçerli kanıtlar farklı ölçme araçlarından elde edilen verilerdir (Bahar vd., 2015). Öğrencilerin matematiğe ilgisini artırmak için güçlü bir motivasyon gerekmektedir (Şimşir Gökalp, 2021). Öğrencilerin duyuşsal özelliklerinin önemsenmesi ders başarısını etkileyecek bir unsurdur. Akademik başarılarıyla ilgili motivasyonun göstergesi olarak bilişsel ve duyuşsal testlerin bir arada değerlendirilmesi gerekmektedir.

Ön-test verilerinde öğretmen adaylarının bilişsel testlere vurgusundan örnek cevaplar;

Ö1: “Çözdüğü sorular bilişsel düzey basamaklarından hangilerini kapsıyor. Süreç ve sonuç değerlendirmelerine katılım nasıl?”

Ö2: “Çoktan seçmeli testlerde yüksek not alması örnek verilebilir. Proje ve performans değerlendirmesinde ise etkin katılımı görevini yapmış olması”

Son-test verilerinde öğretmen adaylarının duyuşsal testlere verdiği önemi gösteren cevaplar aşağıda belirtilmiştir.

Ö3: “Tutum ölçekleri öğrencilere uygulanabilir. Bu ölçeklerden elde edilen sonuçlar değerlendirilip bir sonuca ulaşabilir. Böylece elimizde bir kanıt olur, hem de öğrencilerden motive edilip edilmedikleri öğrenilebilir.”

Ö8: “Özenle hazırlanmış tutum ölçekleri ön-test ve son-test olarak uygulanabilir. Buradan aldığı sonuçlar öğretmene bilgi verir.”

Standart 2. Öğretmenler, öğretim kararları için uygun değerlendirme yöntemleri geliştirme konusunda yetenekli olmalıdır. Bu standardın yeterlikleri ile ilgili olarak ölçme değerlendirme stratejisi belirleme, testin güvenilirliği artırma, üst düzey zihinsel becerileri değerlendirme, madde hazırlama, doğru değerlendirme kararlarına ulaşma kapsamında hazırlanan sorular öğretmen adaylarına sorulmuştur.

Bir ölçme aracından elde edilen sonuçların geçerliği için sadece istatistiksel işlemlerle yorum yapılması mümkün değildir. Geçerlikle ilgili karar verebilmek için birtakım çalışmalara ihtiyaç vardır. Uzmanların testleri incelemesi, başka testlerden elde edilen sonuçlarla istatistiksel işlemlerle kıyaslanması bu çalışmalara örnektir (Demircioğlu, 2015). Öğretmen adaylarının elde edilen puanların geçerliğine karar vermek için farklı ölçme sonuçlarıyla kıyaslama cevabının son-testte artış gösterdiği görülmektedir. Hem ön-testte hem de son testte sadece bir öğretmen adayı puanların geçerliği açısından testin kazanımlara uygunluğunun incelenmesi gerektiğini belirtmiştir. İstatistiksel işlemlerden yararlanma cevabını veren öğretmen adayının sayısının son-testte azalması sınıf içi ölçme değerlendirmede geçerliğin istatistiksel bir orandan ziyade kavramsal anlaşıldığının işareti olabilir.

Tablo 5. Öğretmen Adaylarının Envanterin Ön-test Uygulamasında Standart 2’ ye Yönelik Hazırlanan Sorulara Verdikleri Yanıtlar

Kategoriler	Alt Kategoriler	Ön-test (f)	Son-test (f)
Puanların geçerliğine karar verme	Testin kazanımlara uygunluğunu kontrol etme	1 (Ö1)	1 (Ö4)
	Farklı ölçme sonuçları ile kıyaslama	2 (Ö6, Ö8)	6 (Ö1, Ö2, Ö3, Ö5, Ö7, Ö8)
	Öğrenci yanıtlarını değerlendirme	2 (Ö1, Ö7)	-
	İstatistiksel işlemlerden yararlanma	3 (Ö1, Ö2, Ö4)	1 (Ö4)
Çoktan seçmeli testin güvenilirliğini artıran etkenler	Kazanımlara uygunluk	1 (Ö3)	-
	Uzman görüşü alınması	1 (Ö5)	-
	Teste katılan öğrenci sayısı	1 (Ö7)	-
	Sınav ortamı	1 (Ö4)	2 (Ö5, Ö7)
	Sınav süresi	1 (Ö3)	2 (Ö3, Ö7)
	Madde yazımı	1 (Ö3)	3 (Ö3, Ö4, Ö5)
	Madde güçlüğü	1 (Ö4)	-
	Soru sayısı	2 (Ö3, Ö4)	1 (Ö4)
	Öğrenci seviyesine uygunluk	-	1 (Ö7)
	Çeldiricilerin düzenlenmesi	-	4 (Ö1, Ö2, Ö4, Ö6)
	Düzeltilme formülü	-	2 (Ö2, Ö5)
	Uygun	3 (Ö3, Ö4, Ö6)	6 (Ö2, Ö4, Ö5, Ö6, Ö7, Ö8)
	Uygun değil	5 (Ö1, Ö2, Ö5, Ö7, Ö8)	2 (Ö1, Ö3)

Tablo 5’ te görüldüğü gibi bir öğretmenin belirlediği değerlendirme stratejisine göre puanların geçerliğine nasıl karar vermesi gerektiği sorusuna ön-test ve son-testte sekiz öğretmen adayından 1’i testin kazanımlara

uygunluğunu kontrol etme cevabını vermiştir. Test sonuçlarını farklı ölçme sonuçlarıyla kıyaslama yanıtını ön-testte 2 öğretmen aday, son-testte 6 öğretmen aday vermiştir. Öğrenci cevaplarının değerlendirilmesi ile puanların geçerliğine karar verilmesi gerektiğini sadece ön-testte, 2 öğretmen aday söylemiştir. Puanların geçerliğinde istatistiksel işlemlerden yararlanılması gerektiğini ön-testte 3 öğretmen aday, son-testte 1 öğretmen aday belirtmiştir.

Ön-test verilerinde öğretmen adaylarının istatistiksel işlemlere vurgusundan örnek cevaplar aşağıda sunulmaktadır.

Ö2: “Standart sapma ve madde analizi yaparak”

Ö4: “Ölçme aracının geçerlik ve güvenirliğini hesaplar sonuçlara göre yorum yapar.”

Son-test verilerinde öğretmen adaylarının diğer değerlendirme sonuçlarıyla kıyaslamasına verdikleri örnek cevaplar aşağıda sunulmaktadır.

Ö3: “Öğrencilerin ders içi başarısı ile yaptığı değerlendirmenin sonuçlarını kıyaslayarak belirleyebilir.”

Ö8: “Tek bir ölçme aracıyla öğrenciler hakkında bilgi edinmek zordur. En azından başlangıç, süreç ve sonuç olarak birden fazla test uygulanırsa ve öğrencilerin süreç ve sonuçtaki yanıtları tutarlıysa sonuçların geçerliliğini daha iyi belirlemiş olur.”

Çoktan seçmeli testin güvenirliği artıran etkenler sorulduğunda öğretmen adaylarının verdiği yanıtlar Tablo 5’ te görülmektedir. Öğretmen adaylarından 2’ si ön-testte soru sayısının güvenirliği etkilediğini belirtmiştir. Tablo 5’ te öğretmen adaylarının alt kategorilerindeki cevaplarının frekanslarının kazanımlara uygunluk (1), uzman görüşü alınması (1), teste katılan öğrenci sayısı (1), sınav ortamı (1), sınav süresi (1), madde yazımı (1), madde güçlüğü (1), soru sayısı (2) olduğu görülmektedir. Tablo 5’ te son-testteki alt kategoriler incelendiğinde sınav ortamı (2), sınav süresi (2), madde yazımı (3), soru sayısı (1), öğrenci seviyesine uygunluk (1), çeldiricilerin düzenlenmesi (4) ve düzeltme formülü (2) frekanslarına sahip olduğu görülmektedir.

Güvenirlik ölçme sonuçlarının duyarlı, tutarlı ve kararlı olmasıyla ilişkilidir. Dolayısıyla konu ve davranış bakımından homojen, madde sayısının çokluğu, maddelerin açık ve anlaşılır yazılması, testin iyi koşullarda uygulaması ve objektif puanlama yapılması güvenirliği artıran unsurlardır (Demircioğlu, 2015). Çoktan seçmeli testler hazırlanırken seçeneklerin oluşturulması madde yazımı kadar önemlidir (Yılmaz, 2015). Öğretmen adaylarının son-testte bu unsurlardan daha çok bahsetmiş olduğu dikkat çekmektedir.

Ön-test verilerinde öğretmen adaylarının çoktan seçmeli testin güvenirliğini artırmaya yönelik verdiği cevaplar örnekler;

Ö5: “Testin güvenirliği birden çok uzman tarafından değerlendirilmeyle artırılabilir”

Ö7: “Bence daha fazla kişi üzerinden değerlendirilerek...”

Son-test verilerinden öğretmen adaylarının çoktan seçmeli testin güvenirliği artırma konusunda daha geniş kapsamlı bilgiye sahip olduğu söylenebilir.

Ö3: “Soru sayısı belli bir yere kadar artırabilir. Sorular sade ve anlaşılır sorulabilir. Çeldiriciler amacına hizmet edecek şekilde hazırlanabilir.”

Ö4: “Soruları homojen şekilde hazırlamak, heterojen bir gruba testi uygulamadaki düzeltme formülü kullanmak, standart ortamlarda testi uygulamak güvenirliği artıracaktır.”

Ö7: “Seviyeye uygun hazırlanmış olmalı ve belirli ortamda uygulanmış olmalı ve ayrıca belirli bir süre olmalı. Bu şekilde artırılabilir.”

Standart 3. Öğretmen hem harici hem de kendi hazırladığı değerlendirme yöntemlerinin uygulama, puanlama ve sonuçlarını yorumlama konusunda yetenekli olmalıdır. Bu yeterlik standardı ile ilgili olarak test puanlamalarını yorumlama, uygun puanlama yöntemi tercih etme, öğrenci puanlamalarını değerlendirme ve test uygulamada zaman yönetimi kapsamında hazırlanan sorular öğretmen adaylarına yöneltilmiştir.

Tablo 6. Öğretmen Adaylarının Envanterin Ön-test Uygulamasında Standart 3' e Yönelik Hazırlanan Sorulara Verdikleri Yanıtlar

Kategoriler	Alt Kategoriler	Ön-test (f)	Son-test (f)
Bir testten alınan yüksek puanın değerlendirilmesi	Mutlak başarı ifade etmez	1 (Ö2)	6 (Ö2, Ö3, Ö4, Ö5, Ö6, Ö8)
	Konu öğrenilmiştir	7(Ö1, Ö3, Ö4, Ö5, Ö6, Ö7, Ö8)	2 (Ö1, Ö7)
	Soru-cevap yöntemi kullanma	1 (Ö5)	-
	Derecelendirme ölçeği hazırlama	1 (Ö4)	4 (Ö3, Ö4, Ö7, Ö8)
Proje/ürün değerlendirme stratejisi	Süreç gözlemleme	1 (Ö3)	-
	Puanlama anahtarı hazırlama	2 (Ö1, Ö8)	-
	Performans görevi oluşturma	2 (Ö2, Ö6)	4 (Ö1, Ö2, Ö5, Ö6)
	Doğru cevaba tam puan verme	1 (Ö7)	-

Öğretmen adaylarına 100 puan üzerinden 90 puan alan bir öğrenci hakkındaki yorumları sorulmuştur. Öğretmen adaylarının verdiği cevaplar Tablo 6' da görüldüğü gibi iki alt kategoriye ayrılmaktadır. Ön-testte 7 öğretmen adayı yüksek puan alan öğrencilerin konuyu öğrendiğini, 1 öğretmen adayı bunun mutlak başarı ifade etmediğini belirtmiştir. Son-testte 6 öğretmen adayı bir testten alınan yüksek puanın mutlak bir başarı ifade etmediğini, 2 öğretmen adayı konunun büyük oranda öğrenildiğini ifade etmiştir.

Ön-test verilerinden bu soruya verilen yanıtlardan örnekler aşağıda sunulmaktadır.

Ö3: “Eğer bir öğretmen puanlama işlemini 100 üzerinden yapıyor ve bu sınavda 100 alan bir öğrenci var ise ulaştırılmak istenen hedef davranışlara ulaştırılmış hiçbir eksiği yok manasına gelir. Buradan yola çıkarak 90 alan bir öğrencinin de hedef davranışların kazanılması yolunda birkaç eksiği olduğu söylenebilir. Bu eksikliğin nerede olduğunu, neden ötürü olduğunu keşfedip giderilmesi gerekmektedir.”

Ö4: “Öğrenci hedeflenen kazanımları edinmiş, üst düzey düşünme becerisi gerektiren bir soruda takılıp olabilir. Ölçmede hata türlerinden biri yaşanmış olabilir.”

Ö5: “90 almış öğrenci konuyu kavramıştır. 100 alamamasını sebebi dikkatsizlik, kafa karışıklığı olabilir.”

Son-test verilerinde öğretmen adaylarının çoğunluğu öğrenci başarısını yorumlama konusunda farklı ölçme sonuçlarının da kullanılması gerektiğini ifade etmişlerdir. Bu soruya dair yanıtlarından örnekler aşağıda sunulmaktadır.

Ö1: “Tek bir sınav ile öğrencinin başarısını ölçmek mümkün değildir. Çünkü sınavın içeriğine göre şans başarısı etkili olabilir. Ya da öğrencinin bildiği bir konu vardır. Testin kapsam geçerliği düşüktür böylece de yüksek almış olabilir. Tabii ki bu olumsuzlukların dışında öğrenci gerçekten öğrenmiş ve yapmış da olabilir.”

Ö8: “Soruların düzeyi, öğrencinin genel sınıftaki gözlenen başarısı önemlidir. Sorulara şans başarısı sayesinde doğru cevaplar verilmiş olabilir. Ama genel olarak sorular buna müsait değilse, öğrenci derslerde de başarı gösteriyorsa, sınıfta aktifse, farklı ölçümlerde puanları tutarlarsa konuyu anlamış olabileceği yönünde bir değerlendirme yapılabilir.”

Tablo 6' da öğretmenin konuyla ilgili bir model geliştirme ödevini değerlendirilmesi sorusuna öğretmen adaylarının verdiği cevapların frekanslarının soru-cevap yöntemini kullanma (1), derecelendirme ölçeği hazırlama (1), süreç gözlemleme (1), doğru cevaba tam puan verme (1), puanlama anahtarı hazırlama (2) ve performans görevi oluşturma (2) olduğu görülmektedir.

Performans değerlendirmeleri farklı çözüm yollarına imkan veren bir problem üzerine kurulu, çeşitli yaklaşım ve materyallerin denendiği, üst düzey becerilerinin kullanıldığı bir değerlendirme türüdür. Performans değerlendirmesinin uzun süreye yayılması, birçok beceriyi kapsamaması, grup ve bireysel değerlendirmelerin yapılması, ürüne ve sürece odaklanması, birçok veri toplama tekniğinin kullanılması gerekmektedir (Çepni, 2015). Performans değerlendirmelerinde öncelikle öğretmenin bağlamsal görevleri oluşturma, değerlendirme kriter ve aşamalarını belirlemesi gerekmektedir (Palm, 2008). Öğretmen adaylarının ön-testte verdiği yanıtlar incelendiğinde performans değerlendirmenin aşamalarını kısmen içermektedir. Son-testte verilen yanıtların çoğu performans görevinin planlanma ve değerlendirme süreçlerini birlikte ele aldığı görülmektedir.

Öğretmen adaylarına öğrencilere üçgenin alanını ölçme ile ilgili bir model geliştirmeleri istendiğinde izlenecek ölçme ve değerlendirme stratejisi sorulduğunda ön-testte verilen yanıtlardan örnekler aşağıda sunulmaktadır.

Ö5: “Burada açık uçlu bir ölçme değerlendirme yapılması daha uygun olabilir. Öğrencilere ‘Bu modeli neden tasarladın? Bu model bize ne gösteriyor?’ gibi sorular yapılacak değerlendirme için iyi bir yöntem olabilir.”

Ö7: “Yükseklik ve taban alanını belirleyebilecek materyalleri oluşturmayı başaran öğrenciler için tam puan verilebilir. Böyle bir strateji izlenebilir.”

Öğretim deneyinin ders içeriğinde performans değerlendirme yöntemlerine yer verilmesi ve uygulamalarının yapılmasının etkisi son-testte bu soruya verilen yanıtların tamamında hissedilmektedir.

Ö2: “Performans görevi verebilir. Sadece sonuç odaklı değil süreç odaklı bir değerlendirme yapabilir.”

Ö3: “Analitik rubrik kullanılmalıdır. Süreç sonunda rapor olmalıdır.”

Ö5: “Bu çalışmayı bir performans ödevi olarak değerlendirmek doğru olacaktır. Öğrenciler modeli geliştirirken ara görüşmelerle süreci kontrol altında tutmak not vermeyi daha doğru bir konuma ulaştıracaktır.”

Standart 4. *Öğretmenler, öğrenciler hakkında bireysel kararlar verirken, öğretimi planlarken, müfredat geliştirirken ve okul iyileştirmesi yaparken değerlendirme sonuçlarını kullanma becerisine sahip olmalıdır.* Bu standart yeterlikleriyle ilgili öğrencilerin hazırbulunuşluklarını belirleme, test puanlarını yordama, değerlendirme sonuçlarını kullanma ve eleştirme kapsamında hazırlanan sorular öğretmen adaylarına sorulmuştur.

Matematik sistem ve düzen üzerine kurulu teorilerden oluşan bilim dalıdır (Nasibov & Kaçar, 2005). Matematik eğitimi diğer bilim dallarından farklı olarak kesinlik içeren sayılar, uzay, nesne ve olaylar arasındaki ilişkileri kapsayan belirli matematiksel süreçleri içerir (Baki, 2008). Bu nedenle matematik öğretmenlerin öğrencilerini değerlendirme süreçlerinde referans olarak seçtikleri değerlendirme yöntemleri önem arz etmektedir.

Tablo 7. Öğretmen Adaylarının Envanterin Ön-test Uygulamasında Standart 4’ e Yönelik Hazırlanan Soruya Verdikleri Yanıtlar

Kategori	Alt Kategoriler	Ön-test (f)	Son-test (f)
Matematik dersi değerlendirmelerinde tercih edilen bilgi türü	Norm dayalı bilgi	2 (Ö1, Ö4)	3 (Ö1, Ö4, Ö5)
	Kritere dayalı bilgi	6 (Ö2, Ö3, Ö5, Ö6, Ö7, Ö8)	5 (Ö2, Ö3, Ö6, Ö7, Ö8)

Öğretmen adaylarına matematik dersindeki kesirler ünitesindeki aritmetik işlemler konusunda değerlendirme yaparken norm-dayalı bilgi (her öğrencinin performansının, diğer öğrencilerin performansına göre değerlendirilmesi) veya kritere dayalı bilgi (her bir öğrenciyi öğrenme çıktılarına göre ortaya koyduğu bireysel başarıyı referans alarak değerlendirme) tercihleri sorulmuştur. Tablo 7’ de görüldüğü gibi ön-testte 2 öğretmen adayı norm dayalı bilgi, 6 öğretmen adayı kritere dayalı bilgiye dayalı değerlendirme yapılmalıdır demiştir. Son-testte 3 öğretmen adayı norm dayalı bilgi, 5 öğretmen adayı kritere dayalı bilgi tercih edilmelidir yanıtını vermiştir.

Ön-testte elde edilen verilerden matematik dersi değerlendirmesinde norm-bilgi tercih edilmesine ilişkin yanıtlar aşağıdaki gibidir.

Ö1: “Norm dayalı bilgiyi kullanmalıdır. Çünkü; her bir öğrenci için ayrı ayrı ders planı hazırlayamaz. Her bir öğrencinin kazanımlarda belirli bir düzeye gelmesini bekleyemez. Zamandan da ekonomiklik sağlaması gerekir.”

Ö4: “Norm dayalı bilgi daha yararlı olur. Dersi planlarken öğrencilerin durumunu toplu olarak görülmesi ve buna göre yol izlenmesi gerekir.”

Benzer şekilde son-test verilerinden de öğrencinin bireysel değerlendirme kriteriyle grubun ortalama başarı konusunda kavram karmaşası olduğu görülmektedir. Matematik eğitimi açısından öğrencilere bireysel dönütlerin verilmesi ve öğrenci başarısını ders içeriklerini öğrenebilme ile ilişkilendirilmesi gerekliliği üzerine vurgu yapılmalıdır. Doğru ve etkili şekilde verilen geri bildirimler değerlendirmenin öğrenci ve öğretmen arasındaki iyi iletişimin önemli bir adımıdır (Türnüklü, 2003).

Ö1: “Norm-dayalı bilgiyi kullanmasını tavsiye ederim. Konular toplumun genel durumuna göre ayarlanması gerekir. Her öğrenciye hitap etmeye yetecek eğitim sisteminizde zaman yok.”

Ö4: “Norm-dayalı bilgiyi kullanarak sınıfın genel durumu hakkında bilgi sahibi olabilir. Bu sayede sınıfın tümünü göz önünde bulundurarak bir planlama yapar.”

Ö5:” Bence norm dayalı bilgi bu konu için daha yararlıdır olacaktır. Öğrencilerin genel olarak toplandığı aralık bize bu konuda daha doğru sonuç verecektir.”

Standart 5. *Öğretmenler, öğrenci değerlendirmelerini kullanan geçerli öğrenci derecelendirme prosedürleri geliştirme konusunda yetenekli olmalıdır.* Bu standart yeterlikleriyle ilgili değerlendirme tasarımı yapma, öğrenci başarı sıralaması yapabilme, not verme konularıyla ilgili hazırlanan sorular öğretmen adaylarına sorulmuştur.

Öğretmen adaylarından öğrenci başarı ve performansını en iyi şekilde ortaya koyan ve güvenilir bir değerlendirmeye olanak sağlayan ölçme değerlendirme tasarımı söylemeleri istenmiştir.

Ö7: “Etkinliklerdir. Kavram haritası ve dallandırılmış ağaç olabilir.”

Ö8: “Hazırlanan testteki soru maddelerinin iyi düşünülmüş, hata payının düşük olduğu, her düzeydeki bilgileri sorgulayan sorular hazırlanmalıdır.”

Son-testte öğretmen adaylarını tanılayıcı, süreç ve sonuç odaklı değerlendirme tasarımlarına değindikleri görülmektedir. Dersin başında konunun temelini oluşturan bilgileri ölçen tanılayıcı değerlendirme, dersin işleme sürecinde öğrenmeyi destekleyici öğrenciyi aktif tutan süreç değerlendirme, ders sonunda öğrenmenin gerçekleşip gerçekleşmediğini belirleyen sonuç değerlendirme yapılmalıdır (Uğurlu & Akkoç, 2011).

Ö2: “Tamamlayıcı ve klasik ölçme değerlendirme araçları birlikte kullanıldığında”

Ö7: “Bence süreç değerlendirme testleri yaparsa güvenilir hale getirebilir.”

Ö8: “Bence bizim de uygulamasını yaptığımız şekilde hazırbulunuşluk süreç içinde birden fazla test ve sonuç değerlendirme şeklinde birkaç test uygulaması ve öğrencilerin tutarlılığını ölçmek daha faydalı.”

Öğretim deneyi sonrasında öğretmen adayları geleneksel ve tamamlayıcı ölçme değerlendirme yöntemlerinin bir arada kullanılması gerektiğini belirtmişlerdir. Baştürk ve Dönmez (2011) çalışmalarında matematik öğretmen adaylarının tamamlayıcı ölçme ve değerlendirme yaklaşımına uygun fikirler paylaştıklarını ancak ders işleyişlerine yansıtamadıkları görülmüştür.

Standart 6. *Öğretmenler, değerlendirme sonuçlarını öğrencilere, velilere, ilgili kişilere ve diğer eğitimcilere iletme konusunda yetenekli olmalıdır.* Bu yeterlik standardı ile ilgili olarak test sonuçlarını değerlendirme, başarı farklılıklarını açıklayabilme, öğrenci başarılarıyla ilgili geçerli yorum yapabilme ile ilgili hazırlanan sorular öğretmen adaylarına yöneltilmiştir.

Öğretmen adaylarına bir öğrencinin 100 üzerinden Fen Bilgisi: 75 Sosyal Bilgiler: 75 Matematik: 75 puan aldığı söylenmiştir. Bu öğrencinin notlarına bakarak derslerdeki başarıları için geçerli bir yorum yapmaları istenmiştir.

Ön-test verilerinden öğretmen adaylarının öğrencinin yeterli düzeyde öğrendiğine ilişkin verdiği cevaplara örnekler aşağıda verilmiştir.

Ö5: “Sözel derste de sayısal derslerde de konuyu ana hatlarıyla anlamış. Ama derinlemesine öğrenme sağlanamamış.”

Ö6: “Öğrencinin ortalama bir başarısı var.”

Ö7: “Öğrenci konuyu genel hatlarla biliyor ama bazı eksiklikleri var demektir. Bazı bilgiler yeteri kadar oluşmamış demektir.”

Son-test verilerinin ön-test ile uyumlu olduğu görülmektedir.

Ö1: “Bu durumda üç derste de benzer şekilde yetkin görünüyor. 75 puanın düşük olduğunu düşünmüyorum bence iyi bir öğrenci.”

Ö4: “Öğrenci ulaşmak istenen hedeflerin çoğuna ulaşmış. Eksik olduğu veya zorlandığı kısımlar tespit edilip gelişimi için uygun bir yol izlenmeli.”

Ö8: “Konuların özünü anlamış olabileceğini ama konulara derinlemesine bakmadığını daha öğretmenin anlattıklarıyla sınırlı kaldığını düşünürüm.”

Standart 7. *Öğretmenler, etik dışı, yasa dışı ve diğer açılardan uygun olmayan değerlendirme yöntemlerini ve değerlendirme bilgilerinin kullanımını tanıma konusunda yetenekli olmalıdır.* Bu standart yeterlikleriyle ilgili test bilgilerinin gizliliğini koruma, adil uygulama ve etik kuralların ihlali ile ilgili hazırlanan sorular öğretmen adaylarına sorulmuştur.

Öğretmen adaylarının tamamı ölçme değerlendirmenin etik kuralların ihlali ile ilgili olarak sorulan sorulara doğru yanıtlar vermişlerdir. Bir öğretmen öğrencilerine yaptığı sınav sonuçlarını öğrenmeleri için ders çıkışında

öğretmenler odasından alıp notlarına bakabileceklerini söylüyor. Öğrenciler de diğer arkadaşlarının kâğıtlarını da görerek kendi notlarına bakmalarındaki yanlışlık nedir diye sorulmuştur.

Öğretmen adaylarının en yüksek puan aldığı bu standartta uygun olmayan ön-test yanıtları aşağıda sunulmuştur. Bir öğretmen adayı, öğrencilerin cevap kâğıtlarının görülmesinin yanlış olduğunun gerekçesini açıklarken adil olmayan bir uygulamayı doğruymuş gibi yorumlaması dikkat çekmektedir.

Ö1: “Öğrencinin derse olan çabasından dolayı bazen öğretmenler ufak tefek yanlışları düzeltebiliyorlar. Öğrenciler bunu görüp öğretmenlerine ya da arkadaşlarına tepki gösterebiliyorlar. Veli, okul idaresi bu iş uzayabiliyor. Onun yerine derste soruları çözmesi ya da kendi cevap anahtarını göstermesi daha ideal olur.”

4. Sonuç ve Öneriler

Değerlendirme okuryazarlığı kişinin değerlendirme kavramlarını ve prosedürlerini anlamasına odaklanır. Değerlendirme okuryazarı bir eğitimcinin bir testin güvenilirliğini gerçekten neyin oluşturduğunu ve farklılıklarını nasıl etkilediğini anlaması gerekir (Popham, 2011). Öğretmen adaylarının ölçme-değerlendirme alanına ilişkin anlayışlarına ulaşarak ve değerlendirme okuryazarlıklarına ilişkin göstergeler elde edilerek eksikliklerin tespit edilmesi gerekmektedir.

Araştırmanın bulguları, matematik öğretmeni adaylarının değerlendirme okuryazarlığının düşük seviyede olduğunu ve geliştirilmesi gerektiğini göstermektedir. Öğretmen adaylarının lisans eğitimlerinin sadece bir yarıyılında iki kredilik teorik dersle edindikleri bilgi ve beceriler, değerlendirme yaklaşım ve tekniklerine hakim olmak için yeterli olmamaktadır. Eğitim camiasında değerlendirme bilgi ve becerilerine artan vurguya rağmen, hizmet öncesi eğitim programlarında değerlendirme derslerinin sayısı düşük kalmaktadır. Bu derslerin içeriklerinin öğretmenlik kariyeri için yeterli olmadığı da bilinmektedir (DeLuca ve Klinger, 2010; Koç, 2019). Ayrıca, öğretmen adaylarının staj sırasında değerlendirme uygulamalarına daha aşina olmaları için daha uzun saatlere ihtiyaçları vardır. Öğretmen adaylarının eleştirel düşünebilmeleri, bazı zorluklarla baş etmeyi öğrenebilmeleri ve kendilerini geliştirebilmeleri için çeşitli değerlendirme uygulamalarına maruz kalmaları gerekmektedir (Xu ve He, 2019).

Araştırmanın hem nicel hem de nitel veri bulguları gerçekleştirilen öğretim deneyinin öğretmen adaylarının değerlendirme okuryazarlığı düzeylerini olumlu yönde etkilediğini göstermektedir. Benzer şekilde matematik öğretmen adaylarının ölçme değerlendirmelerini geliştirmeye yönelik yapılan bazı eğitimlerin olumlu sonuçları olduğu görülmüştür. Uğurlu ve Akkoç (2011) yaptıkları çalıştay sonrasında öğretmen adaylarının not verme, ölçme değerlendirme süreçlerindeki aksaklıkları tespit etme, ders sürecini iyileştirme gibi olumlu yönde etkisi olduğu bulmuşlardır. Daha iyi sonuçlara ulaşılması için bu çalışmada uygulanan öğretim deneyi ders içeriklerindeki eksikliklerin giderilmesi ve standartlara uygun hale getirilmesi gerekmektedir.

Öğretmen adayları lisans eğitiminde meslek derslerinde öğrendiklerini gerçek sınıf ortamına aktarmakta zorlanabilirler. Bu nedenle öğrendiklerini nasıl aktaracakları ve değerlendirmeyi daha iyi nasıl yapacakları konusunda onlara destek olmak gerekir (Yan ve Pastore, 2022). Öğretmen adayları değerlendirme okuryazarlığını, öğrenci bilgisini ölçmekten daha çok öğrencinin öğrenmesini kolaylaştırma ve anlamlı bir şekilde değerlendirme yeteneği olarak anlamaları gerekmektedir (Clark, 2015). Staj deneyimleri, deneyimli bir öğretmenin bunu nasıl yaptığı konusunda öğretmen adaylarına ilham verebilir. Öğretmen adaylarına staj deneyimleri sırasında hangi değerlendirme yöntemlerini gözlemledikleri sorulduğunda hepsi geleneksel yöntemlerin isimlerini söylemişlerdir (Özdil & Kaçar, 2022). Teoride faydası vurgulanan yöntemlerin uygulamada tercih edilmemesi öğretmen adaylarının bu yöntemlere verdikleri değeri ve ilgiyi azaltabilir. Bu nedenle staj sürecindeki ders deneyimleri boyunca fakülte-okul iş birliği ve fikir alışverişi aktif olmalıdır. Öğretmen yetiştirme eğitimlerinde öğretmen adaylarına ölçme değerlendirme stratejilerinin uygulama imkanı verilmesi bu alandaki becerilerini geliştirme konusunda önemli bir adım olabilir.

Bu araştırmanın bulgu ve sonuçlarının literatüre katkıda sağlaması ve öğretmen adaylarının değerlendirme eğitimini iyileştirmeye hizmet etmesi hedeflenmiştir. Matematik öğretmeni adaylarının değerlendirme okuryazarlık düzeylerinin düşük olması nedeniyle lisans eğitim programlarında değişikliklere gidilebilir. Değerlendirme derslerinde teorik bilgilerin yanı sıra uygulamalarla öğretim yapılabilir. Öğretmen adaylarının değerlendirme okuryazarlıklarını incelemek için daha fazla çalışmaya ihtiyaç duyulmaktadır. Bu çalışmanın önemli bulguları olmakla birlikte online olarak gerçekleştirilmesi, bir fakültede ve sekiz öğretmen adayı ile gerçekleştirilmesi gibi bazı sınırlılıkları vardır. Benzer bir çalışma farklı fakültelerdeki matematik öğretmen adaylarıyla yüz yüze eğitimde yapılması durumunda öğretmen adaylarının değerlendirme okuryazarlığı olgusu üzerindeki anlayışımızı güçlendirebilir.

Kaynaklar / References

- AFT [American Federation of Teachers], NCME [National Council on Measurement in Education] & NEA [National Education Association]. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: American Federation of Teachers.
- Akın, A., & Kabael, T. (2016). Bir Matematik eğitimi araştırmasına dayalı öğretim deneyi deneyimi. *Eğitimde Nitel Araştırmalar Dergisi*, 4(3), 7-27.
- Akpolat, T., & Genç, T. (2022). KPSS Kursuna devam eden öğretmen adaylarının epistemolojik inançları ile ölçme ve değerlendirme okuryazarlığı arasındaki ilişkiler. *Trakya Eğitim Dergisi*, 12(2), 647-662.
- Alkharusi, H. A. (2011, January). An analysis of the internal and external structure of the teacher assessment literacy questionnaire. *The International Journal of Learning*, 18(1), 515-528.
- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: an exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2), 217-232.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *The standards for competence in the educational assessment of students*. May 2022 tarihinde <https://buos.org/standards-teacher-competence-educational-assessment-students> adresinden alındı
- Arslan Mancar, S. (2019). *Performansa dayalı durum belirlemede puanlayıcılar arası güvenilirlik tekniklerinin karşılaştırılması*. Ankara: Ankara Üniversitesi.
- Azrak, Y., & Yalçınkaya, E. (2019). Sosyal bilgiler öğretmen adaylarının ölçme-değerlendirme okuryazarlık düzeylerinin çeşitli değişkenler açısından incelenmesi. *Anadolu Journal of Educational Sciences International*, 9(1), 27-55.
- Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2015). *Geleneksel-tamamlayıcı ölçme ve değerlendirme teknikleri öğretmen el kitabı*. Ankara: Pegem Akademi.
- Baki, A. (2008). *Kuramdan uygulamaya matematik eğitimi*. Ankara: Harf Eğitim Yayıncılığı.
- Baki, M., Çelik, D., Güler, M., & Sönmez, N. (2018). *Matematik öğretmeni adaylarının öğrenciyi tanıma bilgilerinin incelenmesi: bir ders analizi çalışması*. Kastamonu Eğitim Dergisi, 26(1), 143-152.
- Ball, R. P. (2017). *Virginia principals' knowledge of classroom assessment and support of assessment for learning practices*. Virginia: The College of William and Mary in Virginia The Faculty of the School of Education.
- Baştürk, S., & Dönmez, G. (2011). Matematik öğretmen adaylarının pedagojik alan bilgilerinin ölçme ve değerlendirme bilgisi bileşeni bağlamında incelenmesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 12(3), 17-37.
- Berk, H. (2021). *Bilgisayar ve öğretim teknolojileri öğretmen adaylarının ölçme-değerlendirme okuryazarlık seviyeleri ve STEM tutumlarının incelenmesi*. İstanbul: Marmara Üniversitesi İstanbul Sabahattin Zaim Üniversitesi.
- Beziat, T., & Coleman, B. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25-30.
- Bıkmaz Bilgen, Ö., & Doğan, N. (2017). Puanlayıcılar arası güvenilirlik belirleme tekniklerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 63-78.
- Birgin, O., & Gürbüz, R. (2008). Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki bilgi düzeylerinin incelenmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20, 163-179.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. Paper presented at the meeting of the *Mid-Western Educational Research Association*, (s. 17-18). Columbus, OH.
- Chappuis, S., & Stiggins, R. j. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-43.
- Clark, J. S. (2015). "My assessment didn't seem real": The Influence of field experiences on preservice teachers' agency and assessment literacy. *Journal of Social Studies Education Research*, 6(2), 91-111.
- Czarnocha, B., & Prabhu, V. (2006). Teaching Experiment NYCity Model. *Roczniki Polskiego Towarzystwa Matematycznego. Seria V: Dydaktyka Matematyki*, 29, 251-272.
- Çelik, Z., & Arslan, Y. (2012). Aday Beden eğitimi öğretmenlerinin ölçme ve değerlendirme genel yeterlik algılarının belirlenmesi. *Selçuk University Journal of Physical Education and Sport Science*, 14(2), 223-232.
- Çepni, S. (2015). Performansların değerlendirilmesi. E. Editör Karip içinde, *Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- DeLuca, C., & Klinger, D. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- Demircioğlu, G. (2015). Geçerlik ve güvenilirlik. E. Editör Karip içinde, *Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.
- Dokumacı Sütçü, N., & Bulut, İ. (2016). Ortaokul matematik öğretmenlerinin alternatif ölçme ve değerlendirme tekniklerine ilişkin yeterlik algıları ve bu teknikleri kullanma düzeyleri. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 29, 289-308.

- Ergül, A. Ö. (2019). *Öğretmenlerin Ölçme ve değerlendirme okuryazarlık düzeylerinin bazı değişkenler açısından incelenmesi*. Ankara: Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü.
- Fan, Y.-C., Wang, T.-H., & Wang, K.-H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computer & Education*, 57, 1727-1740.
- Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlilik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- Göktaş, Ö., & Şad, S. N. (2021). Matematik öğretmenlerinin ölçme ve değerlendirme yaklaşımlarına ilişkin algılarının incelenmesi. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), 402-415.
- Gömlüksiz, M. N., & Kan, A. Ü. (2010). Sınıf Öğretmenliği adaylarının alternatif ölçme-değerlendirme yaklaşımlarını tanıma düzeylerine ilişkin bir değerlendirme. Sınıf Öğretmenliği Adaylarının Yeni İlköğretim Programındaki Ölçme-Değerlendirme Yöntemlerini Tanıma Düzeylerine İlişkin Görüşleri (Fırat, Dicle ve Kilis Yedi Aralık Üniversiteleri Örneği). Ankara: I. Ulusal İlköğretim Kongresi.
- Gül, E. (2011). *İlköğretim öğretmen adaylarının ölçme-değerlendirme okuryazarlığı ve ölçme-değerlendirmeye ilişkin tutumlarının belirlenmesi*. Elazığ: Fırat Üniversitesi.
- Gürsoy, G. (2017). Ölçme ve değerlendirme okuryazarlığı: kavramsal bir analiz. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 6(1), 281-316.
- Güven Akdeniz, D. (2021). Amaç, ölçüt ve puanlama: matematik öğretmen adayları tarafından oluşturulan dereceli puanlama anahtarlarının incelenmesi. *Cumhuriyet International Journal of Education*, 10(3), 974-997.
- Harlen, W. (2007). *Assessment of learning*. London: Sage Publications.
- İzci, K., Çalışkan, G., & Aktürk, A. O. (2018). Öğretmen adaylarının sınıf içi ölçme-değerlendirme okuryazarlıklarına etki eden etmenlerin incelenmesi: Bir yapısal eşitlik modellemesi. *Eğitim, Bilim ve Teknoloji Araştırmaları Dergisi*, 3(1), 10-22.
- Kanık, E. A., Temel, G. O., & Kaya, I. E. (2010). Fleiss Kappa ve Krippendorff Alpha uyum katsayılarının örneklem genişliği, değerlendirici sayısı ve kullanılan ölçeğin kategori sayısından etkilenme durumları üzerine bir benzetim çalışması [Effect of Sample size, the number of raters and the category levels of diagnostic test on Krippendorff Alpha and the Fleiss Kappa statistics for calculating inter rater agreement: A simulation study]. *Türkiye Klinikleri Biyoistatistik*, 2(2), 74.
- Karaca, E. (2003). *Öğretmen adaylarının ölçme ve değerlendirme yeterliliklerine ilişkin algıları*. Ankara: Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Karakuş, F. (Bahar 2010). Ortaöğretim matematik dersi öğretim programında yer alan alternatif ölçme ve değerlendirme yaklaşımlarına yönelik öğretmen görüşleri. *Türk Eğitim Bilimleri Dergisi*, 8(2), 457-488.
- Karaman, P., & Şahin, Ç. (2014, Ağustos). Öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının belirlenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 15(2), 175-189.
- Kıral, B. (2020). Nitel bir veri analizi yöntemi olarak doküman analizi. *Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 15, 170-189.
- Kilmen, S., Akın Kösterelioğlu, M., & Kösterilioğlu, İ. (2007). Öğretmen adaylarının ölçme değerlendirme araç ve yaklaşımlarına ilişkin yeterlik algıları. *AİBÜ, Eğitim Fakültesi Dergisi*, 7(1), 129-140.
- Kilmen, S., & Çıkrıkçı Demirtaşlı, N. (2009). Sınıf öğretmenlerinin ölçme ve değerlendirme ilkelerini uygulama düzeylerine ilişkin görüşleri. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 42(2), 27-55.
- Koç, A. (2019). *Fen bilgisi öğretmen adaylarının ölçme değerlendirme okuryazarlık düzeylerinin belirlenmesi*. Antalya: Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü.
- Koç, A., & Bulut, S. (2020). Fen bilgisi öğretmen adaylarının ölçme değerlendirme okuryazarlık düzeylerinin çeşitli değişkenler açısından incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 40(3), 877-904.
- Koh, K. (2014). Developing preservice teachers' assessment literacy: a problem-based learning approach. I. P. (Ed.), *Proceedings of the IDEAS: Rising to Challenge Conference* (pp. 113-120). Calgary, Canada: Werklund School of Education, University of Calgary.
- Lee, J.-E., & Son, J.-W. (2015). Two teacher educators' approaches to developing preservice elementary teachers' mathematics assessment literacy: Intentions, outcomes, and new learning. *Teaching & Learning Inquiry: The ISSOTL Journal*, 3(1), 47-62.
- Mertler, C. A. (2003). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* <https://files.eric.ed.gov/fulltext/ED482277.pdf>
- Mertler, C. A. (2004, Fall). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49-64.
- Mertler, C. A., & Campell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "Assessment Literacy Inventory"*. <https://files.eric.ed.gov/fulltext/ED490355.pdf>
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA, London: Sage Publications.
- Nasibov, F., & Kaçar, A. (2005). Matematik ve matematik eğitimi hakkında. *Kastamonu Eğitim Dergisi*, 13(2), 339-346.

- Önel, F., Dalkınıç, F., Özel, N., Deniz, Ş., Balkaya, T., & Kurt Birel, G. (2020). Ortaokul matematik öğretmenleri ölçme-değerlendirmeyi nasıl yapıyor? Bir durum çalışması. *Kastamonu Education Journal*, 28(3), 1448-1459.
- Özenç, M. (2013). Sınıf öğretmenlerinin alternatif ölçme ve değerlendirme bilgi düzeylerinin belirlenmesi. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 21, 157-178.
- Özdi, G., & Kaçar, A. (2022). Views of secondary school mathematics pre- service teachers about the field of assessment. *Osmangazi Journal of Educational Research*, 9(2), 178-200.
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation*, 13(1), 4.
- 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. (1992). *Teacher competencies questionnaire description*. Lincoln: NE: University of Nebraska.
- Plake, B.S. & Impara, J.C. (1993). *Teacher assessment literacy questionnaire*. University of Nebraska- Lincoln. In cooperation with the National Council on Measurement in Education and the W.K. Kellogg Foundation
- 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.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
- Roslan, R., Nisho, Y., & Jawawi, R. (2022). Analyzing English language teacher candidates' assessment literacy: A case of Bruneian and Japanese universities. *System*, 111, 102934.
- Sabancı, O., & Yazıcı, K. (2017). Öğretmen Adaylarının ölçme ve değerlendirmeye yönelik yeterlik algılarının incelenmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 7(1), 128-153.
- Sak, R., Şahin Sak, İ. T., Öneren Şendil, Ç., & Nas, E. (2021). Bir araştırma yöntemi olarak doküman analizi. *Kocaeli Üniversitesi Eğitim Dergisi*, 4(1), 227-250.
- Salmon, S. J. (1996). *Alternative math assessment: Teachers making sense of assessment in their classrooms*. Buffalo, New York: State University of New York at Buffalo.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Suzuki, K. (1998). *Measuring "To Think Mathematically" Cognitive Characterization of Achievement Levels in Performance-Based Assessment*. Urbana, Illinois: University of Illinois at Urbana-Champaign.
- Şimşir Gökalp, Z. (2021). Matematik motivasyonu. E. Editörler Ertekin, & B. Dilmaç içinde, *Matematiğin Duyuşsal Özellikleri* (s. 171-190). Ankara: Pegem Akademi.
- Tünkler, V. (2019). Sosyal bilgiler öğretmen adaylarının değerlendirme okuryazarlık düzeyi ile değerlendirme öz-yeterlik algılarının incelenmesi. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 37, 319-332.
- Türnüklü, E. B. (2003). Türkiye ve İngiltere'deki Matematik öğretmenlerinin değerlendirme biçimleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 24, 108-228.
- Uğurlu, R., & Akkoç, H. (2011, Temmuz). Matematik öğretmen adaylarının ölçme-değerlendirme bilgilerinin gelişiminin tamamlayıcı-şekillendirici ölçme-değerlendirme bağlamında incelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi*, 2(30), 155-167.
- 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.
- Xu, Y., & He, L. (2019, December). How Pre-service Teachers' Conceptions of Assessment Change Over Practicum: Implications for Teacher Assessment Literacy. *The journal Frontiers in Education*, 4, 145. <https://doi.org/10.3389/feduc.2019.00145>
- Yılmaz, A. (2015). Ölçme değerlendirmede testler. E. Editör Karip içinde, *Ölçme ve Değerlendirme* (s. 152-229). Ankara: Pegem Akademi.
- Yaman, S., & Karamustafaoğlu, S. (2011). Investigating prospective teachers' perceived levels of efficacy towards measurement and evaluation. *Ankara University Journal of Faculty of Educational Sciences*, 44, 53-72.
- Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, 116, 2998-3004.
- Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation*, 74, 101183.
- Yaralı, D. (2017). Öğretmen adaylarının ölçme ve değerlendirmeye yönelik yeterlik algılarının incelenmesi (Kafkas Üniversitesi Örneği). *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 487-504.