Journal of Theoretical Educational Science, 17(2), 324-350, April 2024
Kuramsal Eğitimbilim Dergisi, 17(2), 324-350, Nisan 2024
[Online]: http://dergipark.org.tr/akukeg
DOI number: http://doi.org/10.30831/akukeg.1276067

Program Evaluation Study for Measurement and Evaluation Course in Distance Education*

Uzaktan Eğitimde Ölçme ve Değerlendirme Dersine Yönelik Bir Program Değerlendirme Çalışması

Muhammed AKINCI** **Erdoğan KÖSE***

Received: 26 May 2023 **Research Article** **Accepted:** 5 February 2024

ABSTRACT: This study aims to evaluate the online measurement and evaluation course in teacher training programs during the COVID-19 process. In the study, we sought answers to two primary research questions: What are the opinions of teachers and school administrators regarding their measurement and evaluation competencies? Does the online "measurement and evaluation" course have the qualities of an effective program in the "antecedents, transactions, and outcomes" dimension? We structured the research into two phases within a multistage evaluation design framework. The findings show that there were problems and positive aspects in all dimensions of the program. For example, adapting teacher training programs developed before COVID-19 to distance education processes was challenging. In distance education, some practices contradict the modern teaching and assessment approach. Such problems were reflected in teachers' acquisition of measurement and evaluation competencies. The achievement test we applied to the observed groups also confirmed these findings. For this reason, responsible organizations should not ignore the fact that we cannot renounce distance education. During program development, they should reconsider how the teachers will acquire measurement and evaluation competencies and how we will measure and evaluate in distance education.

Keywords: Measurement, evaluation, distance education, teacher training, program.

ÖZ: Bu çalışmanın amacı, COVID-19 sürecinde öğretmen yetiştirme programlarında yer alan çevrimiçi eğitimde ölçme ve değerlendirme dersinin değerlendirilmesidir. Çalışmada iki temel araştırma sorusuna yanıt aranmıştır: Öğretmenlerin ve okul yöneticilerinin ölçme ve değerlendirme yeterliklerine ilişkin görüşleri nelerdir? Çevrimiçi "ölçme ve değerlendirme" dersi "girdiler, işlemler ve çıktılar" boyutlarında etkili bir programın sahip olması gereken nitelikleri taşıyor mu? Araştırma çok aşamalı bir değerlendirme tasarımı çerçevesinde iki aşamalı olarak yapılandırılmıştır. Araştırma bulguları, programın tüm boyutlarında sorunların ve olumlu yönlerin olduğunu göstermektedir. Örneğin, COVID-19 öncesi geliştirilen öğretmen yetiştirme programlarını uzaktan eğitim süreçlerine uyarlamak güçtür. Uzaktan eğitimde, bazı uygulamalar modern öğretim ve değerlendirme anlayışıyla çelişmektedir. Bu tür sorunlar öğretmenlerin ölçme ve değerlendirme yeterliklerini kazanmalarına yansımıştır. Gözlem yapılan gruplarda uygulanan başarı testi de bu bulguları doğrulamıştır. Bu nedenle sorumlu kuruluşlar uzaktan eğitimden vazgeçmenin artık mümkün olmadığını göz ardı etmemelidir. Ayrıca bu kuruluşların program geliştirme süreçlerinde uzaktan eğitimde ölçme ve değerlendirmenin nasıl yapılacağının yanı sıra öğretmenlerin ölçme ve değerlendirme yeterliklerini nasıl kazanacağını yeniden düşünmeleri gerekmektedir.

Anahtar kelimeler: Ölçme, değerlendirme, uzaktan eğitim, öğretmen yetiştirme, program.

* This study was produced from the Ph.D. dissertation titled *Evaluation of Measurement and Evaluation in Education Curriculum* prepared by the first and corresponding author of this research.

**Corresponding Author*: Asst. Prof. Dr., Recep Tayyip Erdoğan University, Rize, Türkiye, muhammed.akinci@erdogan.edu.tr, https://orcid.org/0000-0002-5001-2080
*** Prof. Dr., Akdeniz University, Antalya, Türkiye, ekose@akdeniz.edu.tr, https://orcid.org/0000-0003-0426-0267

Citation Information
Akinci, M., & Köse, E. (2024). A program evaluation study for measurement and evaluation course in distance education. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science], 17*(2), 324-350.

ISSN: 1308-1659

People use the concept of evaluation when making decisions about the quality of any business, program, or object. Comments or decisions about the quality of an education system, teaching, or programs related to this system are also associated with evaluation. In this context, evaluation in education refers to the process of considering and deciding the success of learning and teaching according to specific criteria based on the measurement results (Morrow Jr et al., 2000; Secolsky & Denison, 2018). In determining the quality of the teaching activity, we frequently come upon concepts such as testing, measurement, evaluation, and assessment. People use these terms interchangeably and sometimes even mix them up (Mehrens & Lehmann, 1991). Measurement is the process of using measuring tools to determine an object's quantity or value its features by assigning numeric or symbolic values. (Morrow Jr et al., 2000; Mislevy, 2018; Scriven, 1981). Based on this definition, it is conceivable to express that the role of measurement is to provide decision-makers with accurate information about the feature to be evaluated (Mehrens & Lehmann, 1991). From this point of view, considering measurement and evaluation processes complement each other, the accuracy of the measurement process may also affect the accuracy of the evaluation process and the decisions to be taken. Accordingly, teachers have an essential role in measuring and evaluating the quality of teaching. Teachers' role demonstrates these people must acquire measurement and evaluation competencies. Researchers have been working on teachers' measurement and evaluation competencies for years (Chan & Luk, 2021; DeLuca et al., 2016; Holmes, 1971; Plake, 1993; Zhang, 1996). This condition may indicate that the subject has maintained its importance and popularity over the years.

Measurement and Evaluation Education in Teacher Training Programs

Various organizations that develop standards for the quality of teachers and teaching in the world consider measurement and evaluation among the core teacher competencies (DfE. UK, 2021; DfE. US, 2021; ISO, 2021; QCT, 2021). Likewise, in the research country, measurement and evaluation are commonly included among teacher competencies and teacher training standards (EPDAD, 2020; MNE, 2017; MNE, 2019). Over the years, different courses have existed in teacher training programs to acquire measurement and evaluation competencies in Türkiye. For the first time in this field, a measurement and evaluation course was figured in the program of two-year education institutes that trained classroom teachers in 1974 (MNE, 1975). The related course continued to be conducted by teacher training institutions affiliated with the Council of Higher Education (CHE) as a three-credit course with the regulations made in 1980, 1982, and 1985 (MNE, 1980; CHE, 2007). CHE replaced the "measurement and evaluation" course in teacher training programs with a four-credit Instructional planning and evaluation course with the 1997 regulation (CHE, 1998). The Council removed the relevant course from the programs with the 2007 regulation and brought back a three-credit measurement and evaluation course with new contents according to the progressive approach (Ada & Baysal, 2013; CHE, 2007). The last update in teacher training undergraduate programs occurred in 2018, and they reduced the credits of the measurement and evaluation course to two with more straightforward contents. All these changes may lead to future stability issues in the measurement and evaluation training provided to teachers in Türkiye.

The COVID-19 pandemic has taken the entire world under its influence in a short time. For this reason, many educational institutions have switched to online teaching activities during the pandemic. Besides its various advantages in terms of time, place, and economy, distance education has considerable challenges. Assessing student achievement is among these difficulties (Conrad & Openo, 2018; Kearns, 2012; Kim et al., 2008). After COVID-19, distance education has also become widespread in teacher training institutions. In addition to the concern of *how we will carry out the measurement and evaluation processes in distance education*, this situation has also revealed the question of *how pre-service teachers will acquire measurement and evaluation competencies*. In current studies, there is a tendency to seek new ways to overcome assessment problems in distance education (Barthakur et al., 2022; Pavličet al., 2022; Raje & Stitzel, 2020). However, it is a matter to contemplate whether pre-service teachers can adequately acquire the measurement and evaluation competencies through distance education and to what extent the measurement and evaluation training offered to teachers responds to their needs. One of the best ways to find answers to these questions may be an evaluation of online measurement and evaluation course in teacher education.

Measurement, Evaluation and Program Evaluation

Evaluation is an expression that people use to determine the quality of the program as much as to assess the quality of teaching activities. This process, in which we determine the quality of the program and decide its future, is referred as program evaluation (Fitzpatrick et al., 2011; Mertens & Wilson, 2019; Stufflebeam & Coryn, 2014). Today, various institutions increasingly use program evaluation to meet the demand for information on performing public and non-profit schools and private instructional programs (Grayson, 2018). This case discloses that the information needs of the stakeholders regarding the quality of the implemented programs are increasing, and program evaluation is one of the main mechanisms in meeting this need.

Some studies have been under the title of program evaluation for measurement and evaluation discipline in teacher training programs. However, in these studies, researchers examined various practices, approaches, and trends towards the relevant course conducted rather than a systematic program evaluation process (Safrit, 1990; Veal, 1990). Today, some studies in teacher training, measurement, evaluation, and program evaluation focus on the practice of measurement in program evaluation or program-based measurement (Alkharusi, et al., 2011; Graves, 2010; Harris, 2017; Tang, 2012). Different researchers have various studies also on program-based measurement and evaluation (Christ & Desjardins, 2018; Fuchs, 2017; Noble, 2012; Shapiro & Gebhardt, 2012; Tindal, 2013). In these studies, researchers evaluate applications for using program-based measurement and evaluation in different disciplines in the historical process. Another type of study in which the concepts of measurement and evaluation and program evaluation are a matter of research is studies that adopt the use of measurement and evaluation in program evaluation (Harmon et al., 1998; McDowell, 1992; Tusing & Breikjern, 2017). These studies include various measurement and evaluation approaches and methods researchers can employ in the program evaluation process.

Various studies in Türkiye examine situations such as teachers' measurement and evaluation competencies, perceptions, and needs, sometimes under the title of evaluation. (Duman, 2019; Karakuş & Turhan Türkkan, 2017; Karaman and Şahin 2014; Koç, 2019; Sabancı and Yazıcı 2017; Sevimel Şahin, 2019; Yaralı, 2017). However, in these studies, no direct evaluation research was found regarding the measurement and evaluation course in teacher training programs. Other studies examining measurement and evaluation education for teachers in various contexts in Turkey were either not conducted within the scope of program evaluation or did not focus on a specific teaching area.

The importance of measurement and evaluation as a teaching competence, the possible instabilities in the research country in this regard, limited studies on the subject in the literature, and the online education needs arising from situations such as COVID-19 require the evaluation of teachers' measurement and evaluation training programs. Given the importance of this aspect of teacher education, we believe that a comprehensive evaluation of measurement and evaluation program is necessary to equip teachers with the skills they need to provide quality education to their students. Thus, this study will provide the opportunity to facilitate the identification of shortcomings in the pertinent programs, and the derived insights will provide a foundation for updating these programs to strengthen their contribution to the teacher education system.

Aim

This study aims to evaluate the online measurement and evaluation course in teacher training programs during the COVID-19 process in Türkiye. In the first step, we aim to describe the situation of teachers and school administrators regarding the measurement and evaluation competencies. The primary purpose here is to reveal the measurement and evaluation needs in the field where the teaching profession is carried out. Based on this, the next step is to determine whether the course meets the needs in the field in the dimensions of "antecedents, transactions, and outcomes" as required by the program evaluation procedures adopted in the study. In line with these aims, the research pursued answers to the following questions:

1. What are the opinions of teachers and school administrators regarding their measurement and evaluation competencies?
2. Does the "measurement and evaluation" course have the qualities of an effective program in the "antecedents" dimension?
3. Does the "measurement and evaluation" course have the qualities of an effective program in the "transactions" dimension?
4. Does the "measurement and evaluation" course have the qualities of an effective program in the "outcomes" dimension?

Method

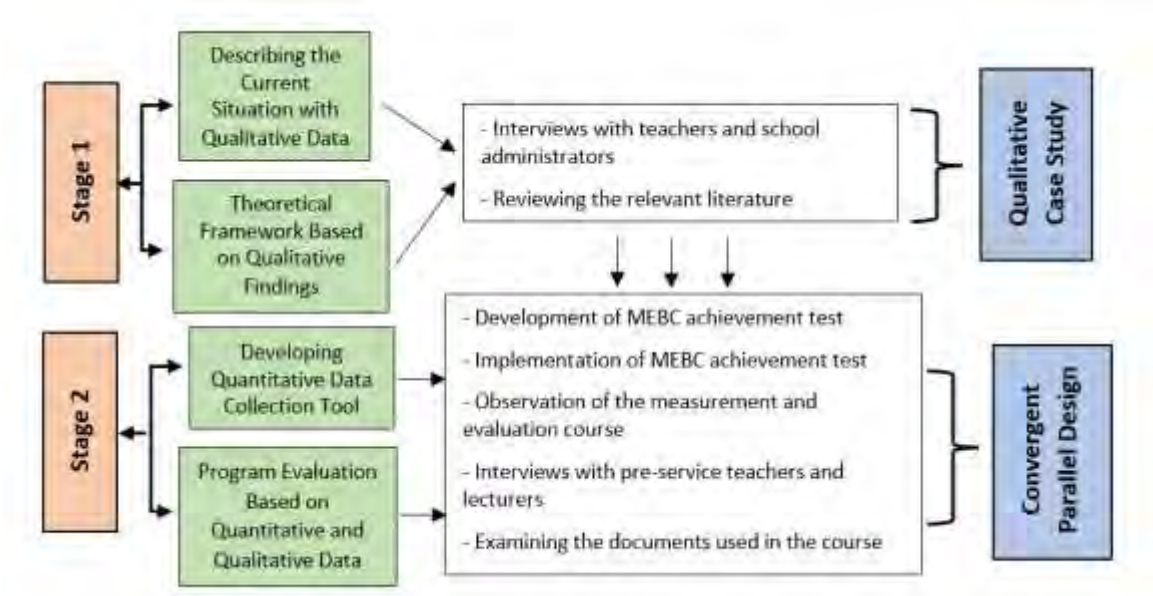
Research Design

This study is a program evaluation research using a multistage evaluation design. Multistage evaluation design is a mixed-method design that we utilize to evaluate the effectiveness of a program or project (Creswell, 2015). Multistage evaluation design, also known as mixed method "program evaluation design,"

comprises one or more core designs added to the steps in an evaluation setup conducted to evaluate the success of an intervention or a program (Creswell & Plano-Clark, 2018). Under related design, this research used a qualitative case study and mixed convergent parallel design. We examined the measurement and evaluation course as a single "case." This type of case study that deals with a whole situation is a singular case study design (Stake, 2005). Moreover, the purpose of convergent parallel design is to collect both quantitative and qualitative data simultaneously, combine these data, compare the results, and explain the differences in the results (Cohen et al., 2007; Patton, 2015). Figure 1 demonstrates the execution stages of the multistage evaluation design used in the research.

Figure 1

Execution Stages of the Multistage Evaluation Design.



The study was conducted in two stages, in which program evaluation procedures were employed in addition to the methodological design. This approach allowed us to evaluate the effectiveness of the program in a comprehensive manner, ensuring that all relevant factors were considered.

Program Evaluation Procedures

We used Stake's Responsive Program Evaluation Model to evaluate the measurement and evaluation course program. We benefited from the "Description Matrix" of Stake's Congruence-Contingency Model to collect descriptive data regarding the evaluated program. Stake's Responsive Model is an evaluation approach or model that also includes the Congruence-Contingency Model that he developed before (Stufflebeam, 1983). The Congruence-Contingency Model is a participant-oriented model developed in 1967 that aims to conceptualize the program evaluation process and make it understandable. The model deals with the dimensions of *antecedents*, *transactions*, and *outcomes* in program evaluation (Stake, 1967; Stufflebeam & Coryn, 2014). Today, researchers also use the Congruence-Contingency Model's data collection approach in describing the program's activities in the responsive model (Stufflebeam & Coryn, 2014). However, although Stake has laid its foundations in the Congruence-Contingency Model, Responsive Evaluation is more informal, pluralistic, and process-

oriented (Fitzpatrick et al., 2011). Stake (2011) expressed that the responsive model adopts an approach that compromises some measurement precision to increase the findings' usefulness. Because responsive evaluation also uses informal data collection processes that require natural communication with stakeholders (Stake, 2014; Stake, 2013). For this reason, we took measures to strengthen this weakness of the model. For this purpose, the program evaluation standards prepared by the Joint Committee on Standards for Educational Evaluation (JCSEE) were used during the program evaluation process. Moreover, the qualitative findings of the study were supported by quantitative data, and an achievement test was developed in this context.

Data Collection Procedures

Qualitative Data Collection

In the qualitative data collection processes, we used semi-structured interview and observation forms for teachers, school administrators, instructors, and pre-service teachers. Interview forms comprise 5 to 8 questions for each interviewer and sub-questions that may change according to the interview flow. The interview questions cover opinions about the model's dimensions used in program evaluation procedures. In the observation form, there are the following variables that we think will help to observe teaching activities in distance education processes:

1. The design of the distance education platform
2. Student-teacher interaction
3. Teaching strategies, approaches, methods, and techniques
4. Tools, resources, and materials
5. Assessment methods and techniques

In preparing the interview and observation forms, we received the expert opinions of individuals in the fields and disciplines of curriculum and instruction, measurement and evaluation, and qualitative research methods at three universities. Moreover, we subjected the interview forms to pilot applications and rewrote some questions because they lacked intelligibility. In addition, we asked the interviewees to confirm whether there was any problem with the interview content. Of the 67 interviewees who were asked for participant confirmation, 47 responded, and two lecturers, a teacher, and a pre-service teacher made additions to the questions that they answered incompletely because of anxiety, stress, and inattention during the interview.

Researchers assumed the role of the observer as a participant under the non-participant observation approach in order not to affect the research area and not be affected by the research area during the observations. In this context, the possibility of not being physically present in the teaching environment and not interacting, thanks to the distance education process, contributed to not affecting the research area.

Quantitative Data Collection

Measurement and Evaluation Basic Concepts (MEBC) Achievement Test was another measure we took to strengthen the part of the Responsive Model that compromises some precision in measurement. For this reason, we applied an achievement test to the pre-service teachers who took the measurement and evaluation

course to evaluate their learning with a standardized tool. We adopted the goal selection and writing principles of (Miller et al., 2009) in preparing test items. In this context, we set 10 goals in different cognitive learning stages, including the seven-week objectives and contents in the courses until the midterm week. To ensure content validity, we prepared a pool of 84 questions, including items for each objective. A preliminary test, comprising 30 questions in a multiple-choice format with five choices, was developed using items from the abovementioned pool. The test was then made accessible on online platforms such as Google Forms and Microsoft Forms to facilitate applicability in an online learning environment. Therefore, a question distribution was adopted to ensure content validity, with at least two questions for each outcome. Ultimately, we made a 30-minute preliminary application of the draft test to predict and provide psychometric qualities in 2 groups (132 pre-service teachers) that were equivalent to the observed groups.

Table 1

Item Statistics Regarding the Pre-Application of the Achievement Test.

n	$\bar{X}$	*ss*	Skewness-Kurtosis	Number of Items Discarded	Item Difficulty	Item Discrimination	KR20
132	54.80	3.89	0.086 - 0.916	7	0.54	0.40	0.70

We examined the skewness and kurtosis coefficients to determine whether the distribution of the achievement test results was normal, which revealed a range of values between 0.086 and 0.916. The fact that these values are between +1 and -1 means that the scores do not significantly deviate from the normal distribution (Büyüköztürk, 2014). For this reason, we performed the remaining analyzes assuming that the test scores indicated normal distribution. Item difficulty of 0.50 and around item discrimination of 0.30 and above is reasonable in a test (Güler, 2018). In cases where item discrimination is below 0.20, practitioners should remove the item from the test and correct the ones above this value (Atılgan, 2018). Therefore, we removed seven items with item discrimination below 0.20 from the test, corrected four items with a corresponding value between 0.23 and 0.28, and included them in the test. After these procedures, the mean scores obtained became 54.8 out of 100, the item difficulty 0.54, and the item discrimination 0.40. Eventually, we applied the KR20 test to determine the reliability of the achievement test scores and calculated it as 0.70. It is satisfactory for the KR20 reliability coefficient to be above 0.50 in tests performed for the first time under 50 items (Salvucci et al., 1997). After all these processes, the preliminary application of the Measurement and Evaluation Basic Concepts (MEBC) Achievement Test was completed, and we shaped the test in its final form as 23 items with five options.

Participants, Data Collection and Analysis

This study is a two-stage research because of its methodological design. For this reason, there are various sampling methods for collecting qualitative and quantitative data at each stage. Ensuring the representativeness of the sample population is necessary to realize valid and reliable data collection processes in research (Miles et al., 2014). In

this respect, we used different sampling methods to form the correct sample groups according to the data needs.

Table 2

Type of Data Needed, Sampling Methods Used, and Participants.

Stages	Transactions Taken	Data Collection Tools	Sampling Method	N	Sample Characteristics
Stage 1	Describing the Current Situation with Qualitative Data Theoretical Framework Based on Qualitative Findings	Semi-structured interviews	Criteria Sampling	16 Teachers, 8 School Administrators	Adopted Criteria *(Different school types, teaching levels, branches, and professional experience)*
Stage 2	Developing Quantitative Data Collection Tool	Achievement test pre-application	Typical Case Sampling	132 Pre-service Teachers	Participant Features *(Typical two groups that took the measurement and evaluation course)*
	Program Evaluation Based on Quantitative and Qualitative Data	Observations	Study group	484 Pre-service Teachers	Participant Features *(6 different branches and classes where the measurement and evaluation course was carried out)*
		Achievement test actual application	Study group	364 Pre-service Teachers	Participant Features *(All groups that took the course during the observation period)*
		Semi-structured interviews	1. Maximum Diversity Sampling	36 Pre-service Teachers	Participant Features *(3 teacher candidates from each group who took the lesson online during the observation period and face-to-face in the previous term)*
			2. Study group	7 Lecturers	*(Lecturers teaching the course)*

During the initial phase of our study, we conducted semi-structured interviews with teachers and school administrators employed in public schools across Türkiye. Our sample selection involved individuals from various educational levels and types, including preschools, primary, secondary, and high schools. These schools were categorized into three groups to ensure a balanced representation of the student population, and participants were selected to ensure that each group did not exceed 25% of the total number of participants. In addition, teaching branches were classified into seven groups based on their field and professional characteristics, with at least three or four participants included from each branch.

In the second stage, we collected data from pre-service teachers who took the online measurement and evaluation course in four faculties of three different state universities in Türkiye and from the lecturers of the course. In this context, by observing the online courses for 14 weeks, we collected data from 484 pre-service teachers from six different branches (Science Education, Theology Bachelor's Program, Music Education, Preschool Education, Special Education, Guidance and Psychological Counseling) and classes in four faculties where the measurement and evaluation course was conducted by four instructors. We pre-applied the online achievement test on 132 pre-service teachers who took the course in the previous semester and applied the decisive test to 364 pre-service teachers from six different groups during online courses. In addition, we conducted semi-structured interviews with 36 pre-service teachers from different departments (Science Education, Theology Bachelor's Program, English Language Teaching, Mathematics Education, Music Education, Preschool Education, Special Education, Guidance and Psychological Counseling, Social Studies Education, Elementary Education, Turkish Language Teaching) who took the measurement and evaluation course and seven lecturers who led the course in the faculties where the observations were made. Four instructors were the academics whom we observed in their lessons.

NVIVO 10 was used in qualitative data analysis. In this context, we coded the interview and observation data through the related software and classified them into themes. We calculated the Cohen Kappa coefficient of agreement for the sake of consistency among coders to determine the reliability of these data and to see whether they correctly coded the contents of the research questions. The value we calculated varied between 0.63 and 1.00. According to Landis & Koch (1977), an agreement between 0.61 and 0.80 is satisfactory, while an agreement between 0.81 and 1.00 means a near-perfect fit.

We analyzed the quantitative data collected by the MEBC achievement test using the Test Analysis Program (TAP) and IBM SPSS Statistics 25. Initially, we assessed the normal distribution of the test scores by examining the skewness and kurtosis coefficients. To investigate potential variations in the test scores concerning factors such as department and gender, we employed Independent Samples T-Test, One-Way ANOVA, and Mann-Whitney U tests, contingent upon fulfilling the normal distribution assumption. We checked the Levene statistic regarding the homogeneity of variances before examining the test results for the department variable to which we applied One-Way ANOVA and found the significance level of the Levene value for the quotient variable to be lower than .05. For this reason, we used the Games-Howell Post Hoc Test to see which group averages differed significantly in the ANOVA test. Researchers commonly prefer the Games-Howell Test as it is used in cases where variances and sample numbers are not equal, is considered conservative to protect the margin of error, and provides the significance value between the mean scores of the groups (Armstrong et al. 2000). We also calculated the effect size for the variables with a significant difference according to the test results.

Findings

Opinions of Teachers and School Administrators on their Measurement and Evaluation Competencies

The interviewers mentioned three measurement and evaluation competency levels: "low, medium, and high."

Table 3

Descriptive Statistics on Measurement and Evaluation Competency Views of Teachers and School Administrators

Theme	Measurement and Evaluation Competence	Teachers			School Administrators		
		Participants	f	%	Participants	f	%
Codes	High	Teacher I Teacher M	2	12.5	Principal A Principal D Deputy Principal A	3	37.5
	Medium	Teacher A Teacher B Teacher C Teacher D Teacher E Teacher K Teacher L Teacher N Teacher 0 Teacher P	10	62.5	Principal B Principal E Deputy Principal B Deputy Principal C	4	50
	Low	Teacher F Teacher G Teacher J Teacher H	4	25	Principal C	1	12.5
	Total		16	100%		8	100%

Teachers consider themselves moderately competent at 62.5%, at a high level of 12.5%, and at a low level of 25% in measurement and evaluation. Respectively, school administrators consider themselves moderately competent at 50%, high at 37.5%, and low at 12.5%. This finding means school administrators assess themselves as more qualified than teachers in measurement and evaluation. However, it is probable to claim that teachers and school administrators assume themselves to be moderately sufficient.

Quality of the "Antecedents" Dimension of the Measurement and Evaluation Course

We examined the program's *antecedents* in terms of objectives, contents, readiness levels, expectations, attitudes, motivations of pre-service teachers, and instructive quality. The syllabuses and participant views have been the primary data sources of the study on this subject.

Table 4

Interview Codes and Descriptive Statistics Regarding the Quality of Program Antecedents

	Lecturers		Pre-service Teachers	
Theme	Codes	f	Codes	f
Quality of Antecedents	Limiting content criteria	7	Low student readiness	28
	Employment motivation source	7	Field incompatible content	25
	Low student readiness	6	Negative student attitude	19
	Positive lecturer attitude	6	Successful lecturers	12
	Negative student attitude	5	Low student motivation	10
	Limiting objectives criteria	4	Field-specific expectations	7
	Lecturer quality issues	3	Objectives unknown	3
	Total	38	Total	104

In the interviews, many pre-service teachers expressed that their measurement and evaluation knowledge infrastructure and readiness levels are insufficient. Similarly, lecturers consider the readiness level of teacher candidates inadequate. Especially, pre-service teachers with no background in mathematics, such as music and theology, stated that they had fears and prejudices that the course would challenge them at the beginning of the semester. According to the interviewees, these and similar cases are issues that create negative attitudes towards the course. Instructors think that measurement and evaluation, one of the basic skills in teachers' employment, is the most important source of motivation for the course. However, pre-service teachers consider this assumption a source of stress that reduces their motivation. In addition, although the instructors state that there are sometimes problems with the quality of the lecturers who conduct the course, the teacher candidates have positive views on this subject.

We examined the objectives in the syllabuses of the courses, considering the principles of goal selection and writing. Thus, we detected some problems, such as some objectives showing over one learning product, ending with different incompatible statements such as "will do, does, can use or uses" and pointing to the learning process instead of the learning product. We also received the opinions of the lecturers and pre-service teachers regarding the objectives of the program. Lecturers consider CHE regulations and Bologna and accreditation processes to be limiting factors in setting objectives. Pre-service teachers, on the other hand, avoided answering the questions about the objectives.

The observations and document examinations related to the course contents indicated that the weekly course contents presented to pre-service teachers largely matched the planned contents in syllabuses. However, since some contents were not completed in the scheduled times, the instructors could not present essential subjects such as "test and scale development," "statistical operations on the measurement results," "item analysis," and "evaluation of student achievement" and "graphics and standard test scores" in some groups. Instructors who conducted the course in over one field presented the same content in different groups. This situation reveals that there is no field-specific regulation in the course contents. At this point, the question of whether

the measurement and evaluation course meets the information needs of different teaching fields on this issue appears in mind. Both instructors and pre-service teachers criticized this issue. There were criticisms of pre-service teachers, especially in preschool, special education, music teaching, and guidance and psychological counseling (GPC) about the measurement and evaluation course's contents, were not suitable for their fields.

Quality of the "Transactions" Dimension of the Measurement and Evaluation Course

We included the distance education platforms, the teaching methods and techniques used, the duration of the course, the tools and materials, the communication and participation processes, and the measurement and evaluation methods in the *transactions* dimension of the program. In this context, we also received the lecturers' and pre-service teachers' opinions regarding this dimension.

Table 5

Interview Codes and Descriptive Statistics Regarding the Quality of Program Transactions.

	Lecturers		Pre-service Teachers	
Theme	Codes	f	Codes	f
Quality of Transactions	Limited lesson time	7	Methods and techniques	27
	Distance education limitations	6	Source-material	22
	Methods and techniques	6	Distance education limitations	21
	Source-material	3	Limited lesson time	15
	Interaction and participation	3	Homework	14
	Class size	2	Interaction and participation	11
			Class size	3
	Total	27	Total	113

In the research process, the measurement and evaluation course was mainly carried out as a live course over distance education platforms. We made course observations on the same platforms.

The most preferred platforms for live lessons were "Google Meet, Google Classroom, Adobe Connect, and Microsoft Teams," and for recorded lectures, "Microsoft PowerPoint Video." There were differences between universities in the preference for distance education platforms. "University 1" mainly preferred the Google Meet platform, "University 2" Google Meet and Adobe Connect platforms, and "University 3" Adobe Connect and Microsoft Teams platforms. In addition, 97.8% (5895 min.) of the lecturers taught their lessons live and 2.2% (180 min.) over the recording. Moreover, participation was higher in groups where attendance was compulsory.

Table 6

Observation Findings.

University	Groups	Distance Education Platform	Way of Conducting the Course	Attendance	Participation%	Teaching Methods and Techniques	*f*	Assessment Methods: *Midterm Exam* Action	$\bar{x}$	*Final Exam* Action	$\bar{x}$
1	GPC 1	Google Meet Google Classroom	875 min. live lesson	+	85.61	Lecturing Question-answer Discussion Case study	15 15 8 2	Exam	72	Exam	77
	Theology	Google Meet Google Classroom Microsoft PowerPoint Video	760 min. live, 120 min. recorded lesson	+	74.74	Lecturing Question-answer Discussion Case study	15 14 9 2	Exam	60	Exam	84
2	GPC 2	Google Meet Adobe Connect Microsoft PowerPoint Video	695 min. live, 60 min. recorded lesson	-	60.17	Lecturing Question-answer Discussion	14 13 7	Exam	70	Exam	62
3	Science Education	Adobe Connect	1275 min. live lesson	-	58.03	Lecturing Question-answer	14 13	Homework	68	Homework	69
	Music Education	Adobe Connect	1470 min. live lesson	-	58.36	Lecturing Question-answer	16 15	Homework	65	Homework	73
	Special Education and Pre-School Education	Microsoft Teams Adobe Connect	770 min. live lesson	-	62.26	Lecturing Question-answer Discussion	12 12 6	Homework	-	Homework	-

The pre-service teachers and lecturers also presented various views on the online course. Both pre-service teachers and lecturers stated that distance courses limit teaching and assessment efficiency to a large extent compared to face-to-face classes. Although the interviewees identified the distance education processes as inefficient, the instructors tried to make their lessons more productive by using various teaching methods in the live lectures. In the distance education processes, the instructors used lecturing, question-answer, discussion, and case study teaching methods. Apart from these methods, no teaching methods and techniques were preferred in the courses, and there was no remarkable difference in methodical choices among universities. In the interviews, the instructors expressed that the distance education processes and student

readiness level were effective in deciding on teaching methods and techniques. Pre-service teachers, like lecturers, expressed their views that the distance education process was a limiting factor for the teaching methods and techniques. While "University 1" and "University 2" preferred to make exams in measurement and evaluation, "University 3" preferred homework. According to the average scores, the group with the highest midterm exam scores is GPC 1, with a class average of 72, and the theology group has the lowest average with 60 points. In the final exam, inconsistent with the previous findings, theology was the group with the highest average of 84 points, while GPC 2 had the lowest score with an average of 62 points. The instructor did not share the scores of the special education and preschool education groups with the observer due to ethical concerns. This contradictory situation confirmed the necessity of a standard achievement test for all groups.

Quality of the "Outcomes" Dimension of the Measurement and Evaluation Course

In the interviews, the pre-service teachers were asked whether they considered themselves competent in this subject after completing the measurement and evaluation course.

Table 7

Descriptive Statistics on Measurement and Evaluation Opinions of Pre-Service Teachers

		Pre-service Teachers Taking the Course Online																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	f	%
Competence Opinions	Yes	+												+	+		+	+		+		+	7	33.3
	No		+	+	+	+	+	+	+	+	+		+			+					+		12	57.1
	Not Sure											+							+				2	9.5
		Pre-service Teachers Taking the Course Face to Face																					TOTAL	
		22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	f	%					f	%
	Yes		+	+	+	+			+	+	+						7	46.6					14	38.9
	No	+					+	+				+	+	+	+		7	46.6					19	52.8
	Not Sure															+	1	6.6					3	8.3

According to the table, 52.8% of the 36 pre-service teachers interviewed consider themselves unqualified in measurement and evaluation, 38.9% competent, and 8.3% did not express any opinion on this issue. In addition, pre-service teachers who took the course online consider themselves less competent. Another instrument for determining pre-service teachers' measurement and evaluation competencies was the achievement test.

The average scores of pre-service teachers who participated in the achievement test differ in terms of some variables.

Table 8

Distribution of MEBC Achievement Test Scores by Department and Gender Variables.

					Gender Variable			
	Department Variable		Normality Tests		Female		Male	
Departments	N	$\bar{x}$	Skewness	Kurtosis	N	$\bar{x}$	N	$\bar{x}$
GPC 1	65	71.31	-.611	.271	54	72.39	11	66.01
GPC 2	30	63.19	-.946	1.405*	19	66.82	11	56.92
Theology	75	56.75	.370	-.714	52	56.27	23	57.84
Science Education	35	58.39	-.738	.298	33	57.97	2	65.25
Music Education	33	51.51	.239	-.841	19	52.63	14	49.99
Pre-School Education	78	62.88	-.249	-.921	67	63.67	11	58.11
Special Education	48	73.47	-.611	-.406	44	73.43	4	73.92
Total	364	63.08	-.443	-.408	288	64.29	76	58.53

The special education and GPC 1 groups attained the top scores on the achievement test, achieving an arithmetic mean of more than 70 points. Specifically, the special education group obtained a score of 73.5, while the GPC 1 group achieved a score of 71.3. Conversely, the music education and theology groups obtained the lowest scores on the achievement test, with scores of 51.5 and 56.7, respectively. Furthermore, the scores on the achievement test varied based on the department and gender variables. We employed the Independent Samples One-Way ANOVA test to ascertain the significance of this difference concerning the department variable.

Table 9

Independent Samples One-Way ANOVA Results for the Department Variable.

	Sum of Squares	sd	Mean Square	F	p
Between Groups	17780.707	6	2963.451	12.313	.000
Within Groups	85921.040	357	240.675		
Total	103701.747	363			

According to the table, there is a significant difference in the scores from the achievement test in terms of the department variable, $F(6.357) = 12.313$, $p<.05$. We examined the Games-Howell Post Hoc Test results to determine which groups had this difference.

Table 10

Games-Howell Post Hoc Test Results Regarding the Department Variable

(1) Department	(2) Department	Mean Difference (1-2)	p	(1) Department	(2) Department	Mean Difference (1-2)	p
GPC 1	Science Education	12.92110*	.003	Special Education	GPC 1	2.16170	.954
	Music Education	19.80023*	.000		Science Education	15.08280*	.000
	Pre-School Education	8.42821*	.012		Music Education	21.96193*	.000
	Special Education	-2.16170	.954		Pre-School Education	10.58990*	.001
	Theology	14.55805*	.000		Theology	16.71975*	.000
	GPC 2	8.12205	.247		GPC 2	10.28375	.078
Science Education	GPC 1	-12.92110*	.003	Theology	GPC 1	-14.55805*	.000
	Music Education	6.87913	.696		Science Education	-1.63695	.999
	Pre-School Education	-4.49289	.845		Music Education	5.24218	.809
	Special Education	-15.08280*	.000		Pre-School Education	-6.12985	.257
	Theology	1.63695	.999		Special Education	-16.71975*	.000
	GPC 2	-4.79905	.914		GPC 2	-6.43600	.576
Music Education	GPC 1	-19.80023*	.000	PRD 2	GPC 1	-8.12205	.247
	Science Education	-6.87913	.696		Science Education	4.79905	.914
	Pre-School Education	-11.37203	.066		Music Education	11.67818	.160
	Special Education	-21.96193*	.000		Pre-School Education	.30615	1.000
	Theology	-5.24218	.809		Special Education	-10.28375	.078
	GPC 2	-11.67818	.160		Theology	6.43600	.576
Pre-School Education	GPC 1	-8.42821*	.012				
	Science Education	4.49289	.845				
	Music Education	11.37203	.066				
	Special Education	-10.58990*	.001				
	Theology	6.12985	.257				
	GPC 2	-.30615	1.000				

* p<0.05

The scores of GPC 1, science education, music education, preschool education, and theology groups from the MEBC Achievement Test differed significantly in favor of the GPC 1, and the scores of special education, science education, music education, preschool education, and theology groups differed in favor of special education. Since the achievement results significantly differed in terms of the department variable, we calculated the Eta square (η2) effect size and found the (η2) value to be 0.171. This result means that the department variable explains 17% of the variance in the scores obtained from the achievement test. After checking the normality scores, we used the Independent Samples t-Test and Mann-Whitney U tests to determine whether the achievement test scores differed significantly according to the gender variable.

Table 11

Independent Samples t-Test Results Regarding the Gender Variable

Departments	Variable / Gender	n	$\bar{x}$	ss	sd	T	p
GPC 1	Female	54	72.39	11.21	63	1.68	.096
	Male	11	66.01	12.42			
Theology	Female	52	56.27	15.15	73	-.39	.695
	Male	23	57.84	17.70			
Science Education	Female	33	57.97	17.04	33	-.59	.557
	Male	2	65.25	6.15			
Music Education	Female	19	52.63	21.09	31	.38	.701
	Male	14	49.99	16.60			
Pre-School Education	Female	67	63.67	16.80	76	.98	.327
	Male	11	58.11	20.15			
Special Education	Female	44	73.43	11.61	46	-.08	.935
	Male	4	73.92	7.94			

There was no significant difference in terms of gender variable in the scores obtained from the MEBC Achievement Test. Since the scores of the GPC 2 group did not indicate a normal distribution, we revealed the circumstance of the achievement test scores of this group in terms of the gender variable with the Mann-Whitney U Test.

Table 12

Mann-Whitney U Test Results Regarding Gender Variable of GPC 2 Group.

Department	Variable / Gender	n	Mean Rank	Sum of Ranks	U	p
GPC 2	Female	19	16.89	321.00	78.00	.251
	Male	11	13.09	144.00		

According to Mann-Whitney U Test results, there was no significant difference in terms of gender variable in the achievement test scores of the GPC 2 group.

Conclusion, Discussion, and Implications

According to the research findings, like current teachers and school administrators, pre-service teachers who took the measurement and evaluation course face-to-face and online did not consider themselves competent in measurement and evaluation. This situation reveals that pre-service measurement and evaluation training provided to teachers in Türkiye may be problematic regardless of whether it is online. From this point of view, it is conceivable to consider the research results in two different contexts. The first is the results regarding the measurement and evaluation program, and the other is the results about the measurement and evaluation training carried out with distance education.

As the relevant programs were developed before COVID-19, the developers of teacher training programs in Türkiye conceivably did not consider the distance education processes while creating the measurement and evaluation course. There are various problems with the programs, both related and unrelated to this situation. The most common ones were the problems in selecting and organizing the objectives and contents of the program. For example, there were problems in preparing the program's objectives regarding compliance with target selection and writing principles, especially in the Bologna information package web pages. This case reminds the opinion that the lecturers edited the related web pages without being taken seriously enough. There are similar difficulties in executing Bologna processes in higher education in different countries (Curaj et al., 2015; Pires Pereira et al., 2021; Valeyeva et al., 2015). In addition, the responsible actors did not sufficiently consider the institutional and individual characteristics in determining and regulating the objectives and contents. All faculties where the research occurred carried out similar processes, especially in determining the contents of the program. Ünver (2016) similarly expressed that they do not consider institutional differences in terms of instructors and opportunities in determining the course content of teacher training programs in Türkiye. In addition, faculties did not arrange these contents according to teaching fields and presented similar contents to all pre-service teachers.

According to the participants, the readiness level of pre-service teachers for measurement and evaluation was low. Various studies have also expressed the inadequacies of pre-service teachers in their pre-learning in Türkiye (Kozikoğlu & Kayan, 2018; Yenen & Durmaz, 2019). At this point, there may be problems regarding the admission of qualified students to teacher training programs. In this context, there are problems in the selection and employment of teacher candidates in Türkiye (Akdemir, 2013; Kutluca Canbulat, 2014).

Pre-service teachers in non-mathematics fields such as theology and music education had fears and prejudices towards the lesson. Güvendir & Özkan (2016) also reached similar findings in their study. In addition, the achievement test scores of pre-service teachers in these departments were lower, and the department variable considerably affected this issue. Ergül (2019) examined teachers' measurement and evaluation literacy and figured that visual arts, music, physical education, and theology teachers scored lower than computation-based departments, such as mathematics

teaching. The first two groups with the highest scores on the achievement test were special education and GPC 1. In GPC undergraduate programs, a statistics course is offered to pre-service teachers before the measurement and evaluation course in Türkiye (CHE, 2018). It is expectable that the groups who have taken the statistics course will get high scores on the measurement and evaluation achievement test. Furthermore, the special education department that received the highest score on the achievement test was the department that accepted students with the highest points in 2018 among the groups in the faculty (OSYM, 2021). This situation indicates that not considering the field variable in teacher candidates' measurement and evaluation training may lead to two problems. The first is that teaching fields with no mathematics background may have difficulties, especially in the statistical dimension of measurement and evaluation. The second problem is the uncertainty of how a statistical-based measurement and evaluation approach will work, especially in sports, art, or preschool fields, where performance-based learning should be evaluated.

Besides the problems related to the program's structure, we reached various positive and negative results regarding the online conduct of the measurement and evaluation course. The faculties conducted mainly the lessons on the "Google Meet and Classroom, Adobe Connect, and Microsoft Teams" distance education platforms, of which 97.8% were live. These platforms have been among the most preferred distance education and online conference platforms in recent years (Hurst, 2020; Koçoğlu, 2020). The pre-service teachers considered the instructors successful in teaching the course in distance education. In the study of Özer & Turan (2021), in which they took the opinions of pre-service teachers about distance education, the participants viewed distance education as successful in transferring theoretical contents. In addition, the participants stated that the online measurement and evaluation course was successful regarding the resources and materials. Different researchers have emphasized the importance of digital content in distance education in learning performance (Bae et al., 2009; Jena & Devi, 2020). This situation may be an indicator that the course was successful in these aspects and that both students and instructors were ready for distance education.

Besides their successful aspects, the online courses had some negative qualities. For example, both pre-service teachers and lecturers criticized the distance education process because of its limitations in terms of communication and participation. Kaysi & Aydemir (2017) noticed that distance education processes negatively affect students' communication with each other and the instructor. Compulsory attendance and computer and internet facilities in distance education processes are among the main factors affecting participation (Aydın & Dalkılıç, 2018; Durak et al., 2020). We observed that the participation level was higher, especially in the groups that took attendance. In addition, the participants think the course duration was too short for completing the contents, including contents specific to the field and practice. The subject of course duration in distance education is an issue that we should not discuss only in terms of time limitations. During the COVID-19 process, some studies emphasized that the course duration is vital in terms of interest and motivation, especially in lessons conducted in an atmosphere independent of the classroom environment (Özkara, 2021; Tiedt et al., 2021).

Participants think that distance education processes are limiting in teaching and assessment processes. However, appropriate teaching and assessment methods are critical for course success (Köse, 2012; Peters, 2013). Remarkably, the limitations in measurement and evaluation are worrisome for a course whose primary purpose is to provide relevant skills. In this context, some faculties made simultaneous exams for measurement and evaluation, while others used housework as a measurement and evaluation tool. Although pre-service teachers consider homework necessary and effective in the teaching process, they think it was not an appropriate method for measurement and evaluation. However, the reason some faculties preferred homework instead of exams might be the difficulties caused by distance education processes. It is because exam applications in the distance education process have problems, especially in ensuring exam security (Al-Shalout et al., 2021; Kınalıoğlu & Güven, 2011). For this reason, instructors may have had difficulties in evaluating student success in going beyond homework.

Not considering different needs and contexts in the measurement and evaluation program development process can make things even more difficult in distance education. Some suggestions that are supposed to contribute to the acquisition of measurement and evaluation competencies in distance education, which might become indispensable in the coming years, are:

1. Responsible individuals and institutions should consider the institutional characteristics and differences in the development and implementation of the programs and should not ignore the possibility of conducting the relevant programs online.
2. There must be clear criteria that do not confuse in determining the objectives and contents of the program, and the instructors who are experts in the field should be more trusted and empowered.
3. While arranging the objectives and contents of the program, those responsible should consider the teaching areas and student characteristics, such as readiness, interest, attitude, and motivation to terminate the fears and prejudices of pre-service teachers in different teaching fields.
4. The course duration should be arranged to allow the presentation and application of field-specific and general theoretical knowledge, considering the dynamics of distance education processes.
5. The lecturers should increase the variety of teaching methods and techniques to ensure communication, interaction, attendance, and participation in both distance and face-to-face courses.
6. Responsible institutions should serve technical support to students with limited technological opportunities in distance education.
7. The lecturers should use measurement and evaluation instruments that will not conflict with what they teach in the lessons.

Acknowledgements

This research was not funded by any organization. The study was produced from the Ph.D. dissertation named “Evaluation of measurement and evaluation in education curriculum” which was written by the first and corresponding author. The second author is the academic advisor of the first author. In this context, the first author's contribution to the study may be considered 60% and the academic advisor's 40%.

Statement of Responsibility

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in the study. In this context, all ethical and legal responsibilities related to the study belong to the researchers.

Conflicts of Interest

First Author declares that he/she has no conflict of interest. Second Author declares that he/she has no conflict of interest.

Author Bios:

Muhammed Akıncı is currently working as a faculty member at Recep Tayyip Erdoğan University, Faculty of Education. He has academic and administrative experience in subjects such as program development in education, teaching strategies, methods and techniques, evaluation of teaching and program, quality management and accreditation in education, program standards, and adaptation of program evaluation standards to Turkish. In this context, he is actively engaged in research related to Curriculum Development and Evaluation, concentrating on developing and implementing standards for evaluating educational practices and designing curricula. Moreover, his scholarly inquiries extend to the fields of teacher education, particularly concerning the acquisition of professional competencies. He has been giving courses on research methods in education and curriculum development since 2020.

Erdoğan Köse is a Professor at Akdeniz University, Faculty of Education in Turkey. He has academic experience in the fields and subjects of program development and evaluation in education, educational sociology, teaching principles and methods, and measurement and evaluation. As deputy dean and head of department, he also has administrative experiences in different faculties of education in Turkey. He also served as an academic advisor to many graduate and doctoral students.

References

Ada, S., & Baysal, Z. N. (2013). *Pedagojik-androgojik formasyon ve Türkiye'de öğretmen yetiştirme.* Ankara: Pegem Academy Publication.

Akdemir, A. S. (2013). A history of teacher training programmes and their problems in Turkey. *Electronic Turkish Studies*, *8*(12), 15-28.

Al-Shalout, M. I., Rasmi, M., & Hassan, M. A. (2021). Evaluation of E-exam during Covid-19. *Psychology and Education Journal*, *58*(1), 4604-4612.

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, *39*(2), 113-123.

Armstrong, R. A., Slade, S. V., & Eperjesi, F. (2000). An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthalmic and Physiological Optics*, *20*(3), 235-241.

Atılgan, H. (2018). Değerlendirme ve not verme. In Atılgan, H. (Ed.), *Eğitimde Ölçme ve Değerlendirme* (11th Edition) (pp. 315-362). Anı Publication.

Aydın, Ö., & Dalkılıç, F. (2018). University student attendance management system. *Journal of Science and Engineering*, *20*(60), 863-875.

Bae, S. H., Kim, J. H., & Lim, Y. W. (2009). A study on the effect of multimedia online learning contents on learner's performance. *Journal of Korea Society of Digital Industry and Information Management*, *5*(1), 127-139.

Barthakur, A., Joksimovic, S., Kovanovic, V., Richey, M., & Pardo, A. (2022). Aligning objectives with assessment in online courses: Integrating learning analytics and measurement theory. *Computers & Education*, 104603.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı.* (3rd Edition). Ankara: Pegem Academy Publication.

Chan, C. K., & Luk, L. Y. (2021). A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment & Evaluation in Higher Education*, 1-15.

Christ, T. J., & Desjardins, C. D. (2018). Curriculum-based measurement of reading: An evaluation of frequentist and Bayesian methods to model progress monitoring data. *Journal of Psychoeducational Assessment*, *36*(1), 55-73.

Cohen, L., Manion, L., & Morrison, K. (2007). Research methods in education. *Routledge*.

Conrad, D., & Openo, J. (2018). *Assessment strategies for online learning: Engagement and authenticity*. Athabasca University Press.

Council of Higher Education. (2018). *Yeni öğretmen yetiştirme lisans programları.* Retrieved July 15, 2020, from https://www.yok.gov.tr/kurumsal/idari-birimler/egitim-ogretim-dairesi/yeni-ogretmen-yetistirme-lisans-programlari

Council of Higher Education. (2007). *Öğretmen yetiştirme ve eğitim fakülteleri (1982-2007).* Retrieved November 26, 2020, from https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Ogretmen-Yetistirme/2.pdf

Council of Higher Education. (1998). *Eğitim fakültesi öğretmen yetiştirme lisans programları.* Retrieved November 26, 2020, from https://www.yok.gov.tr/Documents/Yayinlar/Yayinlarimiz/egitim-fakultesi-ogretmen-yetistirme-lisans-programlari-mart-1998.pdf.

Creswell, J. W. (2015). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (5th Edition). Pearson Prentice Hall.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd Edition). SAGE Publications, Inc.

Curaj, A., Deca, L., Egron-Polak, E., & Salmi, J. (2015). *Higher education reforms in Romania: Between the Bologna process and national challenges*. Springer Nature.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272.

Durak, G., Çankaya, S., & İzmirli, S. (2020). Examining the Turkish universities' distance education systems during the COVID-19 pandemic. Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, *14*(1), 787-809.

Duman, S. N. (2019). *Hizmet öncesi öğretmen eğitiminde öğretmenlik meslek bilgisi derslerinin değerlendirilmesi* [Doctoral dissertation]. Gazi University.

EPDAD. (2020). *EPDAD teacher education standards.* Retrieved May 4, 2020, from https://epdad.org.tr/icerik/epdads-teacher-education-standards

Ergül, A. Ö. (2019) *Examination of teachers' assessment literacy levels in terms of some variables* [Master's thesis]. Hacettepe University.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines*. (4th Edition). Pearson Education Inc.

Fuchs, L. S. (2017). Curriculum-based measurement as the emerging alternative: Three decades later. *Learning Disabilities Research & Practice*, 32(1), 5-7.

Graves, M. N. (2010). *Curriculum-based measurement attitudes and training practices in teacher education programs* [Doctoral dissertation]. Indiana University.

Grayson, T. E. (2018). Program evaluation, performance measures, and evaluability assessment in higher education. In Secolsky, C., & Denison, D. B. (Ed.), *Handbook on measurement, assessment, and evaluation in higher education* (2nd Edition) (pp. 616-639). Routledge.

Güler, N. (2018). *Eğitimde ölçme ve değerlendirme*. (12th Edition). Pegem Academy Publication.

Güvendir, M. A., & Özkan, Y. Ö. (2016). Determining the perceptions of prospective teachers in relation to measurement and evaluation course through metaphors. *Dumlupınar University Journal of Social Sciences*, *47*, 91-105.

Harmon, R. L., Sheehy, L. M., & Davis, D. M. (1998). The utility of external performance measurement tools in program evaluation. *Rehabilitation Nursing*, *23*(1), 8-11.

Harris, C. (2017). *A program evaluation of a technology based formative assessment for algebra readiness* [Doctoral dissertation]. The College of William and Mary in Virginia).

Holmes, D. S. (1971). The teaching assessment blank: A form for the student assessment of college instructors. *The Journal of Experimental Education*, *39*(3), 34-38.

Hurst, E. J. (2020). Web conferencing and collaboration tools and trends. *Journal of Hospital Librarianship*, *20*(3), 266-279.

ISO, (2021). *Teaching standards.* Retrieved March 2, 2022, from https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100354.pdf

JCSEE. (2018). *Checklist of the program evaluation standards statements.* Retrieved November 26, 2020, from https://wmich.edu/evaluation/checklists

Jena, A. K., & Devi, J. (2020). Lockdown area of COVID-19: How does cartoon based e-contents effect on learning performance of Indian elementary school students with ADHD. *Online Submission*, *8*(4), 189-201.

Karakuş, M. & Turhan Türkkan, B. (2017). Investigating the needs for measurement and evaluation course: A case study on English language teaching program. *Journal of Education and Training Studies, 5*(4), 227-239

Karaman, P. & Şahin, Ç. (2014). Öğretmen adaylarının ölçme değerlendirme okuryazarlıklarının belirlenmesi. *Ahi Evran Ünv. Kırşehir Eğitim Fakültesi Dergisi (KEFAD), 15*(2), 175-189.

Kaysi, F., & Aydemir, E., (2017). Assessment of interaction dimensions in distance education processes. *The Journal of Social Science, 4*(11), 778-790.

Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *Journal of Online Learning and Teaching*, *8*(3), 198-208.

Kınalıoğlu, İ. H., & Güven, Ş. (2011). *Issues and solutions on measurement of student achievement in distance education*. Proceedings of XIII. Academic Informatics Conference, Turkey, 637-644.

Koç, A. (2019). *Fen bilgisi öğretmen adaylarının ölçme değerlendirme okuryazarlık düzeylerinin belirlenmesi* [Master's thesis]. Akdeniz University.

Sabancı, O. & Yazıcı, K. (2017). Öğretmen adaylarinin ölçme ve değerlendirmeye yönelik yeterlik algilarinin incelenmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, *7*(1), 128-153.

Sevimel Şahin, A. (2019). *İngiliz dili öğretmen adaylarının yabancı dilde ölçme değerlendirme okur yazarlığının incelenmesi* [Doctoral dissertation]. Anadolu University.

Koçoğlu, E. (2020). Analysis of course transfer systems used in distance education. *The Eurasia Proceedings of Educational and Social Sciences*, *18*, 180-184.

Kozikoğlu, İ., & Kayan, M. F. (2018). Changes in education faculties from the viewpoints of academic staff: student profile, base score application and student selection. *Kastamonu Education Journal*, *26*(6), 1863-1873.

Köse, E. (2012). Öğretimde ölçme ve değerlendirmenin planlanması. In Karip, E. (Ed.), *Ölçme ve değerlendirme* (pp. 123-152). Pegem Academy Publication.

Kutluca Canbulat, A. N. (2014). An assessment related to instructors' views about teachers eligibility and teacher candidates' selection process. *Cumhuriyet International Journal of Education*, *3*(3), 1-11.

Landis, J, R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

McDowell, C. L. (1992). Standardized tests and program evaluation: Inappropriate measures in critical times. *New Directions for Program Evaluation*, *1992*(53), 45-54.

Mehrens W. A., & Lehmann I. J. (1991). *Measurement and evaluation in education and psychology* (4th Edition). Wadsworth/Thomson Learning

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice*. (2nd Edition). Guilford Publications.

Miles, H., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd Edition). SAGE Publications, Inc.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10thEdition). Pearson Education, Inc.

Ministry of National Education. (2019). *2023 education vision.* Retrieved July 15, 2020, from http://2023vizyonu.MNE.gov.tr/

Ministry of National Education. (2017). *General competencies of the teaching profession.* Ankara: General Directorate of Teacher Training and Development.

Ministry of National Education. (1980). *Journal of papers numbered 2064*. Ankara: Ministry of National Education General Directorate of Publications and Printed Educational Materials.

Ministry of National Education. (1975). *Journal of papers numbered 1840*. Ankara: Ministry of National Education General Directorate of Publications and Printed Educational Materials.

Mislevy, R. J. (2018). On measurement in educational assessment. In Secolsky, C., & Denison, D. B. (Ed.), *Handbook on measurement, assessment, and evaluation in higher education* (2nd Edition) (pp. 37-64). Routledge.

Morrow Jr, J. R., Jackson, A. W., Disch, J. G., & Mood, D.P. (2000). *Measurement and Evaluation in Human Performance* (2nd Edition). Human kinetics.

Noble, A. (2012). *Examining the instructional utility of curriculum-based measurement in achievement evaluations* [Doctoral dissertation]. East Carolina University.

OSYM (2021). 2018-YKS *Yerleştirme sonuçlarına ilişkin sayısal bilgiler.* Retrieved August 27, 2021, from https://www.osym.gov.tr/TR,15288/2018-yks-yerlestirme-sonuclarina-iliskin-sayisal-bilgiler.html

Özer, S., & Turan, E. Z. (2021). Opinions of prospective teachers about distance education due to COVID-19. *Turkish Studies - Education, 16*(2), 1049-1068.

Özkara, B. Ö. (2021). Determining the optimal duration of a single lecture in distance education using facial analysis of instructors. *Turkish Online Journal of Educational Technology-TOJET*, *20*(2), 35-43.

Patton, M. Q. (2015). *Qualitative research and evaluation methods*. (4th Edition). SAGE Publications.

Pavlič, L., Beranič, T., Brezočnik, L., & Heričko, M. (2022). Towards a novel catalog of assessment patterns for distant education in the information technology domain. *Computers & Education*, *182*, 104470.

Peters, O. (2013). *Learning and teaching in distance education: analyses and interpretations from an international perspective*. Routledge.

Pires Pereira, Í. S., Fernandes, E. L., Braga, A. C., & Flores, M. A. (2021). Initial teacher education after the Bologna process. Possibilities and challenges for a renewed scholarship of teaching and learning. *European Journal of Teacher Education*, 1-29.

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

QCT. (2021). *Australian professional standards for teachers.* Retrieved May 3, 2022, from https://www.qct.edu.au/standards-and-conduct/professional-standards

Raje, S., & Stitzel, S. (2020). Strategies for effective assessments while ensuring academic integrity in general chemistry courses during COVID-19. *Journal of Chemical Education*, *97*(9), 3436-3440.

Salvucci, S., Walter, E., Conley, V., Fink, S., & Saba, M. (1997). Measurement error studies at the National Center for Education Statistics (NCES). Washington D. C.: U. S. Department of Education.

Safrit, M. J. (1990). Measurement and evaluation curricula in professional physical education programs—Current trends. *Journal of Physical Education, Recreation & Dance*, *61*(3), 39-41.

Scriven, M. (1981). *Evaluation thesaurus* (3rd Edition). Edge Press.

Secolsky, C., & Denison, D. B. (Ed.) *Handbook on measurement, assessment, and evaluation in higher education* (2nd Edition). Routledge.

Shapiro, E. S., & Gebhardt, S. N. (2012). Comparing computer-adaptive and curriculum-based measurement methods of assessment. *School Psychology Review*, *41*(3), 295-305.

Stake, R. E. (2014). Information science and responsive evaluation. *E-Learning and Digital Media, 11*(5), 443-450.

Stake, R. E. (2013). Responsive evaluation IV. In Alkin, M. C. (Ed.) *Evaluation roots. A wider perspective of theorists' views and influences* (pp. 189-197). SAGE Publications, Inc.

Stake, R. E. (2011). Program evaluation particularly responsive evaluation. *Journal of Multi-Disciplinary Evaluation, 7*(15), 180-201.

Stake, R. E. (2005). Qualitative case studies, In Norman K. D. ve Yvonna S. L. (Ed.) *The sage handbook of qualitative research* (3rd Edition) (pp. 443- 465). SAGE Publications, Inc.

Stake, R. E. (1967). *The countenance of educational evaluation*. Department for Exceptional Children, Gifted Children Section.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (Ed.), *Evaluation models: Viewpoints on*

educational and human services evaluation (pp. 117-141). Kluwer Academic Publishers.

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications*. San Francisco: Josey-Bass.

Tang, L. (2012). *A formative program evaluation of treatment integrity practices, assessments and attitudes within a specialized school setting* [Doctoral dissertation]. University of Massachusetts Amherst.

Tiedt, J. A., Owens, J. M., & Boysen, S. (2021). The effects of online course duration on graduate nurse educator student engagement in the community of inquiry. *Nurse Education in Practice*, *55*, 1-8.

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *International Scholarly Research Notices*, *2013*, 1-29.

Tusing, M. E., & Breikjern, N. A. (2017). Using curriculum-based measurements for program evaluation: expanding roles for school psychologists. *Journal of Applied School Psychology*, *33*(1), 43-66.

UK. Department for Education. (2021). *Teachers' standards.* Retrieved May 3, 2022, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1040274/Teachers__Standards_Dec_2021.pdf

US. Department for Education. (2021). *Professional teaching standards.* Retrieved May 3, 2020, from https://www.ed.gov/category/keyword/national-board-professional-teaching-standards

Ünver, G. (2016). Connection between the theory and practice in pre-service teacher education programs in Turkey. *Journal of Higher Education and Science*, *1*, 61-70.

Valeyeva, N. S., Kupriyanov, R. V., & Valeyeva, E. R. (2015, September). Results and challenges of Russia's integration into Bologna Process. In *2015 International Conference on Interactive Collaborative Learning (ICL)* (pp. 404-406). IEEE.

Veal, M. L. (1990). Measurement and evaluation curricula in professional physical education preparation programs - A view from the practitioner. *Journal of Physical Education, Recreation & Dance*, *61*(3), 36-38.

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.

Yenen, E. T., & Durmaz, A. (2019). Investigation of teacher candidates' readiness levels for teaching profession. *OPUS International Journal of Society Researches, 13*(19), 922-940.

Zhang, Z. (1996). *Teacher assessment competency: A Rasch model analysis*. Paper presented at the Annual Meeting of the American Educational Research Association, April 8-12, New York. (ERIC Reproduction Service No. ED 400 322).