PES
www.ijpes.com
International Journal of Psychology and Educational Studies

Developing the Measurement and Evaluation Attitude Scale for Physical Education and Sports Teachers*

Turan Başkonuş[1], Fikret Soyer[2]

[1] *Republic of Turkey Ministry of National Education, Turkey*
[2] *Balıkesir University, Faculty of Sports Science, Balıkesir, Turkey*

ARTICLE INFO

Article History:
Received 25.05.2020
Received in revised form 15.08.2020
Accepted 29.08.2020
Available online 28.09.2020

ABSTRACT

The aim of this study is to develop a scale that measures the attitudes of physical education and sports teachers towards measurement and evaluation. In this study, scale development principles and steps of DeVellis (2017) were used. Initially, a literature review was conducted, 19 physical education and sports teachers were interviewed in written form and focus group discussion were held with 7 teachers. Certain statements were converted into a format that can be used in the scale form and an item-pool of 73 items was created including positive and negative items. Expert opinions were taken for the draft items, corrections were made in the item pool in line with these expert opinions, and the initial form the scale was created including 58 items. This draft scale was applied to 50 physical education and sports teachers. As a result of the pilot implementation, the scale was put into its final draft form with a total of 58 items, 17 of which were negative. As the conclusion of the reliability and validity tests applied to the Measurement and Evaluation Attitude Scale for Physical Education and Sports Teachers, 34 items were removed from the draft scale and the final form of the scale included 4 factors and 24 items with a variance value of 46.545. The Cronbach Alpha reliability coefficient of the scale was calculated as 0.85.

Keywords:
Physical education and sports teacher, measurement, evaluation, attitude

1. Introduction

Measurement and evaluation practices are an important part of the education and training processes. Measurement and evaluation reveals to what extent the learners achieved the targeted learning achievements, and to what extent these targeted achievements were accomplished in terms of the teacher. This process, which is an indispensable part of education, will be more meaningful when it is thoroughly accepted and embraced by the teachers. The first priority in ensuring the effectiveness of measurement and evaluation practices is the teacher. Teachers are required to plan the education process, manage the process, measure and evaluate the success. Therefore, it is important for teachers to have a standpoint about the tools and methods of measurement and evaluation and to believe that different methods are necessary.

Quilter (1998) stated that teachers use a variety of measurement and evaluation methods to make decisions in classroom activities, and that measurement and evaluation knowledge levels and attitudes of teachers are correlated. In this context, the teacher who has a positive attitude towards the measurement will try to achieve the measurement and evaluation competences, and this will ensure that different measurement and evaluation tools and methods are included in the process. According to İnceoğlu (2011: 16), attitude is the internal tendency of the person to react to an object or situation in the outside world. In other words, it is the behavior that the person is expected to display in the face of an event or object. Turgut (1995: 155) defined

*This study was produced from Turan Başkonuş's doctoral thesis. It was presented as an oral presentation at ERPA International Congresses on Education, 10-12 April 2020

[1] Corresponding author's address: Republic of Turkey Ministry of National Education, Ankara, Turkey
Telephone: +90-386-2131126
e-mail: baskonusturan@gmail.com

http://dx.doi.org/10.17220/ijpes.2020.04.010

attitude as a person's behavior displayed towards a situation, object or person in a positive or negative way. In another definition, attitude is defined as the learned tendency whether to use the right of choice against a psychological element (Tekindal, 2009: 178). Based on all these definitions, we can say that attitudes cannot be observed directly but can be revealed by one's behaviors, therefore, while measuring the attitudes of individuals, their behaviors towards any subject or object should be observed. Keller (2017: 19) reported that success requires more than having a great attitude, success in life starts and ends in attitude, and other principles cannot be put into action without a positive attitude. According to Şişko and Demirhan (2002:205), if a person has a positive attitude towards an object or situation, she/he approaches that object or situation, but if she/he has a negative attitude towards an object or situation, she/he walks away from that object or situation, giving negative reactions. But attitudes can change, new ones can be gained in time.

From the standpoint of the teacher, the attitudes of teachers about measurement and evaluation can be revealed by the methods they follow in the process and by the measurement and evaluation practices they use. For example, we can say that teachers who use different measurement and evaluation tools and methods in the process and who include diversity and flexible practices in the measurement and evaluation process have a positive attitude towards measurement and evaluation; however, we can say that the teachers, who do not include different measurement and evaluation tools and methods, who attempt to use the same and standard measurement and evaluation tool for everyone, and who complete the education process with methods comprising of only a verbal exam and a written exam of open-ended-questions, have a negative attitude towards measurement and evaluation. In the education process, traditional (written exams, oral exams, right-wrong exams, gap-filling etc.) and alternative assessment (rubric, project and performance assignments, observation forms, checklist, self and peer assessment forms, etc.) and evaluation methods are used in physical education and sports lessons as in other lessons. The assessment and evaluation process and these assessment and evaluation methods included in this process are an integral part of the education process.The assessment and evaluation process becomes even more meaningful when it is fully accepted by the teacher. Issues such as teachers' perspectives towards assessment and evaluation tools and methods, their attitudes towards the assessment and evaluation process, and their belief in the necessity of different methods are important. Therefore, it is important to determine the attitudes of teachers regarding measurement and evaluation. With this research, it is aimed to develop a scale that can measure the attitudes of physical education and sports teachers regarding measurement and evaluation, which is an indispensable element of the education system. There is no previous study in the literature presenting a scale to measure the attitudes of physical education and sports teachers regarding measurement and evaluation, which reveals the significance of this study.

2. Method

In this section, the scale development process is explained in detail.

2.1. Scale Development Process

In order to determine the statements that will be included in the scale development process, the scale development resources (Tavşancıl 2014; Baykul 2015; Özgüven 2015; Erkuş 2016; DeVellis 2017) were examined in the literature. For this purpose, in order to measure the attitudes of physical education and sports teachers towards measurement and evaluation, a Likert-type scale was developed based on the scale development principles and steps of DeVellis (2017). The following steps were followed during the scale development process.

- ✓ Clearly Identifying the Structure to be Measured
- ✓ Determining the Measurement Format and Creating the Item Pool
- ✓ Review of the Item Pool by the Experts (Determination of coefficient of concordance)
- ✓ Pilot Application
- ✓ Application of Items to the Scale Development Sample
- ✓ Evaluation of Items and Putting the Scale into the Most Appropriate Form

2.1.1. Clearly identifying the structure to be measured. This measurement tool was developed to reveal the attitudes of physical education and sports teachers about measurement and evaluation.

2.1.2. Determining the measurement format and creating the item pool. A literature review was conducted about the measurement and evaluation in the physical education and sports course, and 19 physical education and sports teachers were asked to write what they think about measurement and evaluation in physical education and sports courses, whether they believe in the necessity of measurement and evaluation, and what their thoughts are on the related questions with reason. In addition, focus group interviews were conducted with 7 (3 female - 4 male) teachers. The focus group interview was conducted by a field specialist who has a proficiency in the subject (with a doctoral degree in sports sciences). The interviews lasted 40 minutes and the participants were asked open-ended questions concerning what they think about measurement and evaluation in physical education and sports course, the problems they face and their suggestions for these problems. Recordings of interviews with physical education and sports teachers were examined for several times. As a result, the specified statements were converted into a format that could be used in the scale form and a Likert-type draft form was designed, creating an item pool of 73 items with positive and negative statements.

2.1.3. Review of the item pool by the experts. At this stage, the draft form was sent to the field experts (1 professor of educational science, 1 psychological counselor, 1 professor of sports sciences) and a linguist (faculty member) for evaluating both the content validity of the scale and determining whether the prepared items were suitable for the measurement tool. In the draft form sent to experts, scale items were designed as "appropriate, appropriate after corrections, and inappropriate'. The coefficient of concordance among the raters was calculated using the Miles and Huberman (1994) consensus formula. The draft form including 73 items was reduced to 58 items after determining the coefficient of concordance among the raters and the coefficient of concordance was determined as 93% for the agreement among the raters for the 58-item draft form. This result points that the items found in the draft form will reveal the attitudes of physical education and sports teachers towards measurement and evaluation.

2.1.4. Pilot application. The draft scale was applied to 50 physical education and sports teachers outside the sample group within the scope of validity for usefulness. As a result of the pilot implementation, the draft form of the scale now included a total of 58 items, 17 of which were negative.

2.1.5. Application of items to the scale development sample. Draft evaluation form was implemented to 222 physical education and sports teachers, who were attending the "Training Program for Educators for the Promotion of the Updated Curriculums" that hosted physical education and sports teachers from every province and held in Aydın province in the 2016-2017 academic year by the Ministry of Education.

Researcher views (Comrey & Lee, 1992; Kline, 1994; as cited in Çokluk, Şekercioğlu & Büyüköztürk, 2016: 206: Erkuş, 2016: 59) and the KMO values in Table 1 indicate that the sample size is sufficient for factor analysis.

Table 1. *Interpretation of the KMO Test Results*

KMO value	Interpretation
0,90	Perfect
0,80	Very Well
0,70	Well
0,60	Medium
0,50	Weak
Below 0,50	Unacceptable

Source: Kalaycı, 2014: 322; Tavşancıl, 2002: 50; Altunışık, Coşkun, Bayraktaroğlu & Yıldırım, 2005: 217

2.1.6. Evaluation of items and putting the scale into the most appropriate form. Statistical analyses were conducted following the application of the items to the scale development sample, the items were revised for the last time and the scale was given its final form for the factor analysis.

Before the decided statistical analyses, revision of the data is of vital significance for the research to obtain valid and reliable results. There are a number of objectives for reviewing research data before multivariate analyses (Mertler & Vannatta, 2005: as cited in Çokluk et al., 2016: 9). These are; examining the accuracy of the data, examining the loss data, elimination of the extreme values, and satisfaction of the assumptions. In this context, the accuracy of the data was examined, all data entered into SPSS were checked and 5 scales

were excluded from the analysis due to incorrect data entry, and 3 scales due to loss values in which almost half the items were left blank.
Before applying statistical analyses, it is necessary to test whether certain assumptions are satisfied. Some multivariate analyses have their own unique assumptions (Çokluk et al., 2016: 10).

The processes performed at this stage are listed below:

2.1.6.1. ***Normality test.***

Table 2. *Normality Test Results of the MEASPEST (draft scale)*

		Statistics	Standard deviation
Mean		201,96	1,224
Mean value at a confidence interval of 95%	Lower limit	199,54	
	Upper limit	204,37	
5% Trimmed Mean		201,55	
Mode		202,00	
Median		201,50	
Variance		320,78	
Standard deviation		17,910	
Minimum		160,00	
Maximum		247,00	
Range		87,00	
Skewness		,293	,166
Kurtosis		-,388	,331

When Table 2 is examined, it is observed that the arithmetic mean of 214 data for the MEASPEST is 201.96 and the standard deviation is 17,910. In addition, the lower and upper limits range from 199.54 to 204.37 in the 95% confidence range. The median for this series is 201.50. According to Kalaycı (2005: 6) and Howitt and Cramer (2011), the mod, median and arithmetic mean must be equal in order for the data to show a normal distribution. According to Büyüköztürk (2014: 40), the basics of the analysis is that the scores should not extremely deviate from the normal. The fact that the coefficient of skewness remains within the limits of -1 to +1 can be interpreted that the points do not show a significant deviation from the normal distribution. According to George and Mallery (2010: 409) and Blest (2003: 175), if the skewness and kurtosis values are between +2 and -2, it points that the data shows normal distribution. The mod, median and arithmetic mean of the draft measuring tool applied to 214 people during the scale development process and skewness (0.293) and kurtosis (-0.388) values indicate that the distribution of the data is close to normal.

2.1.6.2. ***t-Test for the significance of the differences between the item mean scores of the lower 27% and upper 27% groups formed according to their total scores.***

Table 3. *T-Test Results of the Lower 27% and Upper 27% Groups which Responded the MEASPEST (draft scale) for the Item Mean Scores*

Items	Groups	N	Mean	SD	t	p
Item 1	Upper	58	4,3276	,82479	5,982	,000
	Lower	58	3,5000	,65561		
Item 2	Upper	58	4,5690	,59566	10,099	,000
	Lower	58	3,3793	,67089		
Item 3	Upper	58	4,3966	,74785	8,025	,000
	Lower	58	3,3276	,68538		
Item 4	Upper	58	3,8621	1,08334	7,118	,000
	Lower	58	2,5690	,86068		
Item 5	Upper	58	4,2586	,80699	5,678	,000
	Lower	58	3,4138	,79548		

Table 3. *(Continue)*

Item 6	Upper	58	4,2759	,79014	6,814	,000
	Lower	58	3,2414	,84418		
Item 7	Upper	58	3,9138	1,08068	6,150	,000
	Lower	58	2,6897	1,06305		
Item 8	Upper	58	4,0517	,88699	4,221	,000
	Lower	58	3,3621	,87255		
Item 9	Upper	58	4,2759	,66999	5,183	,000
	Lower	58	3,5000	,92243		
Item 10	Upper	58	4,3103	,79927	6,476	,000
	Lower	58	3,3103	,86261		
Item 11	Upper	58	4,1034	1,00332	6,493	,000
	Lower	58	2,8276	1,11036		
Item 12	Upper	58	4,0345	,72464	5,600	,000
	Lower	58	3,2931	,70109		
Item 13	Upper	58	4,1379	,75969	5,301	,000
	Lower	58	3,3448	,84918		
Item 14	Upper	58	3,2414	1,30196	2,971	,004
	Lower	58	2,6207	,91436		
Item 15	Upper	58	3,8966	1,00332	4,785	,000
	Lower	58	3,0345	,93594		
Item 16	Upper	58	2,4655	1,04656	-1,502	,136
	Lower	58	2,7241	,79014		
Item 17	Upper	58	4,2586	,76228	5,926	,000
	Lower	58	3,4138	,77311		
Item 18	Upper	58	4,2586	,57918	6,478	,000
	Lower	58	3,4655	,73067		
Item 19	Upper	58	3,3448	1,35791	3,325	,001
	Lower	58	2,6552	,80681		
Item 20	Upper	58	3,8793	,83933	5,902	,000
	Lower	58	3,0517	,66021		
Item 21	Upper	58	4,2414	,68340	9,439	,000
	Lower	58	3,0172	,71307		
Item 22	Upper	58	2,7931	1,15103	-,834	,406
	Lower	58	2,9483	,82552		
Item 23	Upper	58	4,2241	,89918	5,256	,000
	Lower	58	3,3966	,79339		
Item 24	Upper	58	4,4138	,59337	7,096	,000
	Lower	58	3,3966	,91651		
Item 25	Upper	58	4,1724	,75249	6,497	,000
	Lower	58	3,2586	,76228		
Item 26	Upper	58	3,6379	1,13475	2,871	,005
	Lower	58	3,1379	,68693		
Item 27	Upper	58	3,0345	1,52137	1,493	,138
	Lower	58	2,6897	,88272		
Item 28	Upper	58	4,2759	,52292	8,232	,000
	Lower	58	3,2414	,80154		
Item 29	Upper	58	4,1552	,69590	6,509	,000
	Lower	58	3,2586	,78495		
Item 30	Upper	58	4,2759	,69568	7,981	,000
	Lower	58	3,1552	,81223		

Table 3. *(Continue)*

Item 31	Upper	58	3,1552	1,28147	2,914	,004
	Lower	58	2,5690	,84005		
Item 32	Upper	58	4,1207	,70282	6,307	,000
	Lower	58	3,3103	,68073		
Item 33	Upper	58	4,2069	,55439	6,085	,000
	Lower	58	3,4310	,79719		
Item 34	Upper	58	3,9138	,90388	4,132	,051
	Lower	58	3,2931	,70109		
Item 35	Upper	58	4,1552	,69590	6,071	,000
	Lower	58	3,3103	,79927		
Item 36	Upper	58	3,2586	1,25041	2,567	,012
	Lower	58	2,7414	,88971		
Item 37	Upper	58	4,2586	,60872	7,431	,000
	Lower	58	3,3276	,73480		
Item 38	Upper	58	4,3448	,76208	7,494	,000
	Lower	58	3,2931	,74947		
Item 39	Upper	58	4,1724	,67896	7,504	,000
	Lower	58	3,2414	,65722		
Item 40	Upper	58	4,9138	5,39748	2,428	,017
	Lower	58	3,1724	,84059		
Item 41	Upper	58	3,1897	1,48036	1,307	,194
	Lower	58	2,8966	,85203		
Item 42	Upper	58	3,5862	1,02657	2,565	,012
	Lower	58	3,1379	,84704		
Item 43	Upper	58	3,9138	,73232	5,205	,000
	Lower	58	3,1379	,86751		
Item 44	Upper	58	4,0517	,84651	6,171	,000
	Lower	58	3,1724	,67896		
Item 45	Upper	58	3,1552	1,25379	1,706	,091
	Lower	58	2,8276	,75249		
Item 46	Upper	58	4,1552	,83355	6,456	,000
	Lower	58	3,2414	,68340		
Item 47	Upper	58	3,2414	1,53680	2,293	,024
	Lower	58	2,7241	,76761		
Item 48	Upper	58	2,6207	1,46092	-3,028	,003
	Lower	58	3,2586	,66386		
Item 49	Upper	58	3,9828	,73726	4,996	,000
	Lower	58	3,2931	,74947		
Item 50	Upper	58	4,1034	,64044	6,624	,000
	Lower	58	3,2414	,75650		
Item 51	Upper	58	2,6379	1,08738	,308	,759
	Lower	58	2,5862	,67628		
Item 52	Upper	58	4,3103	,65446	7,765	,000
	Lower	58	3,3103	,73046		
Item 53	Upper	58	4,2586	,60872	8,547	,000
	Lower	58	3,1724	,75249		
Item 54	Upper	58	2,8448	1,07282	1,982	,050
	Lower	58	2,5000	,77799		
Item 55	Upper	58	4,2414	,80154	6,599	,000
	Lower	58	3,3276	,68538		

Table 3. *(Continue)*

Item 56	Upper	58	4,2414	,75650	7,002	,000
	Lower	58	3,2931	,70109		
Item 57	Upper	58	4,1897	,78264	6,909	,000
	Lower	58	3,2586	,66386		
Item 58	Upper	58	2,5000	1,20307	-1,386	,168
	Lower	58	2,7586	,75650		

The ones with a high total score also have a high item mean score if the item is distinctive. Those with low total mean score have also low item mean score (Erkuş, 2016: 146).

When Table 3 is examined, it is observed that the t value is not significant in the items numbered 16, 22, 27, 34, 41, 45, 51, 54, 58, and that the 48th item is not good in distinguishing the lower and upper groups. Scale items other than these items can be said to be effective in distinguishing between teachers with low attitudes and teachers with high attitudes towards measurement and evaluation.

2.1.6.3. ***Evaluating the reliability of test items using item total correlations.*** As a result of the item analysis, the reliability of scale items was determined by using item-total correlations. The results of the item analysis are given in Table 4.

Table 4. *Item Analysis Results of the MEASPEST (draft scale)*

Items	N	r	p	Items	r	p
Item 1	214	,382	,000	Item 31	,214	,002
Item 2	214	,564	,000	Item 32	,408	,000
Item 3	214	,503	,000	Item 33	,420	,000
Item 4	214	,424	,000	Item 34	,285	,000
Item 5	214	,397	,000	Item 35	,392	,000
Item 6	214	,414	,000	Item 36	,196	,004
Item 7	214	,392	,000	Item 37	,450	,000
Item 8	214	,297	,000	Item 38	,504	,000
Item 9	214	,395	,000	Item 39	,447	,000
Item 10	214	,405	,000	Item 40	,278	,000
Item 11	214	,352	,000	Item 41	,120	,080
Item 12	214	,388	,000	Item 42	,221	,001
Item 13	214	,413	,000	Item 43	,377	,000
Item 14	214	,244	,000	Item 44	,454	,000
Item 15	214	,369	,000	Item 45	,182	,008
Item 16	214	-,153	,025	Item 46	,472	,000
Item 17	214	,458	,000	Item 47	,200	,003
Item 18	214	,467	,000	Item 48	-,259	,000
Item 19	214	,229	,001	Item 49	,314	,000
Item 20	214	,457	,000	Item 50	,407	,000
Item 21	214	,586	,000	Item 51	-,10	,880
Item 22	214	-,044	,518	Item 52	,472	,000
Item 23	214	,434	,000	Item 53	,568	,000
Item 24	214	,501	,000	Item 54	,083	,229
Item 25	214	,436	,000	Item 55	,453	,000
Item 26	214	,183	,007	Item 56	,491	,000
Item 27	214	,118	,085	Item 57	,495	,000
Item 28	214	,502	,000	Item 58	-,128	,061
Item 29	214	,422	,000			
Item 30	214	,521	,000			

The item-total correlation coefficients are classified as the *very good item* for r ≥ 0.40, the *good item* for 0.30 ≤ r ≤ 0.39, the *item that can be tested if necessary or after correction* for 0.20 ≤ R ≤ 029, and as the *item that should not be tested* for r ≤ 0.19 (Büyüköztürk, 2014: 183).

When Table 4 is examined, it is observed that the correlations of the items in the draft attitude scale that are numbered 8, 14, 16, 19, 22, 26, 27, 31, 34, 36, 40, 41, 42, 45, 47, 48, 51, 54, and 58 are at a low level. It is observed that, for the scale items other than these, the item-total correlations vary between 0.314 - 0.586 and t values are significant. This result can be interpreted that the items on the scale are intended to measure the same behavior. Furthermore, this finding explains that items in the scale have the ability to distinguish the levels of attitudes that physical education and sports teachers have.

2.1.6.4. ***Kaiser-Mayer-Olkin (KMO) and Barlett tests.*** It is an index that compares the magnitude of coefficients of observed correlation with the magnitude of coefficients of part correlation. The KMO ratio should be above 0.5 (Kalaycı, 2014: 322: Pallant, 2003: Tavşancıl, 2002: 50; Altunışık et al, 2005: 217). Therefore, relevant values were studied before Exploratory Factor Analysis (EFA).

The KMO and Barlett results of the 39 items, which remained after the exclusion of 19 items from the draft scale as a result of the analyses, are presented in Table 5.

Table 5. *Kaiser-Mayer-Olkin (KMO) Sample Measurement and Barlett's Test Results*

KMO Sample Measurement Value Sufficiency =		,855
Barlett Test	Proximate Chi-Square Value	2933,471
	sd	741
	p	,000

When Table 5 is examined, it is observed that the results of the Kaiser-Mayer-Olkin (KMO) test and the Barlett test are significant. This shows that there are high correlations between variables, in other words, that the attitude scale is suitable for factor analysis.

3. Findings

3.1. Exploratory Factor Analysis Results

Factor analysis is a multivariate statistic that aims at discovering a few number of new, unrelated, and conceptually meaningful variables (factors, dimensions) by bringing together p number of variables that are associated with each other (Büyüköztürk, 2014: 133). In exploratory factor analysis, it is aimed to define and summarize data by grouping the variables associated with each other (Tabachnick and Fidell, 2015: 614).

As a result of the statistical significance, factor analysis was performed to determine the factor structure of the MEASPEST. The stability of a factor is related to the factor loads carried by its items. In order for a factor to be considered stable, it is emphasized in the literature that item factor loads must be above 0.30 and should consist of at least three items (Hogarty, Hines, Kromrey, Perron and Mumford, 2005; MacCallum, Widaman, Zhang and Hong, 1999; Pituch and Stevens, 2016). Büyüköztürk (2014: 134) also states that if the item factor load value is 0.45 or higher, it indicates a good result, and this limit value can be reduced to 0.30 for smaller number of items. It is also stated that there is a relationship between the factor load value and the sample size. In this case, for a sample of 200 people, the acceptable factor load should be .40, and for a sample of 250 individuals, it should be .35 (Hair, Black, Babin, Anderson and Tatham, 1998: 115).

Common factor variances of items of the scale are presented in Table 6.

Table 6. *Common Variance Values of the Items of the MEASPEST (draft scale)*

	Initial Value	Subtraction Value		Initial Value	Subtraction Value
Item1	1,000	,649	Item25	1,000	,626
Item2	1,000	,730	Item28	1,000	,585
Item3	1,000	,559	Item29	1,000	,487
Item4	1,000	,651	Item30	1,000	,563
Item5	1,000	,587	Item32	1,000	,580
Item6	1,000	,536	Item33	1,000	,579
Item7	1,000	,712	Item35	1,000	,505
Item9	1,000	,517	Item37	1,000	,658
Item10	1,000	,630	Item38	1,000	,432
Item11	1,000	,679	Item39	1,000	,616
Item12	1,000	,646	Item43	1,000	,595
Item13	1,000	,555	Item44	1,000	,443
Item15	1,000	,505	Item46	1,000	,521
Item17	1,000	,582	Item49	1,000	,615
Item18	1,000	,507	Item50	1,000	,664
Item20	1,000	,632	Item52	1,000	,689
Item21	1,000	,678	Item53	1,000	,568
Item23	1,000	,589	Item55	1,000	,665
Item24	1,000	,625	Item56	1,000	,674
			Item57	1,000	,570

According to Table 6, it is observed that the common factor variances of the items in the draft attitude scale vary between .432- .730. This indicates that the factor variance of the items is high.

Table 7. *The Factor Count and Explained Variance Percentage of the MEASPEST (Draft Scale) Based on the Eigenvalue Statistics*

Component	Initial Eigenvalues			Eigenvalues After Reduction		
	Total	Variance Contribution Percentage	Cumulative Percentage of the Variance Contribution	Total	Variance Contribution Percentage	Cumulative Percentage of the Variance Contribution
1	9,542	24,468	24,468	9,542	24,468	24,468
2	2,628	6,739	31,207	2,628	6,739	31,207
3	1,871	4,799	36,005	1,871	4,799	36,005
4	1,671	4,285	40,290	1,671	4,285	40,290
5	1,539	3,946	44,236	1,539	3,946	44,236
6	1,323	3,392	47,627	1,323	3,392	47,627
7	1,223	3,135	50,762	1,223	3,135	50,762
8	1,179	3,022	53,784	1,179	3,022	53,784
9	1,156	2,965	56,749	1,156	2,965	56,749
10	1,072	2,748	59,497	1,072	2,748	59,497
11	,993	2,545	62,041			
12	,928	2,380	64,421			
13	,905	2,321	66,742			
14	,880	2,256	68,998			
15	,844	2,164	71,162			
16	,790	2,026	73,188			
17	,779	1,996	75,185			
18	,715	1,834	77,018			
19	,684	1,753	78,771			

Table 7. *(Continue)*

20	,631	1,617	80,388
21	,605	1,551	81,939
22	,562	1,442	83,381
23	,540	1,384	84,766
24	,523	1,340	86,106
25	,507	1,300	87,405
26	,484	1,242	88,647
27	,462	1,185	89,832
28	,453	1,161	90,993
29	,447	1,147	92,140
30	,429	1,101	93,241
31	,400	1,026	94,266
32	,358	,918	95,185
33	,344	,882	96,067
34	,319	,818	96,885
35	,293	,752	97,637
36	,271	,694	98,331
37	,236	,605	98,935
38	,217	,555	99,490
39	,199	,510	100,000

When Table 7 is examined, it is observed that the 39 items (variables) analyzed are grouped under 10 factors whose eigenvalue is greater than 1, and that these 10 factors explain 59,497% of the variance of the scale. The scree plot and the effect of factors on the total variance are important when deciding the number of factors (Tabachnick & Fidell, 2015: 649, Çokluk et al, 2016: 230). Examining the variance values before deciding the number of the factors, it was determined that the contribution of the first four components to the variance was higher and the contribution decreased after the fourth component. However, when the line graph based on the eigenvalue (*Figure 1*) is examined, a significant decrease is observed after the 4th interval, demonstrating that the scale is suitable for 4-factor Structure.

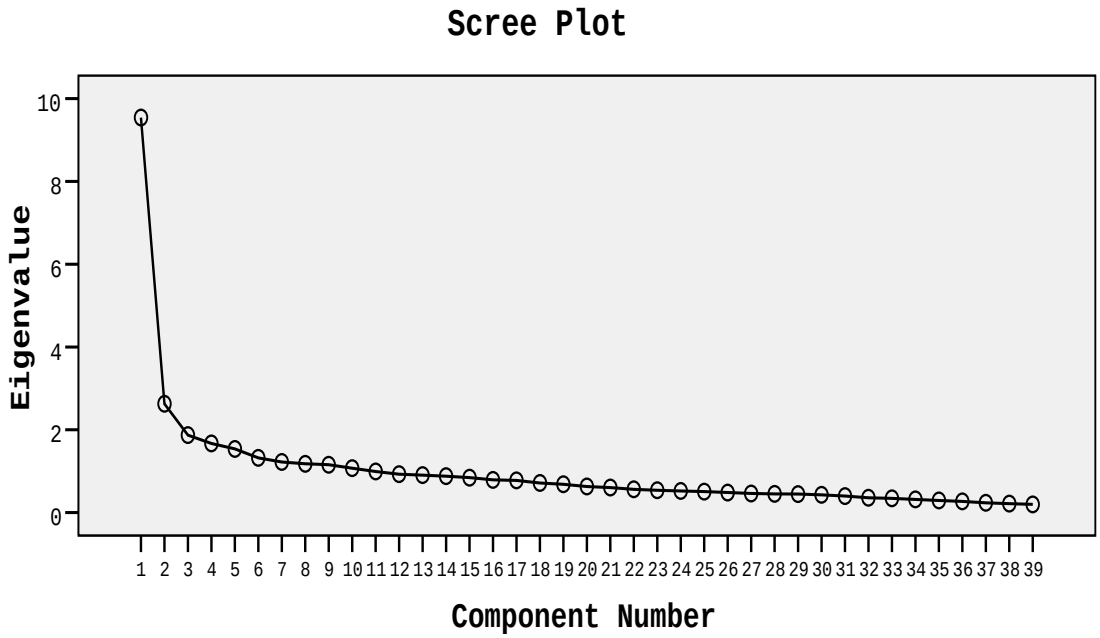

Figure 1. Scree Plot of the MEASPEST (draft scale) Items Delineated based on their Eigenvalues

After deciding about the four-factor structure of the final attitude scale concerning the measurement and evaluation for the physical education and sports teachers, the analysis was conducted based on 0.20 overlapping, the items numbered 49, 17, 43, 28, 50, 18, 12, 5, 5, 25, 9, 13, 24, 30 were excluded respectively,

ultimately obtaining a 24-item final scale form. The explained variance for the resulting 24-item and 4-component final scale form is given in Table 8.

Table 8. *Explained Total Variance of the MEASPEST*

	Initial Eigenvalues			Eigenvalues After Reduction		
Component	Total	Variance Contribution Percentage	Cumulative Percentage of the Variance Contribution	Total	Variance Contribution Percentage	Cumulative Percentage of the Variance Contribution
1	6,020	25,085	25,085	6,020	25,085	25,085
2	2,198	9,159	34,244	2,198	9,159	34,244
3	1,563	6,513	40,757	1,563	6,513	40,757
4	1,389	5,788	46,545	1,389	5,788	46,545
5	1,129	4,704	51,249			
6	1,099	4,580	55,829			
7	1,034	4,309	60,138			
8	,889	3,704	63,842			
9	,878	3,660	67,503			
10	,811	3,381	70,884			
11	,764	3,182	74,065			
12	,708	2,949	77,014			
13	,666	2,775	79,789			
14	,641	2,669	82,459			
15	,574	2,393	84,852			
16	,519	2,162	87,014			
17	,482	2,007	89,021			
18	,470	1,958	90,979			
19	,444	1,849	92,828			
20	,415	1,730	94,558			
21	,386	1,607	96,165			
22	,351	1,464	97,629			
23	,304	1,266	98,895			
24	,265	1,105	100,000			

The explained variance being within the limits of 40% and 60% is considered sufficient in the multifactorial designs in the Social Sciences (Scherer, Wiebe, Luther and Adams 1988: as cited in Tavşancıl, 2014: 48; Henson and Roberts, 2006: 402; Vieira, 2011: 65). When Table 8 is examined, it is observed that the total variance of the final attitude scale for physical education and sport teachers, consisting of 4 factors and 24 items, is 46.545. The factor loads and distributions of the final attitude scale are indicated in Table 9.

Table 9. *Items and Factor Loads of the MEASPEST*

		Component			
	New Item No	1	2	3	4
Item 20	10	,701			
Item 21	11	,699		,325	
Item 44	19	,608			
Item 46	20	,579			
Item 32	14	,574			
Item 37	16	,557			
Item 35	15	,546			
Item 39	18	,545			
Item 38	17	,521			
Item 23	12	,459			
Item 29	13	,438			

Table 9. *(Continue)*

Item 55	22		,729		
Item 56	23		,712		
Item 57	24		,702		
Item 52	21		,617		
Item 2	2			,738	
Item 1	1			,678	
Item 10	7			,572	
Item 15	9			,527	
Item 3	3			,518	
Item 6	5			,507	
Item 7	6				,819
Item 4	4				,776
Item 11	8				,774
Cronbach Alpha	.85	.82	.75	.69	.75

At the end of the validity and reliability studies of the MEASPEST, 34 items were removed from the draft scale and the MEASPEST was put into its final form, consisting of 4 factors and 24 items with a variance value of 46,545. The naming of the factors was based on the classification of the attitudes and values competence field of the General Competences of the Teaching Profession prepared by the Ministry of National Education (MoNE) General Directorate of Teacher Training and Development (2017). Attitudes and Values competence field includes the general attitudes and values of the teaching profession, approaches to the student, national spiritual and universal values, communication and cooperation, and personal and professional development sub-competencies (MEB, 2017: 16). In this context, based on the competence classification of the MoNE, the 4 factors of the 24-item MEASPEST were named as *personal and professional development* (Factor 1), *communication and cooperation* (Factor 2), *student* (Factor 3) and *negativity* (Factor 4).

Under factor 1; the items numbered 10 (,701), 11 (,699), 19 (,608), 20 (,579), 14 (,574), 16 (,557), 15 (,546), 18 (,545), 17 (,521), 12 (,459) and 13 (,438) have the highest factor loads. These items were named as *personal and professional development factor*. All of the items are about following current studies in measurement and evaluation, interest in different measurement and evaluation applications, and the desire to learn new knowledge about the subject. In particular, having certain values for the teachers in the items such as keeping pace with the innovation, research, experimentation and development, made it easier to name this factor as *personal and professional development*. Certain statements in the Attitudes and Values teacher competence field of the MoNE (2017: 16) such as *teachers are doing their profession fondly and willingly*, and *teachers are engaging in activities aimed at improving themselves personally and professionally*, are similar to the naming of the Factor 1.

Under factor 2; the factor loads of the items numbered 22 (,729), 23 (,712), 24 (,702) and 21 (,617) are predominantly clustered. This factor was named as *the communication and cooperation factor*. The items are about the professional conduct and practices of teachers. Certain statements in the Attitudes and Values teacher competence field of the MoNE (2017: 16) such as *being open to sharing knowledge and experience with colleagues* and *using effective communication techniques* are in line with the items grouped under Factor 2.

Under factor 3; the items numbered 2 (,738), 1 (,678), 7 (,572), 9 (,527), 3 (,518) and 5 (, 507) have the highest factor loads. These six items are about the learning process. Especially measurement and evaluation applications and student learning (incentive, fostering interest, teaching to take responsibility, etc.) are the main points highlighted in these items. Therefore, this factor was named as *the student factor*. Certain statements in the Attitudes and Values teacher competence field of the MoNE (2017: 16) such as *displaying attitudes that support developments of students* and *defending that each student can learn* are in parallel with the items grouped under Factor 3.

Similarly, under the Factor 4; the items numbered 6 (,819), 4 (,776), and 8 (,774) have the highest factor loads. All of these variables are about negative thoughts of teachers about measurement and evaluation in physical education and sports course. In particular, statements such as *considering measurement and evaluation*

applications as educationally worthless and unnecessary have been taken into consideration in naming the factor. Therefore, Factor 4 is called the *negativity factor*.

Cronbach Alpha reliability coefficient of the final form of the MEASPEST form was calculated as 0.85. The calculated Cronbach Alpha value is in the range of $0.80 \leq \alpha \leq 1.00$, so the scale is highly reliable (Belt, 2014: 405). The items were formed in a 5-point Likert type design and the responses of the participants were classified as "strongly agree (5)", "agree (4)", "neutral (3)", "disagree (2)" and "strongly disagree (1)". 21 of the items on the scale are positive and 3 were negative. After reversing the negative items (4,6 and 8), the formula *a = Range / number of groups to be decided* was used to determine the group value range of the scale (Taşdemir, 2003). The scoring of the scale is "strongly agree: 4.20-5.00", "agree: 3.40-4.19", "neutral: 2.60-3.39", "disagree: 1.80-2.59", and "strongly disagree: 1.00-1.79".

3.2. Confirmatory Factor Analysis (CFA) Results

CFA is a technique used in the advanced stages of research processes to test a theory about implicit processes. Variables are carefully and meticulously chosen to reveal basic processes (Tabachnick and Fidell, 2015: 614). In order to demonstrate the validity of the proposed model, CFA was performed through the LISREL 9.2 in this study. A single statistical significance test is not sufficient to accurately identify a model obtained with CFA, instead, it is necessary to evaluate it according to many criteria (Schumacker and Lomax, 2004: 83). The value that tests the statistical suitability of the model proposed in CFA and the analysis data is the value x2 (Bacon, 1997: 11). The results of the analysis were examined using the Chi square fit statistics and various fit and error indexes of different perspectives. Second-level confirmatory factor analysis should definitely be performed on multifactor scales (Seçer, 2015). In this context, the second-level CFA was also conducted and the relevant indices, the fit indices selected for comparison, and the required limit are indicated in Table 2.12.

Table 10. *CFA Model Fit Index Values and CFA Results of the MEASPEST*

Fit Index	Acceptable Fit	Perfect Fit	1st Level CFA Results	2nd Level CFA Results
NFI	.90 and above	.95 and above	0,91	0.90
NNFI	.90 and above	.95 and above	0,98	0.97
IFI	.90 and above	.95 and above	0,98	0.98
RFI	.90 and above	.95 and above	0,90	0.89
CFI	.90 and above	.95 and above	0,98	0.98
GFI	.85 and above	.90 and above	0,90	0.89
AGFI	.85 and above	.90 and above	0,87	0.87
RMR	Between .050 - .080	Between .000 - .050	0.046	0.056
REMSEA	Between .050 - .080	Between .000 - .050	0.031	0.034
x^2/ sd	Between 2-3 (inc.)	Between 0 - 2 (inc.)	1.20	1.25

Source: (Bayram, 2010:78; Byrne, 2010:80; Hair, Black, Babin & Anderson, 2010:668; Schumacker &Lomax, 2004: 83; Şimşek, 2007:48).

When Table 10 is examined, it was concluded that, based on the first and second level CFA results of the MEASPEST, the models of the scale items in the relevant structure were suitable. When the model-data fit is examined, it is observed that the acceptability of the data is quite good and the model is confirmed.

4. Conclusion and Discussion

In this study, a scale was developed to determine the attitudes of physical education and sports teachers concerning measurement and evaluation. As a result of the literature review and interviews with the teachers, certain statements were converted into a format that could be used in scale form and the draft form was designed in the 5-point Likert type, creating an item pool with 73 positive and negative items.

While the draft scale included 73 items, it was reduced to 58 items by determining the coefficients of concordance among the raters, and the coefficient of concordance was determined for the 58-item draft form as 93% for the agreement among the raters. The draft scale was applied to 50 physical education and sports teachers outside the sample group concerning the validity for usefulness. As a result of the pilot

implementation, the draft form of the scale was finalized as a total of 58 items, 17 of which were negative. The draft measurement form was applied to 222 physical education and sports teachers on a voluntary basis.

Distinctiveness and reliability of scale materials were examined. As a result of the analyses, items numbered 8, 14, 16, 19, 22, 26, 27, 31, 34, 36, 40, 41, 42, 45, 47, 48, 51, 54 and 58 were removed from the scale.

It was observed that 39 items (variables) were grouped under 10 factors whose eigenvalue is greater than 1, and that these 10 factors on the scale explained 59,497% of the variance. Examining the variance values before deciding the number of the factors, it was determined that the contribution of the first four components to the variance was higher and the contribution decreased after the fourth component. However, the line graph based on the eigenvalue was examined, and a significant decrease is observed after the 4th interval, demonstrating that the scale is suitable for a 4-factor structure. After deciding about the four-factor structure of the final attitude scale concerning the measurement and evaluation for the physical education and sports teachers, the analysis was conducted based on 0.20 overlapping, the items numbered 49, 17, 43, 28, 50, 18, 12, 5, 5, 25, 9, 13, 24, 30 were excluded respectively, ultimately obtaining a 4-component and 24-item final scale form.

At the end of the validity and reliability studies of the attitude scale for physical education and sports teachers concerning measurement and evaluation, 34 items were removed from the draft scale and it was put into its final form, consisting of 4 factors and 24 items with a variance value of 46,545.

The naming of the factors was based on the classification of the attitudes and values competence field of the General Competences of the Teaching Profession prepared by the Ministry of National Education (MoNE) General Directorate of Teacher Training and Development (2017). Attitudes and Values competence field includes the general attitudes and values of the teaching profession, approaches to the student, national spiritual and universal values, communication and cooperation, and personal and professional development sub-competencies (MEB, 2017: 16). In this context, based on the competence classification of the MoNE, the 4 factors of the 24-item scale were named as *personal and professional development* (Factor 1), *communication and cooperation* (Factor 2), *student* (Factor 3) and *negativity* (Factor 4).

Cronbach Alpha reliability coefficient of the final form of the attitude scale for physical education and sports teachers concerning measurement and evaluation was calculated as 0.85. The items were formed in a 5-point Likert type design and the responses of the participants were classified as “strongly agree (5)”, “agree (4)”, “neutral (3)”, “disagree (2)” and “strongly disagree (1)”. 21 of the items on the scale are positive and 3 were negative.

Whether the results of the exploratory factor analysis were fit was examined through the chi-square fit statistics and the RMR, AGFI, GFI, RFI, NNFI, IFI, CFI, GFI, and RAMSEA fit indices. The first and second level CFA results of the attitude scale for physical education and sports teachers concerning measurement and evaluation demonstrated that the models of the scale items for this structure were suitable.

When the literature was examined, it was observed that certain subjects were studied such as attitude scales for the course of measurement and evaluation in education (Yaşar 2014; Aktaş & Alıcı, 2012: Khan, 2006; Akdağ Gürsoy, 2015; Ozan & Köse, 2013), scales developed to measure the attitudes of prospective teachers of other fields regarding the measurement and evaluation (Yıldırım & Öztürk, 2009; İzci, Göktaş and Şad 2014), attitude scales developed for teachers of other fields (Karaduz, 2009; Erdoğdu, 2010; Çalışkan, 2012; Çalışkan & Yazıcı, 2013; Kılıç, 2014), scales about the teachers’ competences on measurement and evaluation (Gelbal & Kellecioglu), views and practices of teachers on alternative evaluation (Güneş, Dilek, Hoplan, Celikoglu, & Demir, 2010; Çalışkan & Sağlam, 2017; Aşık, 2019; Aras, 2020), studies on measurement and evaluations competences of physical education teachers (Tousignant and Siedentop, 1983; Weinberg, 1996; Lai, Wu, Lee, and Jhang, 2018), the frequency of use of measurement and evaluation tools by the physical education teachers (Killoran, 1982), and what the physical education teachers prioritize in selecting the measurement and evaluation tools (Hensley, 1997), which tools the physical education teachers use in their courses (Lund, 1997; Mintah, 2003). However, no measurement tool was encountered that measures the attitudes of physical education and sports teachers. In this regard, it is thought that this study will contribute to the literature and it will fill this gap.

Recommendations

By using the scale developed within the scope of this study, further studies can examine measurement and evaluation attitudes of physical education and sports teachers in different samples. The results of the studies can be evaluated by the Ministry of National Education (Turkey), and projects can be developed for the development of teachers in this field.

References

Akdağ Gürsoy, G. (2015). *Alan ile ilişkilendirilmiş uygulamalı ölçme ve değerlendirme dersinin öğretmen adaylarının ölçme ve değerlendirme okuryazarlık düzeylerine, tutumlarına ve alan bilgilerine etkisi* (Yayınlanmamış Doktora Tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Aktaş, M., & Alıcı, D. (2012). Eğitimde ölçme ve değerlendirme dersine yönelik tutum ölçeği'nin (EÖD-TÖ) geliştirilmesi. *Journal of Qafqaz University, Philology and Pedagogy*, *33*, 66-73.

Altunışık, R., Coşkun, R., Bayraktaroğlu, S., & Yıldırım, E. (2005). *Sosyal bilimlerde araştırma yöntemleri*. Adapazarı: Sakarya Kitabevi.

Aras, E. (2020). *Ortaokul matematik öğretmenlerinin alternatif ölçme değerlendirme araç ve yöntemlerine ilişkin yaklaşımları, yeterlik algıları ve kullanım durumlarının incelenmesi* (Yayımlanmamış yüksek lisans tezi). Uşak Üniversitesi Fen Bilimleri Enstitüsü, Uşak.

Aşık, İ. (2019). *Web tabanlı mesleki gelişim uygulamasının matematik öğretmenlerinin ölçme değerlendirmeye yönelik yaklaşımlarına etkisinin incelenmesi* (Yayımlanmamış doktora tezi). Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul.

Bacon, L. D. (1997). *Using Amos for structural equation modeling in market research*. New York: Lynd Bacon & Associates, Ltd. and SPSS Inc.

Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.

Bayram, N. (2010). *Yapısal eşitlik modellemesine giriş AMOS uygulamaları*. Bursa: Ezgi Kitabevi.

Blest, D. C. (2003). A new measure of kurtosis adjusted for skewness. *Australian & New Zealand Journal of Statistics*, *45*(2), 175-179. doi: 10.1111/1467-842X.00273.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.

Byrne, B. M. (2010). *Multivariate applications series. Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd ed.).* New York, NY, US: Routledge/Taylor & Francis Group.

Çalışkan, H. (2012). Development of the measurement and evaluation self-efficacy perception scale and the examination of the status of social studies teachers. *Energy Education Science and Technology Part B: Social and Educational Studies, 4*(1) Special Issue, 1003-1008.

Çalışkan, H., & Yazıcı, K. (2013). Developing an attitude scale assessment and evaluation and analyzing of social studies teachers attitude levels in terms of several variables. *International Journal of Human Sciences*, *10*(1), 398-415.

Çalışkan, H. & Sağlam, H. I. (2017). Turkish Social Studies Teachers' Opinions about Performance Tasks in Elementary Education. *International Journal of Psychology and Educational Studies*, *4*(1), 1-12.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi.

Devellis, R. F. (2017). *Scale development-theory and applications*. (Çev. T. Totan). Ankara: Nobel Yayıncılık (Orijinal baskı tarihi, 2012)

Erdoğdu, M. Y. (2010). Öğretmenlerin ölçme ve değerlendirmeye yönelik tutumlarının bazı değişkenler açısından incelenmesi. In *International Conference on New Trends in Education and Their Implications* (pp. 11-13).

Erkuş, A. (2016). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi.

Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*(33), 135-145.

George, D., & Mallery, M. (2010). *SPSS for windows step bysstep: A simple guide and reference*. 17,0 update (10a ed.) Boston: Pearson.

Güneş, T., Dilek, N. Ş., Hoplan, M., Çelikoğlu, M., & Demir, E. S. (2010, November). Öğretmenlerin alternatif değerlendirme konusundaki görüşleri ve yaptıkları uygulamalar. *International Conference on New Trends in Education and Their Implications* (pp. 11-13).

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall. doi: 10.1016/j.jmva.2009.12.014.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis (Vol. 5)*: Prentice hall Upper Saddle River, NJ.

Hensley, L. D. (1997). Alternative assessment for physical education. *Journal of Physical Education, Recreation & Dance, 68*(7), 19-24.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological measurement*, *66*(3), 393-416. doi: 10.1177/0013164405282485.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement 65*(2), 202-226. doi: 10.1177/0013164404267287.

Howitt, D., & Cramer, D. (2011). *Introduction to SPSS statistics in psychology: For version 19 and earlier (Fifth edition).* London: Pearson Education Limited.

İnceoğlu, M. (2011). *Tutum algı iletişim*. Ankara: Siyasal Kitabevi.

İzci, E., Göktaş. E., & Şad, S. N. (2014). Öğretmen adaylarının alternatif ölçme değerlendirmeye ilişkin görüşleri ve yeterlilik algıları. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi (KEFAD), 15*(2), 37-57.

Kalaycı, Ş. (2005). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. Ankara: Asil Yayın Dağıtım.

Kalaycı, Ş. (2014). Faktör analizi. Ş. Kalaycı (Ed.). *Spss Uygulamalı çok değişkenli istatistik teknikleri* içinde (s. 321-331). Ankara: Asil Yayınları.

Karaca, E. (2003). *Öğretmen adaylarının ölçme ve değerlendirme yeterlik ilişkin algılar* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Karadüz, A. (2009). Türkçe öğretmenlerinin ölçme ve değerlendirme uygulamalarının "yapılandırmacı öğrenme" kavramı bağlamında eleştirisi. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *22*(1), 189-210.

Kayış, A. (2014). Güvenilirlik Analizi. Ş. Kalaycı (Ed.). Sp*ss Uygulamalı Çok Değişkenli İstatistik Teknikleri* içinde (s. 404-409). Ankara: Asil Yayınları.

Keller, J. (2017). *Attitude is everything*. (Çev. O. Çelik). İstanbul: Beyaz Yayınları (orijinal baskı tarihi 2012).

Kılıç, R. (2014). *İlköğretim I. kademe öğretmenlerinin otantik ölçme ve değerlendirme yöntemleri ile ilgili bilgi, tutum ve görüşlerinin değerlendirilmesi.* (Yayımlanmamış Doktora Tezi). Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul.

Killoran, J. (1992). In defense of the multiple-choice question. *Social Education*, *56*(2), 106-108.

Lai, H. R., Wu, D. M., Lee, P. H., & Jhang, Y. S. (2018). Health literacy teaching beliefs, attitudes, efficacy, and ıntentions of middle school health and physical education teachers. *Journal of School Health, 88*(5), 350-358.

Lund, J. (1997). Authentic assessment: Its development and applications. *Journal of Physical Education. Recreation, and Dance. 68*(7), 25-28.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods 1*(2), 130. doi: 10.1037/1082-989X.1.2.130.

Miles, M., B.,& Huberman, A., M. (1994). *Qualitative data analysis*. Thousand Oaks,Ca: Sage.

Milli Eğitim Bakanlığı. (2017). Öğretmenlik mesleği genel yeterlikleri 10.04.2018 tarihinde http://oygm.meb.gov.tr/meb_iys_dosyalar/2017_12/11115355_YYRETMENLYK_MESLEYY_GENEL_YETERLYKLERY.pdf adresinden indirilmiştir.

Mintah, J. K. (2003). Authentic assessment in physical education: Prevalence of use and perceived ımpact on students' self-concept, motivation, and skill achievement . *Measurement in Physical Education and Exercise Science*, *7*(3), 161–174.

Ozan, C., & Köse, E. (2013). Eğitimde ölçme ve değerlendirmeye yönelik tutum envanterinin Türkçe'ye uyarlanması. *e-Uluslararası Eğitim Araştırmaları Dergisi*, *4*(2), 29-47.

Özgüven, İ. E. (2015). *Psikolojik testler*. Ankara: Nobel Yayın Dağıtım.

Pallant, J. (2003). *SPSS survival manual*. *Versions 10 and 11*. Buckingham: Open University Press

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM"s SPSS (6th ed.)*. New York: Routledge. doi: 10.4324/9781315814919.

Quilter, S. M. (1998). *Inservice teachers' assessment literacy and attitudes toward assessment* (Unpublished doctoral dissertation). University of South Carolina.

Schumacker, R. E., & Lomax, R. G. (2004). *Beginner's guide to structural equation modeling (2nd ed.).* New York: Psychology Press, Taylor & Francis Group. doi: 10.4324/9781410610904.

Seçer, İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci. SPSS ve LISREL uygulamaları*. Ankara: Anı Yayıncılık.

Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş:(Temel ilkeler ve LISREL uygulamaları).* İstanbul: Ekinoks.

Şişko, M., & Demirhan, G. (2002). İlköğretim okulları ve liselerde öğrenim gören kız ve erkek öğrencilerin beden eğitimi ve spor dersine ilişkin tutumları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 23,(23), 205-210.

Tabachnick, B., G., & Fidell, L., S. (2015). *Using multivariate statistics*. (Çev. M. Baloğlu). Ankara: Nobel (Orijinal baskı tarihi, 2013)

Taşdemir, M. (2003). *Eğitimde planlama ve değerlendirme*. Ankara: Ocak Yayınevi.

Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayıncılık.

Tavşancıl, E. (2014). *Tutumların ölçülmesi ve SPSS İle veri analizi.* Ankara: Nobel Yayıncılık

Tekindal, S. (2009). *Okullarda ölçme ve değerlendirme yöntemleri*. Ankara: Nobel

Tousignant, M., & Siedentop, D. (1983*).* A qualitative analysis of task structures in required secondary physical education classes. *Journal of Teaching in Physical Education, 3*(11), 47-57.

Turgut, M. F. (1995). *Eğitimde ölçme ve değerlendirme metotları*. Ankara:Yargıcı Matbaası.

Vieira, A. L. (2011). *Interactive LISREL in practice (Vol. 510).* New York, NY: Springer. doi: 10.1007/978-3-642-18044-6.

Weinberg, H. (1996). Authentic assessment in middle school physical education. *Teaching Secondary Physical Education*, *75*(1), 9-12.

Yaşar, M. (2014). Eğitimde ölçme ve değerlendirme dersine yönelik tutum ölçeğinin geliştirilmesi. *Eğitim Bilimleri Araştırmaları Dergisi*, *4*(1), 259-279.

Yıldırım, F., & Öztürk, K. B. (2009). Türkçe dersi öğretim programının ölçme değerlendirme ögesi hakkında öğretmen görüşleri. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 3*(37), 92-108.

Annex-A: Items of Measurement and Evaluation Attitude Scale for Physical Education and Sports Teachers

	Ek-A: Beden Eğitimi ve Spor Öğretmenleri İçin Ölçme Değerlendirmeye İlişkin Tutum Ölçeği Maddeleri
	Kesinlikle Katılmıyorum (1) **Katılmıyorum (2)** **Kararsızım (3)** **Katılıyorum (4)** **Kesinlikle Katılıyorum (5)**
1	Ölçme değerlendirme uygulamalarının (proje,gözlem vb.) eğitim sürecinde etkili bir öğretim aracı olduğunu düşünüyorum.
2	Farklı ölçme değerlendirme uygulamalarının, öğrencilerde öğrenmeyi daha istekli hâle getirdiğine inanıyorum.
3	Derslerimde farklı ölçme değerlendirme araçlarını kullandığımda öğrencilerin derse ilgisinin arttığını düşünüyorum.
4	Ölçme değerlendirmede kullanılan araçların bilinmesinin eğitsel değeri olmadığını düşünüyorum.
5	Ölçme değerlendirme uygulamalarının, öğrencilerin sorumluluk duygusunu geliştirdiğine inanıyorum.
6	Beden eğitimi ve spor dersinde alternatif ölçme değerlendirme araçlarını gereksiz buluyorum.
7	Zengin ölçme değerlendirme sürecinin, bireysel farklılıklara uygun öğrenme fırsatı sağladığına inanıyorum.
8	Beden eğitimi ve spor dersinde, süreci değerlendirmenin gereksiz olduğunu düşünüyorum.
9	Derslerimde kullandığım farklı ölçme değerlendirme uygulamalarının, öğrencilere kendi kendine ders çalışma alışkanlığını kazandırdığını düşünüyorum.
10	Ölçme değerlendirme ile ilgili güncel çalışmaları takip ederim.
11	Farklı ölçme değerlendirme uygulamalarını kullandıkça işimi daha çok seviyorum.
12	Öğretmenlerin, ölçme değerlendirme konularına hakim olması gerektiğini düşünüyorum.
13	Süreç odaklı ölçme değerlendirme uygulamalarının, öğrencilerin dikkat düzeylerini artırdığını düşünüyorum
14	Farklı ölçme değerlendirme uygulamalarını öğrenmek çok ilgimi çeker.
15	Diğer öğretmenlerin derslerinde kullandıkları farklı ölçme değerlendirme uygulamalarının örnek alınması gerektiğine inanıyorum.
16	Ölçme değerlendirme sürecinin, öğrencilerin güçlü yönlerinin tespitinde önemli rol oynadığına inanıyorum.
17	Farklı ölçme değerlendirme araçlarını derslerimde uyguladıkça, çalışma verimimin arttığını düşünüyorum.
18	Ölçme değerlendirme sürecinin kalıcı öğrenmeyi sağladığını düşünüyorum.
19	Ölçme değerlendirmeyle ilgili yeni bilgiler öğrenmenin mesleki olarak gelişmeye imkân sağladığına inanıyorum.
20	Öğretmenlerin farklı ölçme değerlendirme araçlarını kullanmaları konusunda teşvik edilmeleri gerektiğine inanıyorum.
21	Öğretmen olduktan sonra eğitimde ölçme değerlendirmenin önemini daha iyi anladım.
22	Ölçme değerlendirme süreçlerinde, meslektaşlarımla işbirliği yapmanın önemli olduğuna inanıyorum.
23	Zengin ölçme değerlendirme süreçlerinin, öğrencilerin eleştirel düşünmelerine imkân sağladığını düşünüyorum.
24	Her geçen gün ölçme değerlendirmenin gerekliliğine daha çok inanıyorum.
Kişisel ve mesleki gelişim faktörü: 10., 11., 12., 13., 14., 15., 16., 17., 18., 19. ve 20. maddeler. *İletişim ve işbirliği faktörü*: 21.,22.,23. ve 24. maddeler. *Öğrenci faktörü: 1.*, 2., 3., 5., 7. ve 9. maddeler. *Olumsuzluk faktörü: 4.,6. ve 8. maddeler. (Ters puanlanacak maddeler)*	