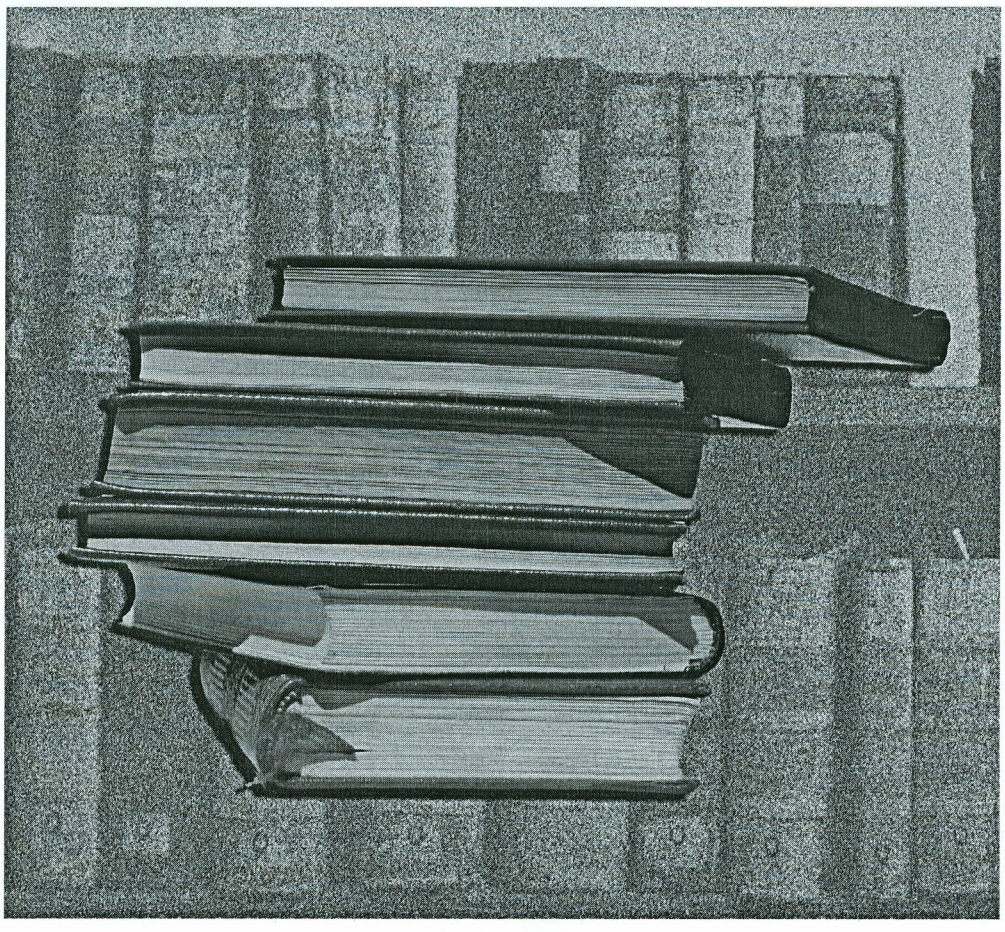
ISSN 1308-7711

ENERGY EDUCATION SCIENCE AND TECHNOLOGY, PART B
Social and Educational Studies



Volume 5 . Issue 1
January 2013

SILA SCIENCE

ENERGY EDUCATION SCIENCE & TECHNOLOGY, PART B

Social and Educational Studies

An International Journal

Publishing Office: SILA SCIENCE, Universite Mah., Mekan Sok. No: 24, Trabzon, Turkey.

Frequency: Energy Education Science & Technology Part B (ISSN 1308-7711) is published 4 times per year.

Abstracted and / or Indexed In: Chemical Abstracts

Aims and scope: Energy Education Science & Technology Part B is dedicated to detailed and comprehensive investigations and appropriate reviews of the interdisciplinary aspects of energy education, chemical education, physical education, mathematical education, biological education, science and education, environmental education, health science and education, all social educational topics, social and behavioral sciences, and political and economic studies.

Publication information: Please address all your requests regarding orders and subscription queries to: A. Demirbas, P. K. 216, 61035-
Trabzon, Turkey. E-mail: ayhandemirbas@hotmail.com

ENERGY EDUCATION SCIENCE & TECHNOLOGY, PART B
Social and Educational Studies

ISSN 1308-7711

Energy Education Science and Technology Part B: Social and Educational Studies
2012 Volume (issue) **Special Issue:** 1003-1008

Development of the measurement and evaluation self-efficacy perception scale and the examination of the status of social studies teachers[†]

Huseyin Caliskan[*]

Sakarya University, Educational Faculty, Sakarya, Turkey

Received: 21 June 2012; accepted: 30 September 2012

Abstract

This study has been carried out for two main reasons: One of them is to develop a scale to determine the self-efficacy levels of teachers regarding measurement and evaluation. Whereas the second reason is to determine the self-efficacy levels of social studies teachers and their status based on various variables regarding measurement and evaluation. The study group consists of a total of 395 social studies teachers. A measurement tool composed of 18 items and 4 factors that explains 64.95% of the total variance has been obtained as a result of the exploratory factor analysis carried out. The factor loads of the scale vary between .45 and .82. It has been determined that both the internal consistency coefficients and the test-re-test reliability coefficients calculated for the whole scale are .93 and that the two half reliability coefficient is .86. As a result of item analysis, it has been determined that the item-total score correlations of the sub-scales vary between .43 and .74 and that all the differences between the averages of the 27% sub-super groups are significant. Based on these results, it can be stated that the scale developed is a valid and reliable tool of measurement. In addition, it has been determined as a result of measurements carried out using the developed scale that the self-efficacy levels of social studies teachers regarding measurement and evaluation is sufficiently high and that there is no difference in terms of the variables of gender, professional seniority and the work residence.

Keywords: Social studies teachers; Measurement -evaluation; Self-efficacy; Scale development.

1. Introduction

Education is one of the key elements that encompasses the whole life span of individuals and has an effect on the determination of the life standards of that individual. Education defined as "the process of ensuring desired behavior or performing desired behavior modifications" [1] or as "The process of shaping and modifying the behavior of individuals" [2] has significant effects on the social, political and economical elements of a country. Accordingly, taking into account the education system and the education policy of a country when trying to determine the development level of a country would enable correct comments to be made.

There are many factors that affect the quality and efficiency of the process of education. According to Kahyaoglu and Yangin [3], the quality and efficiency of education depends on the proper management of all parts of the education system. Yalin [4], defining the process of learning-teaching as a system has also mentioned the necessity of the effective and harmonious arrangement of all elements within the process in order to ensure that education reaches its goals. The quality of education will undoubtedly increase in an environment where factors such as students, teachers, principals, inspectors, family, education materials all interact with each other.

Changes have occurred in the education understanding of this century in the light of advancements in science and technology. Parallel to these changes, learning-teaching process along with the content of measurement and evaluation concepts have also changed. Measuring and evaluation have transformed into advancing activities that give feedbacks to learning and education processes. The objectives of these activities carried out with student centered education understanding are to train creative individuals who can use, apply, comment on and relate information to ease life.

The advancements that have drawn the most attention in the education process have been in the field of measurement and evaluation. Measurement and evaluation applications have tried to determine the extent with which students have acquired the basic information and skills within the curriculum whereas enabling the students and their peers to evaluate their own studies. By this means, students can take active part in the evaluation process and perceive themselves and their surrounding with a subjective view. This in turn creates an environment in which they can analyze their strengths and weaknesses and be successful in life [5].

Measurement is the visual display of observations by means of numbers and other symbols [6]. In other words, it is the expression by the measurer of a specific dimension of a specific entity using a proper measurement tool [7]. Whereas for [8] who defines measurement as a means of depiction, it is the observation of whether a specific object or set of objects have a certain characteristic and its level if there is and the expression of observation results via symbols and especially with numbers. For education, measurement is the thorough analysis of expected behavior changes and the determination of observed and unobserved aspects, the research of the levels individuals display in their behavior changes and lastly the determination of whether the expected behaviors are within the desired limits and in accordance to the desired conditions [9].

Whereas evaluation is to reach certain conclusions based on the comparison of measurement or observation data [10].Evaluation is an interpretation of measurement results. This interpretation is a display of the proficiency of the teacher in addition to being a classification of students as successful and unsuccessful [11].The continuous follow up of the education process via measurements and evaluations enables one to determine problems and make rearrangements [12].Teachers have a great responsibility in the regular measurement and evaluation of student success to ensure a flawless education process. The realization of effective learning by setting to work is possible through the knowledge, skill and attitudes of teachers who prepare students for the 21st century. These knowledge, skill and attitudes comprise the competency of teachers [13]. Kahyaoglu and Yangin [3] emphasize that the bringing up of teachers to the desired levels are related with teacher competency. In order for teachers to use measurement and evaluation methods by the book and in an unbiased and reliable manner, it is required that they be skillful in measurement-evaluation and use this knowledge effectively. It is very important for increasing the quality of training and education that teachers are able to use these measurement and evaluation methods effectively and efficiently. Because it is possible to reschedule the process by way of learning the changes in the knowledge, emotions and thoughts of the students only through the application of correct measurement-evaluation. The teacher should have knowledge on the various special information of the field of measurement and evaluation sufficient enough to be able to use them comfortably and should develop some skills in this field along with positive attitudes. To this end, it is very important to put forth the aptitudes of teachers regarding measurement and evaluation. This study has been carried out for two purposes: One is to develop a scale in order to determine the self-efficacy perception levels of the teachers regarding measurement and evaluation. Whereas the second purpose is to determine the self-efficacy perception levels of social studies teachers regarding measurement and evaluation and also to determine whether there are any

[*]Corresponding author. Tel.: +90-264-614-1033; fax: +90-264-295-7133.
E-mail address: caliskan06@gmail.com (H. Caliskan).
[†]This study was presented in 12th International Educational Technology Conference on July 11-13, 2012.

changes in these levels based on the variables of gender, seniority and the unit of duty.

2. Method

2. 1. Research model

In this study designed as a general scanning model, data was acquired using cross-sectional data acquisition from general scanning models. Cross-sectional scanning model is the approach in which only one measurement is made during the study regarding the properties of the variables to be defined [14].

2. 2. Study group

The study group consists of a total of 395 social studies teachers 111 of whom are female and 284 of whom are males who have participated to the in-service training seminars arranged by the Ministry of Education. Of the participants, 165 work at city schools, 128 at district schools, 102 at town and village elementary schools and the age range is 22-6. This number was accepted to be sufficient due to the fact that the number of participants was over 300. According to Tabachnick and Fidell [15] a total of 300 people are "good" for factor analysis, a total of 500 people "is very good" and a total of 1000 people is "perfect". The study group was subject to the trial "Measurement and Evaluation Self-Efficacy Perception Scale" and the analysis of the scale was made accordingly. Additionally, a different group of 35 teachers was also used in order to ensure test-re-test reliability.

2. 3. Data acquisition tool

First, a relevant literature survey was carried out in order to determine the items making up the scale and theoretical information was examined. In addition, interviews with teachers were conducted and 8 teachers were asked to write an essay regarding how they go about the measurement and evaluation processes. The clues obtained from both the interviews and the essays were combined systematically with information acquired from relevant literature and an item pool of 30 items was prepared by the researcher.

In order for the validity of this form prepared using these written statements, specialists were determined to evaluate the comprehensibility, scope and face validity. The form was presented to 4 psychological consultants along with 3 scholars working on measurement and evaluation and 3 scholars working on Turkish language for the evaluation of comprehensibility, scope and face validity and their opinions were asked. Required corrections and exclusions were made in the scale in accordance with the opinions and criticisms and the trial scale of 25 items was prepared after which the reliability and validity studies were conducted. A five point Likert type scale was used to express the relevant acceptance level for the scale which had no negative items. This reting system was composed as *"I agree completely (5), I agree (4), I partially agree (3), I agree very little (2), I don't agree at all (1)"*. In addition, a guideline was included in the beginning of the scale in order to give information regarding the purpose of the scale, the number of items and the method of answering.

2. 4. Data acquisition and analysis

The trial "Measurement and Evaluation Self-Efficacy Perception Scale" was applied to 395 social studies teachers after which exploratory and corrective factor analyses were carried out using the obtained data as a basis for reliability and validity studies. Whereas exploratory factor analysis aims to explore the factor structure based on the relationships between variables, the corrective factor analysis that examines the accordance between the model and the data tests the hypotheses regarding the relationship between the variables [16]. "Principal component analysis rotated to varimax rotation" was used for the accordance of the scale to construct validity.

The accordance with principal component analysis was determined using Kaiser-Meyer Olkin (KMO) coefficient used to determine whether the sample size is suited to the selected analysis or not and the Barlett Test of Sphericity which gives information regarding whether the data comes from a multi-variable normal distribution or not. The suggestions of Fabrigar, Wegener, MacCallum and Strahan [17] were taken into account when deciding on the analysis method and rotation technique. Various fit indexes such as Chi-Square Fit Test, $\chi 2$/sd, Goodness of Fit Index (GFI), Comparative Fit Index (CFI), Root Mean Square Residuals (RMR or RMS) and Root Mean Square Error of Approximation (RMSEA) were used in order to evaluate the fit of the model in confirmatory factor analysis. In relevant literature it is expected for model data fit that the values of GFI, CFI and AGFI are greater than .90 or that the RMS or the standardized RMS and RMSEA values are smaller than .05, however a value smaller than .08 for RMSEA is also acceptable [15, 18, 19]. The value of $\chi 2$/sd is expected to be between 0 and 2, however a value smaller than 5 is also taken as an acceptable value [20].

In addition, internal consistency (Alpha), Spearman-Brown split-half test and test-re-test reliability coefficients were calculated for the whole scale and the dimensions the factor structure of which have been determined. Whereas for item analysis, the significance of the differences between the corrected item-total score correlation and the item averages of the upper 27% and the lower 27% groups were examined using *t* test. In addition, the arithmetic average and standard deviation values were examined to determine the self-efficacy perceptions of social studies teachers regarding measurement and evaluation and One-Way Anova analysis was carried out in order to put forth whether the self-efficacy perceptions of teachers differed according to different variables or not. SPSS 15 and LISREL 8.7 package software were used for the analyses of acquired data.

3. Results

3. 1. The verification status of the validity and reliability of the scale

Kaiser-Meyer Olkin (KMO) and Barlett Test of Sphericity were carried out in order to determine whether the scale is suitable to factor analysis or not. KMO is a statistical method used to determine whether the data and sample size are suited to and sufficient for the selected analysis or not. A KMO coefficient that is close to 1 means that the data is suitable for analysis. As a result of the analysis that has been carried out, a KMO value of .94 has been found. The selected feature should show a normal distribution in space in order for the parametric method to be used. *Barlett Sphericity* test is a statistical method that can be used to check whether the data come from a multi-variable normal distribution or not. A significant chi-square ($\chi 2$) test statistic obtained as a result of this test shows that the data come from a multi-varible normal distribution. As a result of the analysis carried out during the study, the Barlett sphericity test was determined to be significant ($\chi 2 = 5300.29$; $p < 0.01$).

The results of KMO test measurements should be equal to or greater than .60 and the Barlett sphericity test result should be statistically significant [21]. It was conclude that factor analysis can be carried out since the values obtained as a result of the analyses carried out were in good accordance with the basic assumptions.

Principal component analysis rotated to varimax rotation was used to test the construct validity of the self-efficacy perception scale. Results for the factors have been given in Table 1.

When Table 1 is examined, it is seen that there are four factors with eigen values of greater than 1.00. The contribution of these factors to the total variance is 58.58%. More correct and realistic decisions can be given by looking at the scree plot along with the eigen values. In factor analysis, factors with eigen values equal to or greater than 1 are accepted as dominant factors [22]. Scree plot helps to decrease factors by putting forth the dominant factors [19]. The scree plot can be seen in Fig. 1.

Table 1. Data regarding factors obtained as a result of factor analysis

Factor	Value	Variance Percentage	Cumulative Variance Percentage
1	10.53	42.14	42.14
2	1.74	6.96	49.10
3	1.29	5.17	54.27
4	1.08	4.31	58.58

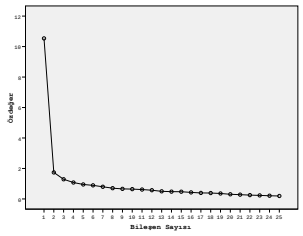


Fig. 1. Scree plot

Since according to the plot in Fig. 1, the slope of rapid decreases does not start to be fixed after the fourth factor, the number of factors can be accepted as four. In addition, the existence of factors with eigenvalues greater than 1 also supports this finding.

In construct validity testing, the sample size has been taken into account [23] and the principle that the load of each variable should be greater than .32 [15] has been accepted as the basic principle for the factor loads of the scale consisting of 25 items. When the four factor structure is evaluated in terms of whether the items meet the overlapping and factor load acceptability levels, it is observed that six items are overlapping (items 6, 7, 8, 10, 11 and 19) and that one item (item 9) is below the acceptable factor load value. As a result of the analysis excluding a total of seven items, it was observed that the contribution to total variance of factor one is 44.44%, factor two is 9.76%, factor three is 6.16%, factor four is 5.59%. Whereas the contribution to the total variance of the determined four factors is 64.95%. Whereas it can be accepted that the total variance for single factor patterns is at least 30% [22], this ratio is expected to be over 41% for multi factor patterns [24]. In this regard, it can be stated that the total variance percentage explained by four factors is sufficient.

The factor pattern of the 18 item scale obtained as a result of the analyses carried out, the factor load values, common factor variances, item-total correlations and internal reliability coefficients have been given in Table 2.

Table 2. Results obtained after factor analysis regarding factors

		Load Value After Rotation					Reliability		
Item No.	Common Factor Variance	Factor 1	Factor 2	Factor 3	Factor 4	Corrected Item-Total Correlation	Internal Consistency	Spearman Brown	Test-re Test
3	0.69	0.80				.58			
2	0.63	0.74				.57			
4	0.66	0.74				.64	.86	.85	.84
1	0.59	0.70				.55			
5	0.56	0.64				.60			
13	0.54	0.51				.64			
24	0.76		0.80			.68			
22	0.77		0.78			.70			
23	0.78		0.77			.74	.89	.83	.83
21	0.75		0.74			.72			
25	0.47		0.58			.56			
15	0.76			.82		.58			
14	0.76			.78		.66	.80	.80	.90
20	0.55			.61		.55			
12	0.56			.45		.65			
17	0.67				.73	.57			
16	0.68				.69	.60	.70	.60	.85
18	0.54				.68	.43			
Reliability coefficients regarding the total scale							0.93	0.86	0.93

When Table 2 is examined, it is observed that six of the scale items (items 1, 2, 3, 4, 5 and 13) have accumulated under factor one, five (items 21, 22, 23, 24 and 25) under factor two, four (items 12, 14, 15 and 20) under factor three and three (items 16, 17 and 18) under factor four. When the items under each factor are evaluated for content and suitability to the structure, the items under the first factor
can be named as the sufficiency of teachers regarding measurement and evaluation "*method and technique determination"*, under the second factor as "*process review according to results*", under the third factor as "*data analysis and comments*" and under the fourth factor as "*giving feedback about the student*". The factor load values regarding the items that comprise the scale vary between 0.51 and 0.80 for the first factor, between 0.58 and 0.80 for the second factor, between 0.45 and 0.82 for the third factor and between 0.68 and 0.73 for the fourth factor. When the common factor variances of each item in this multi-factor structure are examined, it is observed that the values range between 0.47 and 0.78. According to these values, it can be stated that the variables form a homogeneous structure. When the factor load values are evaluated in terms of magnitude, it can be stated that items 12 and 13. are "mediocre" in terms of load values and that the other items range between "good" and "perfect" [15]. All these findings can be shown to be proofs that the construct validity of the scale is acceptable.

The item test correlations regarding the construct validity and homogeneity of the scale have been calculated. It has been determined that the item test correlations of the scale varied between 0.43 and 0.74. These values show that the items represent similar behaviors. Internal consistency (alpha) coefficient regarding the reliability of the scale has been calculated as 0.93. This value shows that the items that make up the scale are in accordance with each other. In addition, since during the trials carried out by taking out an item in turn there was no increase in the internal consistency coefficient that was calculated as 0.93, no item was taken out of the

scale [25]. Also, internal consistency reliability coefficients along with item test correlations for each factor were calculated and the results have been given in Table 2. The internal consistency reliability (alpha) coefficient for the scale in total was calculated as 0.93, the reliability coefficient regarding the first factor as 0.86, the second factor as 0.89, the third factor as 0.80 and the fourth factor as 0.70. Spearman Brown split-half correlation was calculated as 0.86, first factor coefficient as 0.85, second factor coefficient as 0.83, third factor coefficient as 0.80 and fourth factor coefficient as 0.60. Whereas it has been determined that the test-re-test reliability coefficient is 0.93, 0.84 for the first factor, 0.83 for the second factor, .90 for the third factor and .85 for the fourth factor. All these findings can be used as proofs showing that the scale has a satisfactory reliability.

The total scores of 395 teachers obtained from the scale have been arranged in increasing order in order to put forth the distinctiveness features for each of the 18 items of the scale. The total score averages of the teachers in the lower and upper groups have been compared for each item using t test.

When Table 3 is examined, it is observed that the t (sd = 212) regarding the differences in the item scores of the 27% lower and upper groups vary between 9.88 and 16.98. In addition, it has been determined that all items are significant at a level of $p < 0.001$. All values show that the reliability of the items in the scale are high and represent similar behaviors.

Table 3. Comparison of lower and upper groups for each item

Item No	N	Lower Group (%27)		Upper Group (%27)		t
		$\overline{X}$	S	$\overline{X}$	S	
3	107	3.29	0.99	40.62	0.64	110.64*
2	107	3.69	0.78	40.79	0.47	120.40*
4	107	3.08	0.83	40.52	0.6	140.32*
1	107	3.90	0.78	40.85	0.36	110.53*
5	107	3.14	0.90	40.53	0.65	130.03*
13	107	3.18	0.87	40.54	0.57	130.56*
24	107	3.34	0.86	40.66	0.51	130.72*
22	107	3.11	0.86	40.67	0.47	160.44*
23	107	3.16	0.80	40.67	0.47	160.82*
21	107	3.12	0.83	40.68	0.49	160.74*
25	107	3.17	0.94	40.54	0.59	120.86*
15	107	2.27	10.00	40.17	0.79	150.48*
14	107	2.61	0.94	40.45	0.61	160.98*
20	107	3.00	10.08	40.58	0.74	120.49*
12	107	3.06	0.87	40.53	0.61	140.43*
17	107	3.47	0.79	40.61	0.56	120.13*
16	107	3.28	0.89	40.74	0.44	150.19*
18	107	3.96	0.72	40.77	0.45	90.88*

*$p < 0.001$.

An additional confirmatory factor analysis (CFA) has been carried out in order to validate the construct validity of the scale developed via exploratory factor analysis (EFA). The fit indexes of the model obtained as a result of CFA have been examined and it has been determined that the chi-square value ($\chi 2 = 484.93$; N = 395; sd = 125; p = 0.00) is significant. Whereas the fit index values have been determined to be RMSEA = 0.08; NFI = 0.97; CFI = 0.97; IFI = 0.97; RFI = 0.96; GFI = 0.89; AGFI = 0.84 and SRMR = 0.05. These fit index values can be interpreted such that model is a good fit.

In addition, CFA was carried out for the single factor structure of the scale and the chi-square value ($\chi 2 = 898.52$, N = 395, sd = 131, p = 0.00) according to the fit indexes of the model has been determined to be significant. Whereas the fit index values have been determined as RMSEA = 0.13, NFI = 0.94, CFI = 0.94, IFI = 0.95, RFI = 0.93, GFI = 0.78, AGFI = 0.71 and SRMR = 0.07. According to these results, it can be stated that it is better to use the scale as multi-factored.

3. 2. Self-efficacy perception levels of social studies teachers regarding measurement and evaluation

The lowest and highest scores along with average scores, standard deviation values for the factors and the scale as a whole of the self-efficacy perception levels of social studies teachers regarding measurement and evaluation have been given in Table 4.

It is observed that scores close to the highest score that can be taken from all of the four dimensions of the scale have been taken when the replies of the social studies teachers to the "Measurement and Evaluation Self-Efficacy Perception Scale" have been taken into account. It can be stated that the self-efficacy perceptions of social studies teachers regarding all four factors and the total scale are high.

3. 3. The examination of the self-efficacy perception levels of social studies teachers according to various variables

One way ANOVA test was carried out in order to determine whether there was a statistically significant difference in the self-efficacy perception scores regarding gender, career seniority and the unit of duty after which the results obtained have been given in Table 5.

Table 4. Descriptive statistics of the replies of social studies teachers to the scale along with the factors and the whole scale

	N	Min.	Max.	$\overline{X}$	S
Factor1	395	9.00	30.00	24.15	3.80
Factor2	395	5.00	25.00	19.53	3.50
Factor3	395	4.00	20.00	14.39	3.27
Factor4	395	6.00	15.00	12.52	1.85
Total	395	29.00	90.00	70.60	10.37

Table 5. Anova results for the measurement and evaluation self-efficacy perception total scores of social studies teachers according to various variables

Variables		Sum of Squares	df	Mean Square	F	p
Gender	Between Groups	12.19	66	0.19	0.90	0.70
	Within Groups	67.62	328	0.21		
	Total	79.81	394			
Career Seniority	Between Groups	90.99	66	10.38	10.21	0.15
	Within Groups	375.15	328	10.14		
	Total	466.14	394			
Unit of Duty	Between Groups	58.62	66	0.89	0.86	0.77
	Within Groups	339.04	328	10.03		
	Total	397.66	394			

When Table 5 is examined, it has been observed that the total scores of social studies teachers regarding their answers to the "Measurement and Evaluation Self-Efficacy Perception Scale" have not caused any significant changes on the gender, career seniority and unit of duty variables (p>.05). Accordingly, it can be stated that gender, career seniority and unit of duty have no effect on the self-efficacy perception levels of social studies regarding measurement and evaluation.

4. Discussion, conclusion and suggestions

In this study, "Measurement and Evaluation Self-Efficacy Perception Scale" has been developed in order to determine the self-efficacy perceptions of teachers regarding measurement and evaluation. When a literature survey is carried out, studies to develop a scale regarding the measurement and evaluation efficiency have been found [26-34]. However, it has been observed that some of these studies have been carried out on teacher candidates [28, 34], that some have been carried out for efficacy regarding measurement tools [29], that some have been carried out for efficacy regarding alternative tools [26, 30], that some have been carried out before the structural change in the programs [33,34] and that construct validity studies in accordance with psychometric properties have not been carried out in these studies. The objective of this study is to eliminate these deficiencies in the field. First, a five-point Likert type scale has been prepared by making use of opinions of experts and

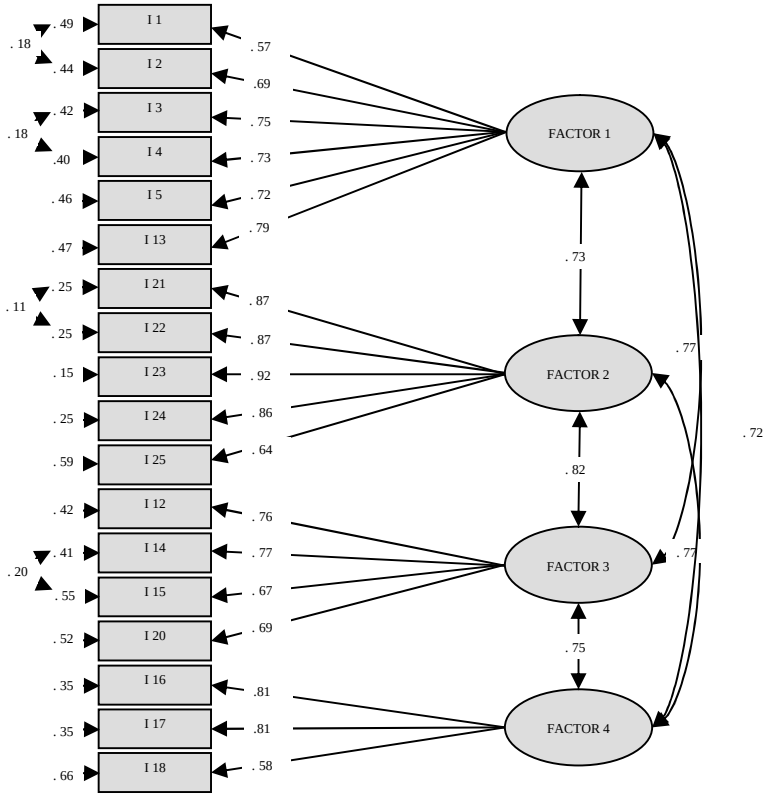


Fig. 2. Path diagram and parameter estimations regarding measurement and evaluation self-efficacy perception scale.

teachers along with data from relevant literature. The prepared scale was examined by experts regarding its comprehensibility, scope and face validity. Items with an agreement of 90-100% were accepted and 5 items that do not fit this criteria were eliminated after which construct validity studies for the 25 item scale were started.

Exploratory and confirmatory factor analyses were carried out for construct validity and it has been determined that the items forming the scale have been collected under four factors. It has been observed that these factors have a structure suited for naming and grouping. The items of the factors have been examined in terms of content and features. Accordingly, the items in the first factor have been named as "*method and technique determination*" regarding the measurement and evaluation of teachers, the second factor has been named as "*process overview*", the third factor has been named as "*data analysis and interpretation*" and the fourth factor has been named as "*giving feedback about the student*". The fact that all of the 18 items that make up the scale have high factor load values for their own factors and low values for the other factors has been accepted as an indication of factor independency. The model fit of the structure obtained by testing via exploratory factor analysis has been examined via confirmatory factor analysis. It has been determined as a result of the exploratory and confirmatory factor analyses that the model comprised of 18 items and four factors is appropriate both institutionally and statistically. In addition, according to the confirmatory factor analysis carried out to test both the multi-factor and single factor structure of the scale, it can be stated that using the multi-factor structure is proper since it has better fit values in comparison with the single factor structure.

As a result of the item analysis carried out, it has been observed that the corrected item-score correlations are at an ideal level. In addition, as a result of the *t* test carried out between the 27% lower and upper group scores, a significant difference has been determined for all items and sub-scales. The fact that the internal consistency, split-half test and test-re-test reliability coefficients are quite high puts forth that the items in the scale measure the same structure in consistence and accordance with each other. Thus, all these results are proofs that the properties measures by the "Measurement and Evaluation Self-Efficacy Perception Scale" are homogeneous and that all the items in the scale measure the same property [35]. In short, it shows that the scale developed is a valid and reliable tool. In this regard, the scale that has been developed can be used in further studies carried out to determine the measurement and evaluation self-efficacy levels of teachers. In addition, the self-efficacy levels of teachers in different fields regarding measurement and evaluation can be determined and it can be examined whether there are any differences or not.

In the study, the self-efficacy levels of social studies teachers regarding measurement and evaluation have also been determined and the status according to various variables have been put forth. It has been concluded that gender, career seniority and unit of duty have no effect on the measurement and evaluation self-efficacy perception levels. In a study carried out by Yaman [27] and Toptas [26] it has been put forth that gender and career seniority has no effect on the measurement and evaluation efficacies of teachers. Similarly, a study carried out by Yaman and Karamustafaoglu [28] on teacher candidates has put forth that gender has no statistically significant effect on the measurement and evaluation self-efficacy perceptions. It has been determined that the self-efficacy perception levels of social studies teachers are quite high both in terms of factors and in general. Gelbal and Kelecioglu [32] have obtained similar results in their study carried out on class and branch teachers regarding their measurement and evaluation teacher efficacies. However, it is also observed that this result of the study is in conflict with many other studies [31, 33-37]. This can be explained by the difference of the studied groups, the fact that it is a group of teachers who are trying to develop and renew themselves by participating in in-service activities and the fact that teachers are better equipped now with the widespread use of information technologies in recent years. According to all these results, it can be stated that in-service training activities to increase the measurement and evaluation efficacies of teachers are effective. To this end, it

can be beneficial for teachers to attend in-service training activities to increase their career efficiencies. In addition, it can also be beneficial to increase the interaction of teachers via the use of computers and internet.

References

[1] Erturk S. Egitimde program gelistirme. Meteksan Publications, Ankara, 1994 [in Turkish].
[2] Tay B. Learning strategies in social studies text books. J Kirsehir Educ Fac 2005;6:209–225.
[3] Kahyaoglu M, Yangin S. Views of prospective teachers in elementary school teaching departments about professional self-efficacy. Kastamonu Educ J 2007;15:73–84.
[4] Yalin HI. Ogretim teknolojileri ve materyal gelistirme. Nobel Publication Distribution, Ankara, 1999 [in Turkish].
[5] Kutlu O, Dogan CD, Karakaya D. Ogrenci basarisinin belirlenmesi performansa ve portfolyoya dayali durum belirleme. Pegem A Publications, Ankara, 2008 [in Turkish].
[6] Turgut MF. Egitimde olcme ve degerlendirme metotlari. Saydam Printing, Ankara, 1988 [in Turkish].
[7] Ural M, Erdogan H, Ural M. Egitimde olcme ve degerlendirme (istatistik uygulamali). Emel Printing, Ankara, 1998 [in Turkish].
[8] Tekin H. Egitimde olcme ve degerlendirme. Yargi Publishing, Ankara, 2004 [in Turkish].
[9] Eyitmis AN. Searching for the proficiency of the teachers at secondary schools in using the techniques of measurement and evaluation in an active way (sample of Kahramanmaras). Kahramanmaras Sutcu Imam University, Master's Thesis, Kahramanmaras, 2007 [in Turkish].
[10] Gumus B. Egitimde olcme ve degerlendirme. Ankara: Kalite Printing, 1977 [in Turkish].
[11] Karahan U. Application of alternative measurement and evaluation methods that are grid, diagnostic tree and concept maps within biology education. Gazi University, Unpublished Master's Thesis, Ankara, 2007 [in Turkish].
[12] Yetkin D, Dascan O. Ilkogretim programi 1–5. Ankara: Ani Publishing, 2006 [in Turkish].
[13] MEB. Ogretmen yeterlilikleri: Ogretmenlik meslegi ozel ve genel alan yeterlilikleri. Directorate of State Books, Ankara, 2008 [in Turkish].
[14] Fraenkel JR, Wallen NE. How to design and evaluate research in education. McGraw-Hill, Boston, 2006.
[15] Tabachnick BG, Fidel LS. Using multivariate statistics. Allyn & Bacon, Needham Heights, MA, 2001.
[16] Daniel LG. Comparisons of exploratory and confirmatory factor analysis. Paper presented at the Annual Meeting of the Mid-south Educational Research Association, Little Rock, AR, 1989.
[17] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 1999;4:272–299.
[18] Gerbing DW, Anderson JC. Monte Carlo evaluations of the goodness-of-fit indices for structural equation models. (Ed: Bollen K.A. & Long J.S.), Testing structural equation models. Sage, Newbury Park, CA, 1993.
[19] Cokluk O, Sekercioglu G, Buyukozturk S. Sosyal bilimler icin cok degiskenli istatistik. Pegem Academy Publishing, Ankara, 2010 [in Turkish].
[20] Schermelleh-Engel K, Moosbrugger H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods Psycholog Res Online 2003;8:23–74.
[21] Jeong J. Analysis of the factors and the roles of HRD in organizational learning styles as identified by key informants at selected corporations in the Republic of Korea. Texas A&M University. Unpublished Doctoral Thesis, Texas, USA, 2004.
[22] Buyukozturk S. Sosyal bilimler icin veri analizi el kitabi. Pegem A Publications, Ankara, 2006 [in Turkish].
[23] Sencan H. Sosyal ve davranissal olcumlerde guvenirlik ve gecerlik. Ankara: Seckin Publications, 2005 [in Turkish].
[24] Kline P. An easy guide to factor analysis. Routledge, London, 1994.
[25] Ozdamar K. Paket programlar ile istatistiksel veri analizi I. Anadolu University Faculty of Science Publications, Eskisehir, 1997 [in Turkish].
[26] Toptas V. Classroom teachers' perceptions about the use of alternative assessment and valuation methods in mathematics courses. Educ Sci 2011;36:205–219.
[27] Yaman S. Teachers' perceptions about their measurement and evaluation practices in science and technology course. Element Educ Online 2011;10:244–256.
[28] Yaman S, Karamustafaoglu S. Investigating prospective teachers' perceived levels of efficacy towards measurement and evaluation. Ankara Univ J Fac Educ Sci 2011;44:53–72.
[29] Yildirim Ekinci H, Koksal EA. Development of elementary science and mathematics teachers' competence in measurement and evaluation scale. Kastamonu Educ J 2011;19:167–184.
[30] Yayla G. Fen ve teknoloji ogretmenlerinin tecrubeleriyle alternatif olcme ve degerlendirme yaklasimlarina yonelik oz yeterlilikleri arasindaki iliski. 2nd International Conference on New Trends in Education and Their Implications, Antalya, Turkey, April 27-29, 2011 [in Turkish].
[31] Kilmen S, Cikrikci Demirtasli N. The perceptions of primary school teachers about their application levels of measurement and evaluation principles. Ankara Univ J Fac Educ Sci 2009;42:27–55.
[32] Gelbal S, Kelecioglu H. Teachers' proficiency perceptions of about the measurement and evaluation techniques and the problems they confront. HU J Educ 2007;33:135–145.
[33] Cakan M. Comparison of elementary and secondary school teachers in terms of their assessment practices and perceptions toward their qualification levels. Ankara University J Fac Educ Sci 2004;37:99–114.
[34] Karaca E. Development of a Likert type competence perception scale towards measurement and evaluation competencies of teacher candidates. Dumlupinar Univ J Soc Sci 2003;9:179–198.
[35] Tavsancil E. Tutumlarin olculmesi ve SPSS ile veri analizi. Nobel Publication Distribution, Ankara, April 27–29, 2002 [in Turkish].
[36] Gunes A. Measurement and evaluation capabilities of class teachers according to their own perceptions. Marmara University, Master's Thesis, Istanbul, 2007 [in Turkish].
[37] Ulutas S. Investigating the competency of teachers in high schools in measurement and evaluation and level of application principles of measurement and evaluation. Ankara University, Master's Thesis, Ankara, 2003 [in Turkish].