

A Developmental Study on Evaluating the Performance of Preschool Education Institution Teachers with 360 Degree Feedback*

Asude Balaban Dagal¹, Rengin Zembat¹

¹Marmara University Ataturk Education Faculty, Turkey

Correspondence: Asude Balaban Dagal, Marmara University Ataturk Education Faculty, Turkey.

Received: April 19, 2017

Accepted: May 18, 2017

Online Published: May 23, 2017

doi:10.11114/jets.v5i6.2365

URL: <https://doi.org/10.11114/jets.v5i6.2365>

Abstract

In accordance with the subject of this research, three different scales to measure preschool teachers' performances with respect to school principals', parents' and children's point of view were developed for the research. 51 principals, 893 parents and 256 children participated in the research.

"Teacher Performance Evaluation Scale – Administrators' Form" (TPESAF), "Teacher Performance Evaluation Scale – Parents' Form" (TPESPF) and "Teacher Performance Evaluation Scale – Children's Form" (TPESCF) are the scales that were prepared and used by the researcher. The data obtained were analysed by "SPSS 13 for Windows". The main results of the research are as follows: Research, an exploratory factor analysis was carried out for each of the scales developed. In order to show the validity of the scales, the overall and sub-test content validity analysis and differentiation power of the (TPESAF), (TPESPF) and (TPESCF) were calculated. Following the factor analysis which gave the scales their final shape, statistical reliability analysis was carried out. In this study, Cronbach Alpha coefficients of the (TPESAF), (TPESPF) and (TPESCF) and their sub-tests' means and standard deviations were calculated. Furthermore the test re-test reliability analysis was carried out. The scales were found to be reliable according to the reliability analysis (Cronbach Alpha Coefficients of Principal's Form: 0,94, Parents' Scale: 0,93, Children's Scale: 0,84).

Keywords: preschool, performance evaluation, performance evaluation in education, 360 degrees, preschool teacher performance evaluation

1. Introduction

The changing world, globalism come with 21st century and competition occurred with the effect of globalism caused changes in management area in education as well as in other fields. These changes not only affected the way of management but also caused to emerge a perception of management including the administrators and also others who are affected by education institutions.

Schools are the institutions which played an important role on systematic training of young generation and transferring the knowledge and culture throughout ages, are still maintaining this important role to significant extent today (Oktay, 1999, p.202). Children are spending a significant part of their days in schools. This situation creates school environment as a common entry point for serving and intervening for a huge number of children (Pahl & Barrett, 2007, p.6).

Evans (1996) claim that preschool education has long term effects on a child's development and the quality of the education given in this period is important (cited; Yavuzer, 2003, p. 61). Also, hundreds of research showed that qualified education given in the early period of life has positive effects on children's self-esteems, motivation levels and social behaviours (Boyd, Barnett, Bodrowa, Leong, & Gomby, 2005, p. 6). It is very important for a preschool institution to be useful, safe and in a position of responding needs; and to be an environment providing movement, play, social interaction, music and art (Oktay, 1999, p. 203). Given the poor quality of much child care, it might instead produce mild negative consequences (Barnet, 2008).

In our country 1.115.818 children are attending preschool education centres according to the data of 2010-2011 education year of the Ministry of Education. Children who spend significant part of the day in schools have to use the

*Based on a doctoral thesis from Marmara University, Ataturk Faculty of Education, Preschool Education.

time spent there to be effective and efficient. This time period is essential for children and should not be wasted. It is possible to ensure this with creating qualified education environments.

A preschool teacher is the closest one to the children in a group in school environment and responsible for the education and training in the school (Zembar, 2005, p.31). In this case, to ensure the preschool institutions to be qualified is highly related to the teachers' personal characteristics, competencies, interest towards his/her profession, knowledge and experience (Oktay & Polat, 2008, p.6).

To understand to what extent education system and school managements are fulfilling their objectives and to reach organizational goals at a higher level, there is a need for supervision. However, it should not be forgotten that, supervision alone may not be enough to reach organizational goals (Taymaz, 1993, p.5). For this reason, correct, versatile and objective evaluations are needed to provide an ongoing development and progression.

In a research it was stated that for public and private schools to fulfil their functions is up to the teachers' performances in these schools (Tamam, 2005, p.24). At this point, it is important to note what performance assessment is. Performance assessment means studies intended for defining one's effectiveness and success level on a specific topic. Performance assessment is reviewing all directions like the one's studies, activities, deficiencies, competencies, redundancies and incompetence regardless of the duty in the institution (Findikci, 1999, p.297).

With performance assessment workers' faith and loyalty in organizational culture increase and their motivation is generated. When related literature is examined it is found out that there is a relationship between performance and motivation (Cicek, 2005, p.27; Barutcugil, 2002, p.43; Vroom, 1970) Brown (2005, p.172) also stated that increasing worker's organizational productivity and motivation are among the reasons for using performance assessment in organizations. Also Tamam (2005, p.24), supports the view that assessing teacher's performance is necessary for determining deficiencies and errors in education, for specifying the problems, motivating teachers and taking precautions for the future. At this point, the institutions, in which children are being raised, are wanted to be open to improvement more than any other institution. For this reason, various plans and programs to ensure the improvement of the institution (school) must be needed (Erdogan, 2008, p. 102).

School is different from non-education institutions regarding the administrator's and staff's characteristics. In most of the non-educational institutions there can be significant differences between the factors like education, intellectual capacity and cultural background of managers and subordinates on behalf of the managers. Whereas in schools there is not such big difference between school administrators and teaching staff (Erdogan, 2008, p. 122). Performance assessment approach which can be applied in such cases is encountered in the literature. This approach is usually named as performance assessment through 360 degree feedback.

This method is an approach where combined assessment is done. It is named 360 degree assessment approach because various people and criterion are being used. This method provides the opportunity to workers and administrators to evaluate themselves and each other. 360 degree assessment method includes workers assessment as well as evaluating administrators' performances by workers, subordinates and chiefs. In addition, the method includes self-evaluation. In 360 degree assessment approach it is possible to obtain various data from different evaluators (Bulut, 2004, p. 7). The fact of having more evaluators provides to have more objective results and it makes the process more defendable. The most significant disadvantage is that the assessing process takes a long time and is costly because of the high number of evaluators (Bingol, 2010, p. 398).

The study is aimed to be a research to provide to take a step to form a performance management system which includes objective assessments for preschool education institutions. In this study, in order to reach the solution of the problem, it is planned to develop performance assessment forms which will give similar results as possible when applied by everyone who knows the application. In Turkey, there are applications of some private schools on performance assessment. However, it is observed that these applications are not scientific and forms are being filled in the framework of personal judgments.

From this viewpoint, creating performance assessment scales that would evaluate preschool teachers' performance and which are considered to be lacking is the main aim of this study.

Answers to the following questions have been searched within the general aim of the research:

1. Is "Teacher Performance Evaluation Scale – Administrators' Form" (TPESAF)" prepared for preschool teachers, valid and reliable?
2. Is "Teacher Performance Evaluation Scale – Parents' Form" (TPESPF) prepared for preschool teachers, valid and reliable?
3. Is "Teacher Performance Evaluation Scale – Children's Form" (TPESCF) prepared for preschool teachers, valid and reliable?

2. Method

2.1 Research Method

Scale development method was used in this study. The aim of the study is to develop the preschool teachers' evaluation scales to score their performance by 360 degree method.

2.2 Study Group

The study group of the research consists of administrators, parents and children aged 4-6 years, who evaluated the forms of preschool teachers working in kindergartens of Ministry of Education's public primary schools, independent public preschools and private preschools. 281 teachers attended for administrator form, 893 parents attended for parent form, and 256 children attended for child form.

After collecting experts' opinions on the forms before starting the study, the districts which would be in the study were chosen by simple random sampling method from Istanbul central districts. At least 4 primary schools, 1 independent public school and 2 private preschools from each districts of Kadikoy, Atasehir, Maltepe, Umraniye, Bakirkoy, Sisli, and Besiktas were participated in the study. Schools were chosen by simple random sampling method.

2.3 Data Collection Tools

"Teacher Performance Evaluation Scale – Administrators' Form" (TPESAF), "Teacher Performance Evaluation Scale – Parents' Form" (TPESAF), "Teacher Performance Evaluation Scale – Children's Form" (TPESCF) which developed by the researcher, were used as forms in the study.

While the forms were prepared, teacher qualifications and parental expectations in the field of preschool education, literature information about the development of performance assessment forms in the field of business, scales used in the field, relevant teacher and managerial opinions were taken as basis.

The general characteristics of **TPESAF**; in the process of developing the form, performance assessment forms of various institutions were examined as well as the literature knowledge. Also, performance assessment forms of two universities outside Turkey (University of California at Santa Barbara, 1997; Government of the Northwest Territories Municipal and Community Affairs, 2009) have been effective in the process of developing the scale. In addition, teacher and managerial opinions on performance assessment method and forms were informally examined in the field and were used during preparation of the forms. In order to develop the data collecting tool, first, opinions from preschool education institution administrators and preschool teachers were asked by using "Question List Regarding the Administrators' and Workers' (teachers) Interpretations about Performance Assessment" which consists 18 questions. Based on the obtained opinions and literature, 10 itemed forms, where several different statements regarding each item are involved, were developed by the researcher to assess teacher performance.

The scale consists of 10 questions. For each question 4 different teacher profile is identified. Each profile is evaluated with various explanations. Administrators are asked to identify and evaluate the profiles which they think their teachers belong to. It was concluded that there were no expert opinions to remove but it was discussed that there were some items to be changed in the profile section. As a result of the expert opinions received, profiles were rearranged according to the questions. This forms the structure validation of the research. As a result of the analysis of other validity and reliability studies, it was found out that the scale consisted one dimension and its internal validity was 0,94.

TPESPF was designed by the researcher as an item pool consisting 91 items. The correlations between the characteristics to be measured and scale items are concerning the validity of the assessment tool. Prior studies are needed to identify the strength of scale item inclosing the characteristic aimed to be measured (content validity) and the item's strength of predicting the structure (construct validity) (McGartland, R.D, 2003; cited Yurdugul, 2005, p.1). Other factors like, scale items being understandable, its conformity for the target group etc., which are affecting validity of the assessment tool are the points to be considered. Consistency/inconsistency between experts' opinions obtained in prior studies is also used for estimate criteria for content and construct validity (Yurdugul, 2005, p.2).

To identify the content validity of the scales Lawshe method was used in this study. Numbers of questions were decreased to 56 from 91 as a result of the views taken from 6 experts within content validity and Lawshe method was used. 35 questions were deleted from the pool. The scale was designed in 4 point likert type scale.

At the end of other validity and reliability analyses the scale was descended to 33 items. At the end of the factor analysis done within the scope of validity analysis, scale was divided into 4 subgroups and internal consistency coefficient of the whole scale was 0,91.

TPESCF is a 3 point likert type scale consisting 19 items. General characteristics of the child form were;

- Simple sentences which children can easily comprehend are prepared.

- Questions not seeking answers indirectly and concern children individually have taken place.
- Questions were designed to be presented by showing facial expressions like in any other child form.
- It was noted that the questions reflect teacher behaviours.

While applying the forms, the question was read from the form to the child and markings were done on the form. However, in order for the child to express his/her opinion comfortably, they were asked to choose one of the cards, which they can relate every question, with “happy”, “indecisive” and “sad” facial expressions on them.

After factor analysis, scale was divided into 3 subcategories and the internal consistency coefficient of the whole scale was found as 0,84.

2.4 Collecting the Data

The distribution of TPESAF, TPESPF and TPESCF was done by the researcher. Information about the purpose and importance of the research and how to apply the forms were explained to the administrators of each school after 40-45 minutes of meetings. Afterwards, with the permissions of the administrator, interviews were done with the teachers and they were informed about the study. With the studies done with children, after permissions taken from the administration and teachers, a place was set outside of the classroom to be alone with the child. Researcher applying the form sat side by side with the child at a small table. The child was taken from his/her classroom by the person who applied the form. Cards with happy-indecisive-sad facial expressions were put on the table. First, the child was asked what s/he saw. Then, it was told them that they were going to have conversation about his/her teacher and this conversation would remain secret between them. The person applying the form asked questions about the child and asked him/her to show the answers through the cards before applying the questions on the form. These questions asked before applying the form were as follows: “Do you like riding bicycles? Do you like eating ice cream?” With the questions child was directed to the cards in this way: “Showing the happy face –I always like riding bicycle. Showing indecisive face –I sometimes like riding bicycle. Showing the sad face –I never like riding bicycle”. These questions can be increased until child understands the structure of questions. However, in our group usually it wasn’t needed to ask the second question. After this step, questions are asked to children by using similar statements appropriate to the questions in the form.

2.5 Analysis of Data

After collecting data for examining the preschool teacher performance assessments in preschool institutions, the data collected were transferred into “SPSS 13.0 for Windows” packet program and analysis is performed according to the following sequence of operations:

In order to test whether TPESAF, TPESPF and TPESCF are valid assessment tools or not, exploratory factor analysis was performed on the collected data. To identify adequacy, Sampling Adequacy Test (Kaiser-Meyer-Olkin-KMO); for adequacy of each item for factor analysis Measures of Sampling Adequacy (MSA) test; to identify whether data is coming from multivariate normal distribution Bartlett’s Test of Sphericity were applied.

While testing whether “TPESAF”, “TPESPF” and “TPESCF” are reliable evaluation tools or not, in order to identify internal consistency coefficient between the scores obtained from the scale, Cronbach Alpha coefficient and if-item-deleted Cronbach Alpha coefficient of each sub-factors and whole scale were measured. In addition, corrected item total correlations were measured in order to see to what extent scale items distinguished the scale scorers. Also, it was examined with the independent groups t-test if the difference between item average scores of lower %27 and upper % 27 groups formed according to total scores obtained from the scale, was significant.

Test-retest method was applied in order to identify the reliability in term of stability of “TPESAF”, “TPESPF” and “TPESCF”.

Data obtained from “TPESAF”, “TPESPF” and “TPESCF” were weighted in terms of evaluators and sub-factors according to the expert opinions and all factors were turned into one single score. To make this weighting, 20 field expert teachers, school administrators, professional managers responsible for staff management were consulted. Opinions at the extreme points were removed and the averages of the remainder were taken. After this point, study was carried out through total performance scores and sub-dimension scores.

3. Findings

3.1 Findings on Validity of “TPESAF”

In the validity study of the scale, content and construct validity were applied. Content validity of the assessment tool was identified by detecting the scale items’ conformity/validity level by taking expert opinions.

Within the scope of “TPASAF” five expert assistant professors from universities were consulted for the content validity. According to five expert opinions it was not needed to remove any items. However, some items were changed in the

direction of the experts. In order to examine construct validity of the scale, exploratory factor analysis was done with SPSS 13.0. Exploratory factor analysis was done with principal components method. As a result of exploratory factor analysis, varimax rotation method was used to obtain interpretable factors which can be named. Sampling Adequacy Test (Kaiser-Meyer-Olkin-KMO) was performed to identify adequacy for factor analysis of data obtained from administrators' assessments of 281 preschool teachers. Measures of Sampling Adequacy (MSA) test was used for each item's conformity for factor analysis. Bartlett's Test of Sphericity was applied to determine if data was deriving from multivariate normal distribution. Data set was found conformable for factor analysis due to KMO value, which determines if the sampling is adequate for factor analysis, was above 0,50 and Barlett test was significant at 0,05 significance level in the obtained findings (KMO=,940, $\chi^2_{\text{Barlett test}}(45)= 1933,326$; $p=,000$). These values showed that data's conformity for factor analysis was on a perfect level and according to KMO value scale variables can predict one another without any errors. Measure of Sampling Adequacy (MSA) values, which defines the items' conformity for factor analysis, were examined. In case of Measure of Sampling Adequacy (MSA) values are under 0,50, the question should be removed from the analysis (Sipahi, Yurtkoru & Cinko, 2006, p.81). However, since the value of any items was not under 0,50, no item has to be removed from the scale at this point. In the factor analysis using varimax rotation to define primary components, "TPASAF" consists of a single sub-dimension and % 64,65 of total variance is explained by one dimension. % 30 and more of the variance explained on single factor scales can be se considered adequate (Buyukozturk, 2007, p.125). On this scale explained variance is well over % 30.

Finally, in construct validity for TPESAF correlation between items were examined. Data obtained for study group is presented in the table below.

Table 1. Correlation Values of Data Obtained from Study Group in Which Construct Validity of "TPASAF" Assessment Form was Applied

		m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	Av.
Program Planning and Organization (m1)	r	1	,677**	,657**	,546**	,617**	,596**	,639**	,566**	,481**	,615**	,797**
	p		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	277	277	277	276	276	277	276	277	277	274	277
Classroom Layout (m2)	r	,677**	1	,713**	,610**	,565**	,603**	,611**	,525**	,479**	,633**	,799**
	P	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	277	281	281	280	280	281	280	281	281	277	281
Classroom Domination (m3)	r	,657**	,713**	1	,570**	,601**	,648**	,650**	,538**	,464**	,574**	,801**
	P	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000
	N	277	281	281	280	280	281	280	281	281	277	281
Child Hygiene (m4)	r	,546**	,610**	,570**	1	,589**	,545**	,547**	,516**	,487**	,569**	,742**
	P	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000
	N	276	280	280	280	279	280	279	280	280	276	280
Communication with Parents (m5)	r	,617**	,565**	,601**	,589**	1	,713**	,667**	,685**	,601**	,603**	,829**
	P	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000
	N	276	280	280	279	280	280	279	280	280	276	280
Communication with Child (m6)	r	,596**	,603**	,648**	,545**	,713**	1	,677**	,616**	,556**	,615**	,820**
	P	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000
	N	277	281	281	280	280	281	280	281	281	277	281
Monitoring Child's Development(m7)	r	,639**	,611**	,650**	,547**	,667**	,677**	1	,654**	,597**	,682**	,841**
	P	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000
	N	276	280	280	279	279	280	280	280	280	276	280
Communication with Administration (m9)	r	,566**	,525**	,538**	,516**	,685**	,616**	,654**	1	,731**	,599**	,808**
	P	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000
	N	277	281	281	280	280	281	280	281	281	277	281
Workplace Adjustment (m10)	r	,481**	,479**	,464**	,487**	,601**	,556**	,597**	,731**	1	,556**	,748**
	P	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000
	N	277	281	281	280	280	281	280	281	281	277	281
Individual Development	r	,615**	,633**	,574**	,569**	,603**	,615**	,682**	,599**	,556**	1	,808**
	P	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000
	N	274	277	277	276	276	277	276	277	277	277	277
Average	r	,797**	,799**	,801**	,742**	,829**	,820**	,841**	,808**	,748**	,808**	1
	P	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	
	N	277	281	281	280	280	281	280	281	281	277	281

As seen in Table 1, correlations between items were found to be significant between 0,464 and 0,829 and correlations between factors and total scale were between 0,742 and 0,841, positive and had 0,01 level of significance ($p<0,01$). All

items were related with each other at $p < 0,000$ level. Since correlations of all items are much higher than 0,30, it can be concluded that items distinguish individuals on a high level.

3.2 Findings on Reliability of "TPESAF"

Whole scale's Cronbach Alpha coefficients and if-item-deleted Cronbach Alpha coefficients were calculated to identify internal consistency coefficient of the scores obtained from the scale. In addition, corrected item-total correlations were measured in order to define to what extent scale items distinguish the scorers. Then, it was tested with independent groups t-Test whether the difference between item average scores of lower %27 and upper %27 groups, which were formed according to total scores obtained from the scale, was significant or not.

As a result of the reliability analysis, internal consistency coefficient of "TPESAF" was found 0,94. As a result of if-item-deleted analysis, it was found out that there was not any change in internal consistency coefficient when an item was removed

Arithmetic average of upper %27 group is ranging from 3,97 to 3,85, whereas arithmetic average of lower %27 group ranges from 2,89 to 2,53. Also, standard deviation of upper %27 group is between 0,352 and 0,81 whereas, lower %27 group is between 0,670 and 0,380.

In item discrimination procedures, when the difference between item averages of lower %27 and upper %27 group formed according to scale's total scores, was analysed with independent group t-Test, discrimination index of each item was found significant over 0,000 level statistically. Any of the items has to be removed from the scale because this finding suggests that reliability of scale items were at high levels and it was distinguishable in terms of characteristics to be measured for scale scorers.

When findings were evaluated, as a result of "TPASAF"'s application performed in study group, it can be said that high levels of internal consistency was found and scale provided reliable results for study group.

In order to identify reliability in terms of stability test-retest method was applied. It was applied in two weeks intervals for 31 teachers who are working in public and private schools in districts stated in Istanbul city method department. Pearson Correlation Coefficient and dependent groups t-Test results were examined to identify stability coefficient of the scale's two separate applications. According to findings obtained from test-retest of "TPASAF" arithmetic average of pre-test is 3,06, arithmetic average of post-test is 3,19, standard deviation is 0,75 in pre-test and is 0,84 in post-test. Pearson Correlation Coefficient value between two applications is 0,89. This finding shows that relationship between two separate applications of the scale is positive and significant at 0,01 level ($p < 0,01$). As a result, findings pointed out that consistency of test-retest of the scale, on a total scale basis, was high. Dependent groups t-Test results obtained from test-retest applications which were done to identify reliability in terms of stability of the scale were examined. It suggested that there was not a significant difference, level of significance was 0,05 ($p > 0,05$), as a result of dependent groups t-Test applied between average scores in total scale basis. This result showed that external consistency which identifies reliability in terms of stability of the scale was found.

3.3 Findings on Validity of "TPESPF"

Content and construct validity were investigated in the validity study of the scale. Content validity of the assessment tool was defined by identifying conformity/validity of the items on the scale according to expert opinions. Some items have been modified only in the way that experts have indicated. As a result of opinions received from 6 experts, within the context of content validity, number of items were decreased to 56 from 91 since some of the questions in the pool were valued under 0,99 using Lawshe method.

In order to examine construct validity of the scale, exploratory factor analysis was done with SPSS 13.0. Exploratory factor analysis was done with principal components method and varimax rotation method was used. Thus, construct validity of the scale was examined by exploratory factor analysis. Sampling Adequacy Test (Kaiser-Meyer-Olkin-KMO) was used to identify adequacy for factor analysis of data obtained from 893 parents' assessments of preschool teachers. Measures of Sampling Adequacy (MSA) test was applied for each item's conformity for factor analysis. Bartlett's Test of Sphericity was performed to determine if data was deriving from multivariate normal distribution. Data set was found conformable for factor analysis due to KMO value, which determines if the sampling is adequate for factor analysis, was above 0,50 and Bartlett test was significant at a 0,05 significance level in the obtained findings. ($KMO = 0,940$, $\chi^2_{\text{Bartlett test}}(528) = 11498,568$; $p = 0,000$). Keiser (1974), claimed that values above 0,50 were acceptable values. Also, values between 0,5 and 0,7 are acceptable, between 0,7 and 0,8 are good, between 0,8 and 0,9 are very good and values above 0,9 are considered as perfect values (Field, 2009, p. 647). These values showed that data's conformity for factor analysis was on a perfect level and according to KMO value scale variables can predict one another without any errors.

Measure of Sampling Adequacy (MSA) values that identifies every each item's conformity for factor analysis were

examined. In the cases where Measure of Sampling Adequacy (MSA) values are under 0,50 the question should be removed from the scale (Sipahi, Yurtkoru & Cinko, 2006, p.81). However, since the value of any item was not under 0,50, no item has to be removed from the scale at this point.

Variance ratios and eigenvalues of “Teacher Performance Assessment Scale Parent Form” are presented in Table 2 below.

Table 2. Variance ratio and Eigenvalue of “TPESPF” and its Sub-Dimensions

Factor	Eigenvalue	Variance	Cumulative Variance
1	11,424	34,619	34,619
2	3,097	9,385	44,004
3	1,571	4,761	48,765
4	1,373	4,160	52,925

As seen in table 2, “TPESPF” was divided into four sub-dimensions in the factor analysis using varimax rotation in order to identify principal components and total variance was % 52,925. As shown in Table 3, variance ratios explained by “TPESPF” sub-dimensions are as follows; variance ratio explained by first factor with eigenvalue of 11,424, is % 34,619; variance ratio explained by second factor with eigenvalue of 3,097, is %9,385; variance ratio explained by third factor with eigenvalue of 1,571, is% 4,761 and variance ratio explained by fourth factor with eigenvalue of 1,373 is %4,160. Variance ratios ranging between %40-%60 in factor analysis are accepted as ideal (Scherer, 1988; cited Erdogan, Bayram and Deniz, 2007, p.6). In this case, total variance of %52,925 was agreed to be ideally acceptable.

Factor weights of the scale as a result of factor analysis, variance ratios explained by factors are presented in Table 3 below.

Table 3. Results of Exploratory Factor Analysis of Data Obtained from Study Group in Which “TPESPF” Construct Validity Was Applied

Item no	Factor no	Factor Weight Values	Item no	Factor no	Factor Weight Values	Item no	Factor no	Factor Weight Values	Item no	Factor no	Factor Weight Values
q25	1	,694	q51	2	,881	q09	3	,702	q39	4	,823
q32	1	,693	q52	2	,842	q11	3	,697	q40	4	,792
q24	1	,687	q50	2	,804	q07	3	,683			
q33	1	,681	q53	2	,762	q16	3	,615			
q31	1	,640	q41	2	,620	q14	3	,573			
q29	1	,634	q04	2	,587	q08	3	,572			
q34	1	,575	q01	2	,530	q10	3	,544			
q27	1	,569	q55	2	,499	q05	3	,517			
q12	1	,555	q22	2	,451	q03	3	,499			
q35	1	,545									
q30	1	,541									
q17	1	,503									
q26	1	,440									

When table 3 is examined, it can be seen that 33 items forming the scale were gathered under 13 items in first sub-factor, under 9 items in second sub-factor, under 9 items in third sub-factor and under 2 items in fourth sub-factor as a result of exploratory factor analysis. Factor weight value is a coefficient which explains items' relationship with sub-dimensions. In literature, to form factor pattern factor weights ranging from 0,30 to 0,40 can be taken as under cutting point (Tavsancil and Keser, 2001, p. 87; cited. Tok, Tok, Mazi, p.130). For this reason, 0,40 was accepted as undercutting point in this study. When first results of factor analysis were examined it was observed that factor weight values of some items were below 0,40 or they had high weight values in both factors. In line with these criteria, factor analysis was applied one more time after each item was deleted from the scale. By this way 22 items were removed from the scale. As a result of the analysis “TPESPF” consisting of four factors and 33 items, was put into its final form. When factors' weight values are examined, it can be seen that factor weight values of items forming “trust on the teacher” factor were ranging from 0,494. to 0,440; items forming “family cooperation” factor were ranging from 0,881 to 451; items forming “parent communication” factor were ranging from 0,702 to 0,499and items forming “continuance of education” factor were ranging from 0,823 to 0,792.

In the construct validity study of “TPESPF”, lastly correlation between factors and correlation between factors and total scale were examined. Findings for study group were presented below in table 4.

Table 4. Correlation Values of Data Obtained from Study Group in Which “TPESPF” Construct Validity Study was Applied

		F1	F2	F3	F4	p
F1	r					
F2	r	0,52**				0,00
F3	r	0,69**	0,44**			0,00
F4	r	0,40**	0,37**	0,31**		0,00
Total	r	0,84**	0,86**	0,74**	0,52**	0,00

**p< 0,01 level (2-tailed).

As seen in Table 4, correlation between factors were ranging from 0,52 to 0,69, between factors and total score were ranging from 0,86 to 0,52 and they were positive and significant with significance level of 0,01 (p<0,01). This finding was interpreted as there was a positive significant relationship between the factors which forms the scale for study group and factor-total scale scores.

3.4 Findings on “TPESPF” Reliability

In order to identify internal consistency coefficient between the scores obtained from the scale, Cronbach Alpha coefficient and if-item-deleted Cronbach Alpha coefficient were measured. In addition, to define to what degree scale items distinguish scale scorers corrected item-total correlations were applied. Then, whether difference between item average scores between lower %27 and upper %27 groups was significant or not was examined with independent groups t-Test.

First, “TPESPF” internal consistency coefficient was found 0,93 after Cronbach- alpha test. When if-item-deleted Cronbach- alpha test was applied, it was seen that there was not any difference in reliability level when any item was deleted so items were left out of the analysis.

Cronbach alpha analysis was repeated for every four dimensions. Results of the analysis can be summarized like; Cronbach Alpha coefficient of first sub-dimension of “TPESPF” was 0,905, second sub-dimension was 0,885; third sub-dimension was 0,834; fourth sub-dimension was found 0,834. However, fourth sub-dimension had 2 sub-dimensions so if-item-deleted Cronbach Alpha coefficient was not examined.

Independent group t-Test results between item average scores, standard deviation, item based arithmetic average of lower %27 and upper %27 groups formed according to TPASPF total scores of study group were done either. According to the results arithmetic average of upper % 27 group is ranging from 3,87 to 1,88 whereas arithmetic average of lower %27 group is between 4,00 and 3.81. On the other hand, standard deviation of upper %27 group is between 1,09 and 0,39 and standard deviation of lower %27 group is ranging from 0,00 to 0,47. Independent groups t-test results between item average scores were significant at 0,05 level (p<0,05) in all items. This finding shows that the reliability of the items on the scale was at high levels and that it can distinguish the scorers in terms of the characteristics wanted to be measured.

When the findings are examined, it can be claimed that high level of internal consistency was detected as a result of the studies done in study group of “TPASPF”. In other words, it can be considered that scale presented reliable results in study group.

In order to test reliability in terms of stability of the scale, test-retest method was applied. It was applied in two weeks intervals for 60 parents who are working in public and private schools in districts stated in Istanbul city method department. Pearson Correlation Coefficient and dependent groups t-test results were examined to identify stability factor between two separate applications of the scale and sub-dimensions. Results are presented in Table 5 below.

Table 5. Pearson Correlation Coefficient and Arithmetic Averages, Standard Deviation Values and Dependent Groups t-Test Results of “TPASPF”s Test-retest Scores

		X	N	Ss	r	p	t	p
Pair 1	Beforetotal	3,70	60,00	0,32	0,676	0,00	0,259	0,797
	Aftertotal	3,69	60,00	0,45				
Pair 2	beforef1	3,87	60,00	0,22	0,433	0,001	0,374	0,71
	afterf1	3,85	60,00	0,38				
Pair 3	beforef1	3,45	60,00	0,58	0,745	0,00	-0,144	0,886
	afterf1	3,46	60,00	0,68				
Pair 4	beforef1	3,91	60,00	0,16	0,449	0,00	1,026	0,309
	afterf1	3,87	60,00	0,36				
Pair 5	beforef1	3,66	60,00	0,53	0,367	0,004	-1,692	0,096
	afterf1	3,78	60,00	0,48				

As seen in Table 5, according to findings from “TPESPF” test-retest applications arithmetic average of pre-test was 3,70, arithmetic average of post-test was 3,69; standard deviation of pre-test was 0,32 and standard deviation of post-test was 0,45. Pearson Correlation Coefficient value between two applications was 0,68. This finding showed that relationship

between two applications of the scale was positive and significant at 0,00 level ($p < 0,01$). As a result, findings pointed out that on a whole scale basis consistency between test-retest applications of the scale were high. Similarly, when "TPESPF" first sub-dimension test-retest application was examined, first sub-dimension's pre-test arithmetic average was 3,87 and post-test arithmetic average was 3,85. Pearson Correlation Coefficient value between two applications was 0,43. A positive and significant relationship at 0,01 level ($p < 0,01$) was found. These findings show that consistency between test-retest applications in the first dimension of the scale are on high levels. When findings are examined for the second dimension of the scale, arithmetic average of second sub-dimension pre-test was 3,45 and arithmetic average of post-test was 3,46. Pearson Correlation Coefficient value between two applications was 0,74. Relationship between two applications was positive and significant at 0,001 level ($p < 0,01$). When third sub-dimension of the scale was examined, pre-test arithmetic average was found 3,91 and post-test arithmetic average was found 3,87. Pearson Correlation Coefficient value between two applications was 0,45. Relationship between two applications was positive and significant at 0,00 level ($p < 0,01$). Finally, when findings for fourth sub-dimension were examined, pre-test arithmetic average for fourth sub-dimension was 3,66 and post-test arithmetic average for fourth sub-dimension was 3,78. Pearson Correlation Coefficient value between two applications was 0,37. Relationship between two applications was positive and significant at 0,004 level ($p < 0,01$).

Dependent groups t-Test results of average scores obtained from test-retest done to identify scale's reliability in terms of stability were examined. Dependent groups t-Test results between average scores on a total scale basis was found significant at 0,05 significance level and pointed out that there was not a significant difference. This result showed that external consistency that defines the reliability in terms of stability was found. When similar findings for sub-dimension of the scale were studied, a significant difference was not found after dependent groups t-Test applied for every each sub-dimension. As a result of this analysis, it can be claimed that every sub-dimension of the scale shows external consistency.

3.5 Findings on "TPESCF" Validity

In the validity study of the scale, content and construct validity were checked. Content validity of the assessment tool was identified by detecting the scale items' conformity/validity level by taking expert opinions. An item pool consisting of 38 questions was formed for children to evaluate their teachers. Questions in the pool were broached to 6 experts. Lawshe method was applied to opinions taken from experts' views. After applying Lawshe method, number of scale items were reduced to 23 since 15 items got values under 0,99.

In order to examine construct validity of the scale, exploratory factor analysis was done. Exploratory factor analysis was done with principal components method. As a result of exploratory factor analysis, varimax rotation method was used to obtain interpretable factors which can be named. By this way construct validity of the scale was examined by factor analysis.

In order to determine data's adequacy obtained from 256 children's evaluations on preschools teachers, for factor analysis, Sampling Adequacy Test (Kaiser-Meyer-Olkin-KMO) was applied. To determine adequacy of each item for factor analysis Measures of Sampling Adequacy (MSA) test was used and to specify whether data are coming from multivariate normal distribution or not, Bartlett's Test of Sphericity was applied.

KMO value which defines whether sampling is adequate for factor analysis or not, was above 0,50 and Barlett test was significant at 0,05 significance level so data set was found conformable for factor analysis ($KMO = 0,849$, $\chi^2_{\text{Barlett test}}(253) = 1212,045$; $p = 0,000$). These values showed that data's conformity for factor analysis was on a perfect level and according to KMO values scale's variables was able to predict each other without any errors.

Measure of Sampling Adequacy (MSA) values, which defines each item's conformity for factor analysis, was examined. In cases where Measure of Sampling Adequacy (MSA) values under 0,50 the question is to be removed from the analysis (Sipahi, Yurtkoru and Cinko, 2006). As a result of MSA analysis, 5th and 8th items were found to be under 0,50 and these questions were removed from analysis due to their low explanatory power.

"TPESCF" variance ratios and eigenvalues are presented in Table 6 below.

Table 6. Variance Ratios and Eigenvalues "TPESCF"s sub-dimensions

Factor	Eigenvalue	Variance	Cumulative
1	2,80	14,75	14,75
2	2,54	13,38	28,14
3	2,43	12,77	40,91

In factor analysis in which varimax rotation was used to specify principal components, "TPASCF" was divided into three sub-dimensions as can be seen in Table 15 and total variance amount is % 40,91. When it is considered that variance ratio between %40 and %60 in factor analysis is accepted as ideal (Scherer, 1988; cited Erdogan, Bayram, Deniz, 2007), total variance found in this study is adequate.

Factor weights obtained from factor analysis and variance ratios explained by factors are presented in Table 7 below.

Table 7. Exploratory Factor Analysis Results of Data Obtained from Study Group in Which “TPASCF” Construct Validity was Applied

Item no	Factor no	Factor Weight Values	Item no	Factor no	Factor Weight Values	Item no	Factor no	Factor Weight Values
q16	1	0,61	q7	2	0,74	q10	3	0,72
q22	1	0,58	q14	2	0,70	q23	3	0,64
q17	1	0,57	q1	2	0,53	q12	3	0,54
q2	1	0,55	q19	2	0,50	q9	3	0,49
q11	1	0,55	q21	2	0,48	q4	3	0,46
q20	1	0,53	q3	2	0,46	q6	3	0,46
q13	1	0,46						

When Table 7 was investigated, it can be seen that, 19 items forming the scale was divided into 7 items under first sub-factor, into 6 items under second sub-factor and into 6 items under third sub-factor as a result of exploratory factor analysis. Factor weight value is a coefficient which explains items' relationship with sub-dimensions. In literature, to form factor pattern factor weights ranging from 0,30 to 0,40 can be taken as under cutting point (Kaplan, 1989; cited Ozguven, 1999, p.104; Buyukozturk, 2007). 0,40 was accepted as undercutting point in this study. When first results of factor analysis were examined it was observed that factor weight values of some items were below 0,40 or they had high weight values in both factors. In line with these criteria, factor analysis was applied one more time after four items were deleted from the scale. After the analysis, “TPASCF” was put into its final form which consists of three factors and 19 items. It can be seen that first sub-factor “motivating the child” consists of 7 items, second sub-factor “close relationship with the child” consists of 6 items and third sub-factor “child's trust in the teacher” consists of 6 items after factor rotation.

In “TPASCF” construct validity study, finally correlations between factors with each other and with the total scale was investigated. Findings obtained for study group are presented in Table 8 below.

Table 8. Correlation values of Data Obtained from Study Group in Which “TPASCF” Construct Validity was Applied

		F1	F2	F3	p
F1	r				
F2	r	0,63**			0,00
F3	r	0,51**	0,45**		0,00
Total	r	0,85**	0,82**	0,81**	0,00

**p<0,001

As seen in Table 8, correlations between factors got values ranging from 0,45 to 0,63, correlations between factors and whole scale got values ranging from 0,85 to 0,81. They were positive and significant at 0,01significance level (p<0,01). This finding was interpreted as there was a positive and significant relationship between factors forming the scale prepared for study group and factor-total scale scores.

3.6 Findings on “TPESCF” Reliability

Whole scale's Cronbach Alpha coefficients and if-item-deleted Cronbach Alpha coefficients were calculated to identify internal consistency coefficient of the scores obtained from the scale. In addition to this, item-total correlations were measured in order to identify to what degree scale items distinguish the evaluators. Then, it was tested with independent groups t-Test whether the difference between item average scores of lower %27 and upper %27 groups, which were formed according to total scores obtained from the scale, was significant or not.

As a result of the reliability analysis, internal consistency coefficient of “TPASCF” was found 0,84. At the same time, when if-item-deleted Cronbach Alpha is applied, the fact that items are not higher than Cronbach Alpha coefficient indicates that the reliability coefficient of the factor will not increase if any item is removed. In this case, removing any item will not have a positive effect on reliability of “TPASCF”. For this reason, it was decided that all items should remain on the scale.

Also reliability analyses were applied for every each of the factors. Results of the analyses are; 0,70 for the first factor, 0,71 for the second factor and 0,70 for the third factor.

Independent groups t-Test results between item-based arithmetic average, standard deviation and item average scores of upper %27 and lower %27 groups formed based on the scores obtained from “TPESCF” total scores and its sub-dimension scores in the study group were done, as well. Arithmetic average of upper %27 group was between 3,00 and 2,63 whereas arithmetic average of lower %27 group was ranging between 2,48 and 1,72. Also standard deviation of upper %27 group was between 0,69 and 0,00, standard deviation of lower %27 group was ranging between 0,79 and 0,57. Independent groups t-Test scores between item average scores was significant at a 0,05 significance level (p<0,05) in all items. This finding shows that the reliability of the items on the scale was at high levels and that it can distinguish

the scorers in terms of the characteristics wanted to be measured. When findings are evaluated, it can be said that a high level of internal consistency was found as a result of "TPASCF"s applications in study group. In other words, it can be considered that scale indicated reliable results in the study group.

Test-retest method was applied to identify reliability in terms of stability of the scale. It was applied in two weeks intervals for 35 children who are attending to public and private schools in districts stated in Istanbul city method department. Pearson Correlation Coefficient and dependent groups t-Test results were examined to identify stability coefficient of the scale's two separate applications. Findings are presented in Table 9 below.

Table 9. Dependent Groups t-Test Results, Standard Deviation, Arithmetic Average of Mean Scores Obtained From "TPASCF" Test-retest Applications on Factor and Total Scale Basis

		X	N	ss	r	p	t	p	df
Pair 1	beforetotal1	2,6391	35	,37066	,782	,000	1,453	,155	34
	aftertotal1	2,5774	35	,38854					
Pair 2	beforef11	2,5143	35	,45335	,706	,000	1,189	,243	34
	afterf11	2,4449	35	,44744					
Pair 3	beforef21	2,6238	35	,43788	,700	,000	-,606	,549	34
	afterf21	2,6571	35	,39594					
Pair 4	beforef31	2,7286	35	,33354	,602	,000	1,260	,216	34
	afterf31	2,6524	35	,43979					

As shown in Table 9, arithmetic average of scores after first application on factor basis is between 2,7286 and 2,5143 whereas arithmetic average on total scale basis is 2,6391. Arithmetic average of scores after last application on factor basis is between 2,6571 and 2,4449 whereas arithmetic average on total scale basis is 2,5774. Besides these, standard deviation of obtained scores after first application on factor basis is between 0,45335 and 0,33354 whereas standard deviation of the obtained scores after first application on total scale basis is 0,37066. Standard deviation of obtained scores after last application on factor basis is between 0,44744 and 0,39594 whereas standard deviation of the obtained scores after last application on total scale basis is 0,38854. Dependent groups t-Test results between mean scores on factor and total scale basis was significant at 0,05 significance level ($p > 0,05$), and showed no significant difference between all factors and total scale scores. This finding showed that reliability in terms of stability of the scale was found.

Teachers' total performance scores obtained from the scales which was turned into forms after validity and reliability analyses and performance scores obtained from every sub-group, were weighted according to the feedbacks received from the experts. While weighting scores were being identified, averages of sub-dimension scores defined by the experts were taken.

Scores obtained from administrator-parent-child forms of each sub-group can be weighted and transform into a single performance score.

4. Result

Three different forms to obtain viewpoints from administrators, parents and children to evaluate teachers, were formed for this study which aimed to develop assessment forms to measure teacher performance. For every form Lawshe method was used by receiving opinions from experts. After Lawshe analyses, construct validity of the scales were presented by applying factor analysis for each form. As a result of factor analysis, it was found out that administrator form consisted of only one dimension, parent form consisted of 4 sub-dimensions and child form consisted of 3 sub-dimensions. Each form also was evaluated by item total correlations and item-deleted analysis. And sufficient findings were obtained as presented in the findings section. Relationship between the sub-scales of parent and child forms and a statistically significant relation was detected. When internal consistency of the scales was considered, it was ascertained that administrator form, parent form and child form were found to be reliable at a high level. Considering these forms to specify teacher performance once again experts were consulted and common sub-factors were evaluated together and weighted average calculation method was applied. Obtaining teacher performance score can be achieved by calculating weight averages of administrator, parent and child forms together with sub-factors and then transforming these obtained scores into arithmetic means to get one single score.

As a result of this study, these valid and reliable scales can be used to evaluate teacher performance. And also, each educational institution can identify deficiencies in institutions by using the "Teacher Performance Evaluation Scales" created in this research. On the other hand it must be noted that, this study was only comprised of administrators'-parents' and children's evaluations. It may be suggested that the study can be enriched by adding "self evaluation of teachers".

References

Barnett, S. (2008). *Preschool Education and Its Lasting Effects: Research and Policy Implications*. New Jersey: Great Lakes Center For Education Research and Practice.

- Barutcugil, I. (2002). *Performance Management*. Kariyer Publishing.
- Bingol, D. (2010). *Human Resources Management*. Istanbul: Beta Publishing.
- Boyd, J., Barnett, S. W., Bodrowa, E., Leong, D. J., & Gomby, D. (2005). *Promoting Children's Social and Emotional Development Through Preschool Education*. New Jersey: National Institute for Early Education Research.
- Brown, A. (2005). Implementing Performance Management In England's Primary Schools. *International Journal of Productivity and Performance Management*, 54(5-6), 468-481. <https://doi.org/10.1108/17410400510604593>
- Bulut, Z. (2004). Companies Performance Evaluation Methods and Practices. *Mevzuat Journal*, 7(79).
- Buyukozturk, S. (2007). *Handbook of Data Analysis for Social Sciences*. Ankara: PegemA Publishing.
- Cicek, D. (2005). *Motivation And Quality Of Work Life (Qwl) In Organizations: A Research About By Determining The Motivation Level Of Managers In Public Organization And Developng Their Qwl*. Institute of Social Sciences. Adana: Cukurova Universitesi.
- Erdogan, I. (2008). *Education and School Management*. Istanbul: Alfa Publishing
- Erdogan, Y., Bayram, S., & Deniz, L. (2007). Web based Instruction Attitude Scale: Explanatory and Confirmatory Factor Analyses. *International Human Being Journal*, 4(2), 1-14.
- Field, A. (2009). *Discovering Statistics Using SPSS*. Singapore: Sage Publications.
- Findikci, I. (1999). *Human Resources Management*. Istanbul: Alfa Publishing.
- Government of the NorthWest Territories Municipal and Community Affairs. (2009). Preschool Teacher Training Needs Assesment Evaluation Form. NorthWest Territories, Canada. March 1, 2009
- Oktay, A. (1999). *Magical Years of Life: Preschool Years*. Istanbul: Epsilon Publishing.
- Oktay, A., & Polat, O. (2008). Being a Teacher in the Process of Modern School or Institution, *Educational Psychology Symposium Book*, (s. 1-17). Istanbul: Istanbul Kultur University.
- Ozguven, I. E. (1999). *Psychological Tests, Third Edition*, Ankara: PDREM Publication.
- Pahl, K., & Barrett, P. (2007). *The Development of Social-Emotional Competence in Preschool Aged Children: The Fun FRIENDS Program*. Australia: Pathways Health and Research Centre.
- Sipahi, B., Yurtkoru, S. E., & Cinko, M. (2006). *Data analysis using SPSS in social sciences*. Istanbul: Beta Printed Publishing.
- Tamam, S. (2005). *Teachers and Pricipals Views About Performance Evaluation within the Perspective of Total Quality Management in Primary Schools*. Institute of Social Sciences, Department of Educational Sciences, Adana: Cukurova University.
- Taymaz, H. (1993). *Audit, Concepts, Priciples, Methods*. Kadioglu Press.
- Tok, S., Tok, M., & Mazi, A. (2008). Evaluation of Analysis And Sound-Based Sentences Methods For Pre-Reading And Writing Education, *Educational Administration: Theory and Practice*, (53), 123-144.
- University of California at Santa Barbara. (1997, August). University Of California, Santa Barbara Performance Evaluation Form. Santa Barbara, California, USA. March 1, 2009 www.childscenter.sa.ucsb.edu
- Vroom, V. (1970). The Nature of Relationship Between Motivation and Performance. E. Deci, & V. Vroom in, *Management and Motivation*. Great Britain: Penguin Books.
- Yavuzer, H. (2003). *Knowing and Understanding the Child*. Remzi Publishing
- Yurdugul, H. (2005). Usage of Content Validity Indices in Scale Development Studies for Content Validity. XIV. *National Congress of Educational Sciences*. Denizli: Pamukkale Unversity Faculty of Education
- Zembat, R. (2005). Quality in Pre-School Education). *Contemporary Issues in Pre-School Education* (s. 25-42). Istanbul: Morpa Publishing.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.