

Development of the rubric self-efficacy scale

Perihan Güneş¹, Özen Yıldırım^{2*} , Miraç Yılmaz³

¹Aksaray University, Education Faculty, Nigde, Turkey

²Pamukkale University, Education Faculty, Denizli, Turkey

³Hacettepe University, Education Faculty, Ankara, Turkey

Abstract: The purpose of this study is to develop a valid and reliable measurement tool determining teachers' self-efficacy regarding rubrics. Especially in educational environments, rubrics are measurement tools used in the assessment phase of student products usually based on higher-order thinking skills. Determination of teachers' self-efficacy regarding rubrics can give researchers an idea on how often and how accurately teachers use such tools. For this reason, the existence of a tool accurately measuring self-efficacy variable is necessary. This study's sample consists of 641 elementary, middle and high school teachers. To determine teachers' self-efficacy levels regarding rubrics, 47-item draft was developed. As a result of validity and reliability analyzes, a 28-item measurement tool with a four-factor structure was obtained. The total scale's and sub-factors' internal consistency is quite high. Using this scale, researchers can examine the relationships between teachers' self-efficacy and various variables that play an important role in education. In addition, comparative studies on the intended use of rubrics can be conducted by determining teachers' self-efficacy levels regarding rubrics.

ARTICLE HISTORY

Received: November 01, 2017

Revised: December 12, 2017

Accepted: December 28, 2017

KEYWORDS

Rubric, Teacher Efficacy Scale Development, Psychometric Properties, Performance Tasks

1. INTRODUCTION

The changes in social needs also bring about changes in the qualities people are required to have. In recent years, societies are in need of individuals who can analyze information, think creatively, impart the information they have learned into their daily lives and do research, and who have a developed critical perspective. Many countries have been constantly changing their curriculum to meet this need. The changes made are not only limited to the teaching approaches but also reflect on measurement and evaluation approaches. The question of how to evaluate these higher-order skills needed and the insufficiency of the available tools (oral exams, written exams, tests, etc.) led to complementary measurement and evaluation approaches, which enable these skills to concretize and thus to be measured, to take center stage. Complementary measurement and evaluation approaches provide performance-based assessments of the process in which the product was produced as well the product itself. Rubrics are one of the most common measurement tools used for this purpose.

*Corresponding Author E-mail: ozen19@gmail.com

Researchers define rubrics in different ways. However, according to the most commonly used definition, rubrics are tools that clearly specify the criteria which will be used to evaluate the observed performance, define the behaviors which the individuals have to exhibit in each criterion, and rank these performances from best to worst (or vice versa) (Andrea & Du, 2005; Andrade et al., 2009; Brookhart, 2013; Popham, 1997; Reddy & Andrade, 2010). Rubrics have three basic characteristics: evaluation criteria, criterion definitions, and scoring strategy (Popham, 2007). Evaluation criteria indicate according to which criteria a performance will be evaluated (Wiggins, 1991). Criterion definitions are detailed descriptions reflecting the performance levels of performance criteria scored from best to worst. Scoring strategies provide information on whether the scoring will be on the performance process or the product (Moskal, 2000).

In recent years, attempts to develop characteristics of higher-order thinking skills in schools and easier evaluation of products and process of these characteristics' popularized rubrics. Rubrics contribute significantly to both the teaching and evaluation process by presenting clear and well-defined criteria for the performance that needs to be exhibited. The most important characteristic of rubrics is that they clearly present teachers' learning objectives to the students. In addition, with the clear criteria presented in rubrics, teachers can provide students with detailed feedbacks about the products' weaknesses and strengths (Andrade, 2005). At the same time, detailed feedback mechanism supports the development of students' peer and self-evaluation skills (Panadero et al., 2016). Clear and well-defined criteria in rubrics allow the performance evaluation process to be transparent and consistent (Jonsson, 2014). This has a positive effect on the reliability of performance evaluation. Rubrics with well-defined performance criteria reduce the risk of different interpretation of the exhibited performance by evaluators (Reynolds et al., 2009) and the risk of incorrect scoring due to different interpretations (Venning & Buisman-Pijlman, 2013). In addition to these, rubrics support the development of psychological structures like self-efficacy and self-regulation which positively affect learning (Panadero & Jonsson, 2013).

Today, thanks to performance-based evaluations, teachers can easily evaluate whether students gained higher-order thinking skills or not at the end of their completed complex performance tasks (making presentation, designing model, writing an original story, etc.). For this reason, it is assumed that teachers have sufficient knowledge to use rubrics in educational settings and interpret the results, and they are expected to use these tools appropriately in schools. However, the studies conducted put forth that teachers have difficulties in how to prepare, implement and evaluate performance-based approaches and that they want to be informed on these issues (Metin 2013; Metin & Özmen, 2010). In this context, it is important to determine teachers' self-efficacy beliefs regarding rubrics, which are among the complementary measurement tools. Therefore, within the scope of this study, it was aimed to develop a tool measuring teachers' self-efficacy regarding rubrics.

Bandura(1977, 1994) defined the term 'self-efficacy', that he expressed as one of the most important factors that have an impact on the human behavior, as the self-belief of an individual in her/his competence or ability of successfully accomplishing a task. Bandura (1994) indicates that the beliefs on our abilities are influential on self-efficacy. The possession of a strong or a weak self-efficacy has an impact on the behavior or performance of an individual (Zimmerman, 2000). A strong self-efficacy belief is a behavior that increases the motivation of an individual with regards to overcoming a problem when a problem is confronted and enables an individual to put an effort. On the other hand, a weak self-efficacy belief prevents an individual to perform a task or finalize it (Jerusalem, 2002). A strong self-efficacy emotion is effected from the experience an individual had, other individuals' experiences, the expressions of an individual to perform a task, and from the emotional state

of an individual in the time that the behavior is displayed (Bandura, 1994). Schwarzer (1993) state that self-efficacy might be associated with various particular fields such as education, social, development and health. Moreover, Bandura (1977) remarked that individuals have different levels of self-efficacy in different fields, in other words, self-efficacy might alter according to the field and situation. For instance, an individual may have a high self-efficacy in a particular field, and low-efficacy in another field.

The belief of self-efficacy has been frequently used in the research studies related to learning and teaching (Özkan, Tekkaya & Çakıroğlu, 2002; Riggs & Enochs, 1990; Tschannen–Moran & Woolfolk–Hoy, 2001; Elias and Loomis, 2002). The self-efficacy of teachers, which is one of the most important factors in terms of learning and teaching also plays an important role. Teacher self-efficacy is the belief that teachers have about their abilities towards difficult or low-motivated students to participate in class and learn (Bandura, 1977). In the literature, there are several studies on teacher self-efficacy (Caprara, Barbaranelli, Steca, & Malone, 2006; Tschannen-Moran, Woolfolk-Hoy & Hoy, 1998; Tschannen–Moran & Woolfolk–Hoy, 2001; Yılmaz et.al 2004;). In this context, examination of the existed beliefs of teachers on applying subsidiary assessment and evaluation instruments. This situation might inform about how often and how correctly teachers use these instruments in in-class applications.

2. METHOD

This study is a scale development study using the basic survey model.

2.1. Study Group

This study was carried out during the 2016-2017 academic year. The scale development phase of the study was conducted with 641 elementary, middle and high school teacher who were knowledgeable about rubrics.

During the first phase of the study, the data obtained from 216 teachers were used in principal factor analysis and the data obtained from the remaining 425 teachers were used in confirmatory factor analysis. 327 (51%) of the participants were female and 314 (49%) were male. When the school levels were taken into consideration, the number of participant elementary school teachers (73.5%) were higher than the number of participant middle and high school teachers (26.5%). In order to increase the study impact, data from 16 different cities from Turkey's seven regions were collected. Convenience sampling method was used to reach the sample.

2.2. Data Collection

The validity and reliability works of the Rubric Self-Efficacy Scale was obtained at the end of the pilot study conducted on the selected sample.

2.2.1 Rubric self-efficacy scale

The self-efficacy scale regarding rubrics was developed similar to the scaling approach based on grading totals developed by Likert (1932). During the scale development, first, literature on self-efficacy was reviewed. As a result of the review, literatures on rubrics and self-efficacy were reached. When the literature was examined, it was seen that there was not a measurement tool determining “*teachers’ rubric self-efficacy*” in Turkish or in another language. Therefore, no direct resource was used while developing the items. In addition the literature review, ten elementary and high school teachers were asked to explain their views on the preparation, implementation and evaluation of rubrics in the classroom and their positive or negative experiences with rubrics. Based on the qualitative data obtained, 47 items on teachers’ preparation, implementation and evaluation of rubrics were developed.

During the scale's pilot study development phase, the items were examined by two measurement and evaluation experts and two Turkish language experts. According to the views taken from them, researchers removed 12 items from pilot study of scale form due to the fact that they did not reflect what they intended for and that they had ambiguities. The other items were organized according to the expert opinions. In pilot application, there were 20 positive statements putting forth teachers' high self-efficacy level regarding rubrics and 15 negative statements emphasizing teachers' low self-efficacy level. Teachers express how much they agree or disagree with the statements by choosing responses of Strongly Agree (5), Agree (4), Neither Agree or Disagree (3), Disagree (2) and Strongly Disagree (1).

2.3 Data Analysis

In the scale development phase, first, principal component analysis technique was used to put forth the state of the data structure and to reduce factor, and later confirmatory factor analysis was used to test the structure. Additionally to prove validity, item-total test correlation and a correlation coefficient from the upper and lower 27% of the total group was tested. Also for reliability testing Cronbach Alfa level of each factors was found.

Before principal component factor analysis, the suitability of the data structure for analysis was examined. Multivariate and univariate extreme values were identified, and 12 people were left out of the analysis because of the unexpected data structure. KMO (Kaiser-Meyer-Olkin) value determines how suited the data structure is for factor analysis based on the sampling adequacy, and Bartlett's Test of Sphericity informs about the state of multivariate normal distribution of the data. Table 1 presents the statistics regarding the KMO and Bartlett's Test of Sphericity.

Table 1. KMO and Bartlett's Test of Sphericity

KMO		.79
Barlet Test	Ki-square	2351.76
	df	59
	p	.00

When Table 1 is examined, the KMO value was found to be 0.79. According to this value, the sample size is at an adequate level to continue factor analysis. Whether the data set met the multivariate assumption or not was checked with Bartlett's Test of Sphericity. The value obtained show that data set met the assumption of multivariate normality ($\chi^2= 2351.76$; $p<0.01$). In addition to these assumptions, multicollinearity problem between the variables was examined with Pearson Product-Moment Correlation, and it was found that there was no multicollinearity. During the principal component analysis, factors with Eigen values greater than 1 were taken into consideration, and items with a factor load of at least 0.32 (Tabachnick & Fidel, 2001) were accepted and selected for the real scale. Cronbach's Alpha value, which determines the internal consistency, that is, how closely correlated the items are with each other and the test, was examined for the internal consistency according to the total scale and sub-dimensions. For item discrimination index, the groups of the upper and lower 27% were compared, and item total test correlations were examined for validity testing.

After verify the scale's structure, confirmatory factor analysis method was used. This phase includes the testing process of the measurement model. By this means, whether the factorized structure is verified as a model or not with the principal component analysis was examined. Before starting the confirmatory factor analysis, the data structure of 425 people different than the principal component analysis was examined, and extreme and missing values were checked. Eight people were excluded from the analysis because of their unexpected data structure in terms of univariate and multivariate extreme values. When the missing data was

examined, it was determined that the missing data structure was 1.25%, and the researchers decided to assign missing data based on the mean. Since the data did not meet the assumption of normal distribution during the confirmatory factor analysis, the data were normalized, and the analysis continued. The confirmatory component analysis allowed that each observed variable showed relationship with only the latent variable under it.

3. FINDINGS

This section of the study includes findings regarding the principal component analysis, item-total test correlation, upper and lower %27 total group analysis, reliability analysis and confirmatory factor analysis of rubric development.

3.1. Principal component analysis of Rubric Self-Efficacy Scale

According to the principal component analysis method done to determine the scale's factor structures, nine factors with Eigen values higher than 1.00 were obtained. These nine factors reflect 65.17% of the total variance. Findings based on Eigen values and the variances they explain are given in Table 2.

Table 2. Eigen values and the explained variances

Factor	Eigen value	Variance %	Total Variance %
1	6.78	19.37	19.37
2	6.03	17.23	36.60
3	2.22	6.35	42.95
4	1.95	5.56	48.51
5	1.35	3.86	52.38
6	1.22	3.49	55.87
7	1.17	3.34	59.22
8	1.06	3.04	62.26
9	1.02	2.91	65.168

The first four factors explain the 48.51% of the total variance. After these four factors, the contribution of other factors on the percentage of the total variance decreases. It is seen that the four-factor structure adequately explain the studied variable. This is presented in the scree plot (Figure 1) showing eigenvalue components.

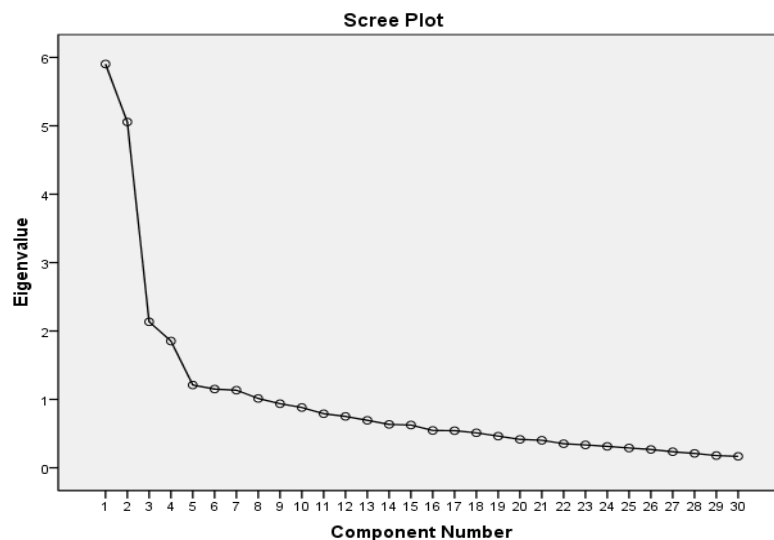


Figure 1. Scree plot

At the scree plot the slope with high acceleration where rapid decreases occur points out to a considerable amount of factor numbers. When Figure 1 is examined, it is seen that after four factors there is a routinized variation. After evaluating Table 2, Figure 1 and dimensions taken into consideration during item writing together, it was decided that the number of factors should be four. However, when the factor load values were examined before rotation, it was determined that factor load values of all items were greater than 0.32, and the smallest value was 0.486 and the greatest value was 0.76.

The step taken into account in the factorization process is the determination of the rotation method used. Varimax has been preferred as a rotation method since it was not expected that there would be a high degree of correlation among the factors that emerge in the principal component analysis. According to findings obtained from the rotation, 11 items under factor 1, nine items under factor 2, ten items under factor 3 and five items under factor 4 were determined. When the distribution of the items according to the factor load values is examined, the lowest load value is 0.43 and the highest load value is 0.79. When the items' cross loading is examined, it is seen that five items are collected under more than one factor and have a high loading value in each factor. The difference between factor load values is less than 0.10. Starting from the first item with the closest load value, the items were removed from the scale.

Table 3. Items' factor load values obtained as a result of factor analysis

Item	Factor 1 Loadings	Factor 2 Loadings	Factor 3 Loadings	Factor 4 Loadings
M19	0.75	0.19		-0.11
M8	0.72	0.20		
M9	0.67		0.27	
M20	0.65	0.22		
M13	0.62	0.19		0.11
M7	0.60	-0.12		
M10	0.59	0.22		
M27	0.59	0.28	-0.13	
M2	0.42	0.21	0.14	0.21
M39		0.78		
M41	0.19	0.77		
M33	0.14	0.70	-0.11	-0.22
M40	0.25	0.68	0.16	
M44	0.11	0.65		
M43	0.39	0.60	0.13	
M29	0.34	0.56	-0.11	
M32	0.18	0.41		-0.28
M12		0.16	0.77	
M15		-0.15	0.70	
M11	0.22		0.69	0.11
M6			0.66	0.18
M26			0.64	0.25
M18		-0.21	0.57	0.22
M24	0.14		0.55	0.26
M37			0.47	0.32
M31			0.14	0.82
M30	0.10		0.12	0.81
M28			0.20	0.79
M42			0.17	0.68
M35			0.22	0.66

The analyses were repeated in this order. Yet, since the cross loadings remained, these five items were not included in the scale. 30 items collected under four factors explain 49.05% of the total variance. The factor load values of the items that were collected under four factors as a result of factor analysis and were decided to be kept in the scale are shown in Table 3 below.

3.2 Item discrimination and examination of the test’s reliability

In order to determine items’ discrimination levels, item-total test correlation coefficients and item discrimination values for groups of the upper and lower 27% were examined. The findings are given in Table 4.

Table 4. Item analysis results

Item	Item-total test correlation N=204	Upper and lower 27% Nupper=Nlower=56
M2	0.478	-7.21***
M6	0.31	-3.68***
M7	0.32	-4.89***
M8	0.49	-7.99***
M9	0.45	-8.20***
M10	0.47	-7.19***
M11	0.32	-5.03***
M12	0.39	-5.92***
M13	0.32	-4.92***
M15	0.10	-1.55
M18	0.19	-2.42*
M19	0.43	-6.19***
M20	0.34	-5.43***
M24	0.32	-5.60***
M26	0.34	-5.20***
M27	0.41	-6.59***
M28	0.56	-6.27***
M29	0.57	-7.18***
M30	0.53	-5.72***
M31	0.55	-6.36***
M32	0.38	-3.78***
M33	0.52	-5.79***
M35	0.33	-3.09***
M37	0.40	-4.57***
M39	0.57	-7.31***
M40	0.64	-8.91***
M41	0.61	-8.13***
M42	0.47	-4.20***
M43	0.63	-8.46***
M44	0.47	-5.36***

***p<0.001 *p<0.05

When item-total test correlations explaining the relationship between the scores from the items and the total score of the scale are examined in determining the item discrimination levels, it is seen that the correlation of item 15 and item 18 with the total has the lowest correlation scores, 0.102 and 0.197 respectively. Item-total test correlation values of the other items range between 0.65 and 0.31. On the other hand, when the difference between item scores’ means among groups of the upper and lower 27% were examined, it was found that item 15 was not discriminative and for item 18 the mean difference between the lower and

upper groups is at a significant level of 0.05, but it was close to each other. Based on the two different discrimination findings, item 15 and 18 were excluded from the scale. When the reliability of the rest 28-item test is examined, the Cronbach's alpha value was determined as 0.85. This indicates that the test measures with high reliability. Reliability factor has also been tested for each factor. While the Cronbach's alpha value of the first factor was 0.80, the Cronbach's alpha value of the second factor was 0.89. The Cronbach's alpha value of the third factor was 0.70 whereas the Cronbach's alpha value of the fourth factor was 0.83. These results showed that sub-dimensions had high reliability (internal consistency).

As a result of the principal components analysis, 28-item scale is grouped under four factors. There are nine items under first factor, eight under the second, six under the third and five under the fourth factor. Factors were named as followed: Factor1 Efficacy of monitoring student development, Factor2 Efficacy of monitoring teaching, Factor3 Efficacy to overcome learning environment difficulties, Factor3 Efficacy of rubric preparation

3.3. Confirmatory factor analysis of the rubric self-efficacy scale

In the third phase of the study, structural validity of the items that were reduced to four factors through principal component analysis was tested with the confirmatory factor analysis. Table 5 was developed based on the results of the measurement model.

In Table 5, standardized loadings provide information about the correlation between the each observed variable and the latent variable that it is related to. While M8 (0.72) shows the highest correlation with Factor 1, M7 (0.47) shows the lowest correlation. Variability in Factor 1 is explained the most by the M8 ($R^2=0.52$) variable. M41 shows the highest correlation with Factor 2 whereas M32 shows the lowest correlation with Factor 2. For this reason, the variable with the highest R^2 coefficient is the M41(0.48) variable. When Factor 3 is examined, it is determined that M11 (0.65) shows the highest correlation with Factor 3 whereas M37 (0.38) shows the lowest. Most of the variability ($R^2=0.42$) in Factor 3 is explained by M11. When Factor 4 is examined, it is seen that M30 has the highest (0.79) correlation coefficient and M11 has the lowest (0.52). In this factor, most of the variance is explained by the M30 ($R^2=0.62$) variable. When the error variances of the observed variables in the measurement model are examined, it found that error variances changed between 0.37 and 0,87. However, observed t values are calculated as significant for all variables ($p<0.00$). According to the findings obtained from the first examination of the measurement model, nothing disturbs the model's fit.

As a second examination, goodness of fit indices obtained from the measurement model were checked. Table 6 provides information on goodness of fit indices. When the significance of value X^2 that reveals the difference between the observed and expected matrices is examined, it was found that this value was significant ($p<0.00$). This can be due to the size sample size in the study. For this reason, the examination of goodness of fit indices continued. In terms of model goodness of fit indices, the indices other than GFI and AGFI have a good fit value whereas GFI (0.88) value and AGFI (0.86) value show a weak fit.

Table 5. Results of the measurement model

Factor/Item	Standardized Loadings	t-value	R ²
Factor1			
M2	0.52	10.61	0.27
M7	0.47	9.54	0.22
M8	0.72	16.04	0.52
M9	0.65	14.04	0.42
M10	0.57	11.97	0.32
M13	0.53	10.93	0.28
M19	0.70	15.47	0.49
M20	0.69	15.24	0.48
M27	0.58	12.25	0.34
Factor2			
M29	0.58	12.15	0.34
M32	0.36	7.07	0.13
M33	0.56	11.45	0.31
M39	0.70	15.38	0.49
M40	0.67	14.34	0.45
M41	0.69	14.89	0.48
M43	0.63	13.46	0.40
M44	0.61	13.93	0.37
Factor3			
M6	0.58	12.86	0.34
M11	0.65	12.98	0.42
M12	0.58	11.51	0.34
M24	0.47	8.92	0.22
M26	0.57	11.25	0.32
M37	0.38	7.08	0.14
Factor4			
M28	0.71	15.47	0.50
M30	0.79	17.89	0.62
M31	0.75	16.73	0.56
M35	0.55	11.34	0.30
M42	0.52	10.49	0.27

Table 6. Goodness of fit indices for measurement model

Fit indicates	Criteria	Value	Goodness
X ² /sd	≤ 3	811.04 / 344 =	Good
RMSEA	0.05 ≤ RMSEA ≤ 0.08	0.057	Good
SRMR	0.05 ≤ SRMR ≤ 0.10	0.060	Good
NFI	≤ 0.90	0.91	Good
NNFI	≤ 0.90	0.94	Good
CFI	≤ 0.90	0.94	Good
GFI	≤ 0.90	0.88	Weak
AGFI	≤ 0.90	0.86	Weak
CN	≥ 200.00	216.29	Convenient sample

In addition to the goodness of fit, modifications suggested by the analysis outputs were checked. In modifications, it is suggested that M19 in Factor 1 is mapped to Factor 3 and factor 4, and M37 to Factor 2 and Factor 4. In the order of adjustments, M37 that lowered the X^2 value the most and then the exclusion of the item M19 from the model were examined. According to the findings, while there was a slight improvement only in the X^2/sd value, there was no improvement in the other goodness of fit indices AGFI and GFI value did not reach the desired fit level. For this reason, researchers decided that both of the items would remain in the model. No modifications were made during the confirmatory factor analysis.

4. DISCUSSION AND CONCLUSION

Many countries have been basing their educational programs on the constructivist approach, and they consider the use of process-based complementary measurement and evaluation tools like portfolios, projects, performance tasks and concept maps in addition to the use of traditional product-based measurement and evaluation tools like paper-and-pencil tests including open-ended, short and multiple choice questions as significant. In order for these complementary measurement and evaluation tools to be used effectively and efficiently, teachers must have knowledge and full competence in this area. For the determination of knowledge and self-efficacies, measurement tools that measure these qualities are also needed. By this means, more objective results can be reached. Within the scope of this study, a scale determining teachers' self-efficacy beliefs regarding rubrics was developed.

First, content validity of the scale developed by taking into the validity and reliability analysis was reached by taking the opinions of experts. Content validity is one of the leading validity analyses giving information on whether a scale measures based on an intended characteristic or not (Cronbach, 1990). Principal component analysis was done to determine the structure of the scale, and it was determined that the self-efficacy structure was grouped under four factors and these factors clearly explain nearly 50% of the variable. In multivariate designs, explained variance ratio is expected to be 40.00% and 60% (Scherer, Wiebe, Luther & Adam, 1998 cited in Tavşancıl, 2005). The obtained findings support the aforementioned resources. Moreover, when the slopes of the scree plot were examined, the four-factor structure can clearly be seen. Giving information about the internal consistency coefficient of the scale, Cronbach's alpha values show high reliability for both the scale itself and its sub-dimensions. This indicates that the items on the scale have a high correlation with each other. Cronbach (1970) stated that the scale would have high internal consistency if the alpha level on the scale is greater than 0.70.

According to the discrimination of the groups of the upper and lower 27% done to determine the item discrimination index levels, it was found that only two items did not differentiate between the positive and negative attitudes. This was determined by low correlation coefficient number, and they were not included in the scale. It is recommended that if the value of discrimination is lower than 0.19, the items should be revised, and if they cannot be adjusted, they should not be included in the scale (Kelley, 1939). The validity of the items was also examined by item-total test correlation. A low correlation suggests that the item should be removed from the scale (Cureton, 1966; Guilford, 1953). According to this finding, the same two items had the low correlation and they were removed from the scale.

Finally, the scale's structure validity was examined with confirmatory factor analysis, and it was determined that the structure put forth in the principal components analysis was reached again. When the analysis results were examined, it was seen that the chi square value was significant. Since the chi square value is affected by the sample (Byrne, 2003), goodness of fit indices was checked. The indices other than GFI and AGFI have a good fit value whereas GFI value and AGFI value show a weak fit. When the literature is examined, an AGFI value

greater than 0,85 can be considered as a good fit (Raykov & Marcoulides, 2006; Schermelleh-Engel Moosbrugger & Müller, 2003; Vieira, 2011).

The developed scale can be used for the purposes of related researches and institutions. Particularly, researchers working on complementary measurement and evaluation methods can examine the relationships between teachers' self-efficacy regarding rubrics and student performance according to different variables.

Acknowledgement

This work was supported by Aksaray University Scientific Research Projects Coordination Unit. Project Number: 2015-096

5. REFERENCES

- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30
- Andrade, H. G., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3). Retrieved from <http://PAREonline.net/getvn.asp?v=10&n=3>
- Andrade, H., Wang, X., Du, Y., & Akawi, R. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287-302.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84 (2), 191-215.
- Bandura, A. (1994). *Self-efficacy*. In V.S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior*, (4), 71-81. New York: Academic Press.
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.
- Byrne, B. M. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: a study at the school level. *Journal of School Psychology*, 44, 473-490.
- Cronbach, L. J. (1970). *Essentials of psychological testing*. 3rd Edition, Harper & Row.
- Cureton, E. E. (1966). Corrected item-test correlations. *Psychometrika*, 31, 93-96.
- Elias, S., & Loomis, R. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *Journal of Applied Social Psychology*, 32(8), 1687- 1702.
- Guilford, J. P. (1953). The correlation of an item with a composite of the remaining items in a test. *Educational and Psychological Measurement*, 13, 87-93
- Jerusalem, M. (2002). Theroretischer Teil - Einleitung I, *Zeitschrift für Pädagogik*, 44, 8-12
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39 (7), 840-852. doi:10.1080/02602938.2013.875117
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17-24.
- Likert, R. (1932). A Techniques for the measurement of attitudes. *Archives of Psychology*, 140, 5-53
- Metin, M. (2013). Öğretmenlerin performans görevlerini hazırlarken ve uygularken karşılaştığı sorunlar. *Kuram ve Uygulamada Eğitim Bilimleri*, 13(3), 1645- 1673.

- Metin, M., & Özmen, H. (2010). Fen ve teknoloji öğretmenlerinin performans değerlendirmeye yönelik hizmet içi eğitim (HİE) ihtiyaçlarının belirlenmesi. *Kastamonu Eğitim Dergisi*, 18(3), 819-838.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7(3),1-5
- Özkan, Ö., Tekkaya, C., & Çakıroğlu, J. (2002). Fen bilgisi aday öğretmenlerin fen kavramlarının anlama düzeyleri, fen öğretimine yönelik tutum ve öz-yeterlik inançları. V. Ulusal Fen Bilimleri Eğitimi Kongresi, ODTÜ, Ankara.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. doi:10.1016/j.edurev.2013.01.002.
- Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). *Scaffolding self-regulated learning through selfassessment and peer assessment: Guidelines for classroom implementation*. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). Cham: Springer International Publishing.
- Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, 55(2), 72-75.
- Popham, W. J. (2007). *Classroom Assessment: What Teachers Need to Know*. Pearson Education, 5th Edition, USA.
- Raykov T & Marcoulides G. A. (2006). *Fundamentals of structural equation modeling. A first course in structural equation modeling*. 2nd ed. London: Lawrence Erlbaum Associates; p.1-3, 41-3
- Reddy, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35, 435- 448. doi:10.1080/02602930902862859.
- Reynolds, J., Smith, R., Moskovitz, C., & Sayle, A. (2009). BioTAP: A Systematic Approach to Teaching Scientific Writing and Evaluating Undergraduate Theses. *BioScience*, 59 (10), 896–903. doi:10.1025/bio.2009.59.10.11
- Riggs, I. M. ve Enochs L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625-637.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Schwarzer R. (1993). General perceived self-efficacy in 14 cultures. Retrieved on 5-June 2007, at URL <http://Web.Fu-Berlin.De/Gesund/Publicat/Ehpscd/Health/World14.Htm>
- Tabachnick, B. G., & Fidel, L. S. (2001). *Using multivariate statistics*. Fourth Edition. Needham Heights, MA: Allyn & Bacon.
- Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Nobel Yayınları, Ankara
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68, 202-248.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education*, 17, 783-805.
- Venning, J., & F. Buisman-Pijlman (2013). Integrating assessment matrices in feedback loops to promote research skill development in postgraduate research projects. *Assessment and Evaluation in Higher Education*, 38 (5): 567–579.
- Vieira, A. L. (2011). *Preparation of the analysis. Interactive LISREL in practice*. 1st ed. London: Springer.

- Wiggins, G. (1991). Standart, not standardization: Evoking quality student work. *Educational Leadership*, 48(5), 18-25.
- Yılmaz, M., Köseoğlu, P., Gerçek, C., Soran, H. (2004). Yabancı dilde hazırlanan bir öğretmen öz-yeterlik ölçeğinin Türkçeye uyarlanması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 27, 260-267.
- Zimmerman, B.J. (2000). Self-Efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82–91. doi:10.1006/ceps.1999.1016

Table Appendix 1. Rubric Self-Efficacy Scale

Rubric Self-Efficacy Scale		Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
1	I believe, I have sufficient information on how to use rubrics in the classroom.					
2	I may have difficulties while preparing rubrics despite my experience in teaching.					
3	I believe use rubrics even if students have negative attitudes towards them.					
4	I can easily evaluate student performances with rubrics.					
5	I believe I have sufficient information on how to prepare rubrics.					
6	I can do applications using rubrics even if the students have not used them before.					
7	I may have difficulties while preparing rubrics even if I have theoretical knowledge.					
8	I may have difficulties while doing rubric applications because it takes a lot of time.					
9	If I encounter a problem while preparing a rubric, I can overcome it.					
10	I can increase student achievement by preparing effective rubrics.					
11	I can use rubrics effectively in group work.					
12	I may have difficulty in explaining the purpose of using rubrics to the students.					
13	I may have difficulty in adapting a rubric I have found from other sources to the subject.					
14	I can increase student interest towards the subject by rubric applications.					
15	I may have difficulty in scoring according to different rubric types.					
16	With rubrics, I can easily determine students' shortcomings during the learning process.					
17	I may have difficulty in determining the rubric type appropriate to the subject matter.					
18	While preparing the rubric, I may have difficulty in determining the learning objectives.					
19	Even if I have challenges in classroom management, I can do rubric applications.					
20	I can fairly evaluate the written exams with a rubric.					
21	While preparing the rubric, I may have difficulty in deciding on the behaviors to be measured.					
22	I may have difficulty in doing rubric applications in crowded classrooms.					
23	I believe I can improve myself in preparing rubrics by using different sources.					
24	I believe the students can easily understand the rubrics I prepare.					
25	With rubric applications, I believe I can lessen students' anxieties about learning the subject.					
26	Even if I try, while preparing rubrics, I may have difficulty in coming up with detailed definitions measuring student behaviors.					
27	I can prepare rubrics that would make students come up with quality works.					
28	If I try, I can get students gain the ability to use rubrics.					

Dereceli Puanlama Anahtarına Yönelik Öz-Yeterlilik Ölçeğinin Geliştirilmesi

Perihan Güneş¹ Özen Yıldırım^{*2}  Miraç Yılmaz³

¹Aksaray Üniversitesi, Eğitim Fakültesi, Türkiye

²Pamukkale Üniversitesi, Eğitim Fakültesi, Denizli, Türkiye

³Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara, Türkiye

Öz: Bu araştırmanın amacı, öğretmenlerin Dereceli Puanlama Anahtarları (DPA)'na yönelik öz yeterliklerini belirleyen geçerli ve güvenilir ölçme aracı geliştirmektir. Dereceli puanlama anahtarları özellikle eğitim ortamlarında öğrencilerin üst düzey zihinsel becerilerine dayalı ortaya koydukları ürünlerin değerlendirilmesi aşamasında kullanılan ölçme araçlarıdır. Öğretmenlerin DPA'ya yönelik öz yeterliklerinin belirlenmesi, onların bu tür araçları ne sıklıkla ve doğru olarak kullandıkları konusunda araştırmacılara bir fikir verebilir. Bu nedenle öz yeterlik değişkenini doğru olarak ölçen bir aracın varlığı gereklidir. Temel araştırma modeline dayalı olan bu araştırmanın örneklemini, ilköğretimde ve ortaöğretimde çalışan 641 öğretmen oluşturmaktadır. Öğretmenlerin dereceli puanlama anahtarına yönelik öz yeterlilik düzeylerini ortaya koymak için 47 maddeden oluşan taslak form hazırlanmıştır. Ölçeğin geçerliğini belirlemek için kapsam geçerliği, temel bileşenler analizi, madde toplam korelasyonu, alt-üst %27'lik gruplar için madde analizi ve doğrulayıcı faktör analizi yapılmıştır. Ölçeğin güvenilirliği ise iç tutarlılık olarak incelenmiş ve Cronbach alfa değeri ile test edilmiştir. Analizler sonucunda 28 maddeden oluşan ve dört faktörlü bir yapıya sahip ölçek elde edilmiştir. Ölçeğin tamamının ve alt faktörlerinin iç tutarlılık katsayısı oldukça yüksektir. Bu ölçeğe dayalı olarak araştırmacılar öğretmenlerin öz yeterliği ile eğitimde önemli rol oynayan farklı değişkenler arasındaki ilişkiler incelenebilirler. Ayrıca öğretmenlerin DPA'ya yönelik öz yeterlilik düzeyleri belirlenerek DPA'ların kullanım amaçlarına dayalı karşılaştırmalı çalışmalar yapılabilir.

MAKALE SÜRECİ

Gönderim: 01 Kasım 2017

Düzeltilme: 12 Aralık 2017

Kabul: 28 Aralık 2017

ANAHTAR KELİMELER

Rubrik, Öğretmen

Özyeterliliği, Ölçek

Geliştirme,Psikometrik

Özellikler, Performans

Görevleri

1. GİRİŞ

Toplumsal ihtiyaçlardaki değişmeler ihtiyaç duyulan insan niteliklerinin de değişmesini beraberinde getirmektedir. Son yıllarda toplumlar bilgiyi analiz edebilen, yaratıcı düşünebilen, öğrendiği bilgileri günlük hayata aktarabilen, eleştirel bakış açısı gelişmiş, araştırma yapabilen bireylere daha çok ihtiyaç duymaktadırlar. Birçok ülke bu ihtiyacı karşılamak için öğretim programlarını sürekli değiştirmektedir. Yapılan değişiklikler sadece öğretim yaklaşımları ile

*Sorumlu Yazar E-mail: ozen19@gmail.com

sınırlı kalmayıp, ölçme değerlendirme yaklaşımlarına da yansımaktadır. İhtiyaç duyulan bu üst düzey becerilerin nasıl değerlendirileceği ile mevcut araçların (sözlü sınavlar, yazılı sınavlar, testler vb.) yetersiz kalması, bu becerilerin somut biçimde ifade edilmesini ve dolayısıyla ölçülmesini sağlayacak tamamlayıcı ölçme ve değerlendirme yaklaşımlarının ön plana çıkmasını sağlamıştır. Tamamlayıcı ölçme değerlendirme yaklaşımları, ürünün yanı sıra ürünün ortaya çıkma süreci hakkındaki öğretmenlere bilgi sunmaktadır. Örneğin öğrencinin kaynaklara nasıl ulaştığı, hangi kaynaklardan yararlandığı, bu kaynakları nasıl kullandığı hakkında ayrıntılı bilgilere ulaşılabilir. Dereceli puanlama anahtarı (DPA) bu amaçla kullanılan en yaygın ölçme araçlarından biridir.

Araştırmacılar tarafından DPA'lar çok farklı şekillerde tanımlanmakla birlikte en yaygın kullanılan tanım, gözlenen performansın hangi ölçütlere göre değerlendirileceğini açıkça belirten, bireyin her bir ölçütte göstereceği davranışları tanımlayan ve bu performansları belirli değere göre sıralayan araçlardır (Andrea ve Du, 2005; Andrade ve diğerleri, 2009; Brookhart, 2013; Popham, 1997; Reddy ve Andrade, 2010).

DPA'lar değerlendirme ölçütleri, ölçüt tanımları ve puanlama stratejisi olmak üzere üç temel öğeye sahiptir (Popham, 2007). Değerlendirme ölçütleri, sergilenmesi gereken performansın hangi ölçütlere göre değerlendirileceğini göstermektedir (Wiggins, 1991). Ölçüt tanımları, performansta ele alınan her bir ölçüte dayalı performans düzeyleri de dikkate alınarak yazılmış detaylı tanımlardır. Puanlama stratejileri, puanlamanın sürece mi yoksa sonuca mı dönük yapılacağını hakkında bilgi vermektedir (Moskal, 2000).

Son yıllarda özellikle eğitim ortamlarında üst düzey zihinsel özelliklerin geliştirilmek istenmesi ve bu özelliklerinin ürünlerinin ve sürecinin DPA'lar sayesinde daha kolay değerlendirilebilmesi, bu ölçme araçlarını popüler hale getirmiştir. DPA'lar öğrenciden beklene performansla ilgili açık ve iyi tanımlanmış ölçütler sunarak hem öğretme hem de değerlendirme sürecine önemli katkılar sağlamaktadır. DPA'ların en önemli özelliği öğretmenlerin öğrenme ile ilgili hedeflerini öğrencilere açık bir şekilde sunmasıdır. Bunun yanı sıra DPA'larda sunulan açık ölçütler ile öğretmenler, öğrencilere çalışmalarının zayıf ve güçlü noktaları hakkında detaylı geri bildirimler verebilmektedir (Andrade, 2005). Aynı zamanda detaylı geribildirim mekanizması öğrencilerin akran ve öz değerlendirme becerilerinin gelişmesini desteklemektedir (Panadero ve diğerleri, 2016). DPA'larda ölçütlerin açık ve iyi tanımlanmış olması performans değerlendirme sürecinin şeffaf ve tutarlı olmasına olanak sağlamaktadır (Jonsson, 2014). Bu durum performans değerlendirmenin güvenilirliği üzerinde pozitif etkiye sahiptir. İyi tanımlanmış performans ölçütlerine sahip DPA'lar sergilenen performansın değerlendiriciler tarafından farklı yorumlanma (Reynolds ve diğerleri, 2009) ve yorumlanma farklılığından kaynaklanan hatalı puanlama riskini düşürmektedir (Venning ve Buisman-Pijlman, 2013). Tüm bunların yanı sıra DPA'lar öğrenmeyi olumlu etkileyen öz yeterlilik ve öz düzenleme gibi psikolojik yapıların gelişmesini de desteklemektedir (Panadero ve Jonsson, 2013).

Günümüzde performansa dayalı değerlendirmeler sayesinde öğretmenler, öğrencilerin karmaşık performans görevleri (sunum yapma, model tasarlama, özgün bir hikâye yazma vb.) yerine getirerek üst düzey zihinsel becerileri kazanıp kazanmadığını rahatlıkla ölçebilir. Bu nedenle eğitim ortamlarında öğretmenlerin DPA'ları kullanma ve sonuçları yorumlama konusunda yeterli bilgiye sahip oldukları varsayılmakta ve bu araçları okullarda uygun bir şekilde kullanmaları beklenmektedir. Oysaki yapılan çalışmalar öğretmenlerin performansa dayalı yaklaşımların nasıl hazırlanacağı, uygulanacağı ve değerlendirileceği konusunda zorlandıklarını ve bu konularda bilgilenmek istediklerini ortaya koymaktadır (Metin ve Özmen, 2010; Metin 2013). Bu tür bir ölçme aracını kullanırken, bilgi eksiklikleri olan veya zorlanan öğretmenlerin, sınıf içi uygulamalarında bu araçtan yararlanma sıklıklarının da az olması beklendik bir durumdur. Bununla birlikte DPA'lar sadece öğretmenin öğrenci

performansı hakkında bilgi edinmesini değil, öğrencinin de görevine dayalı süreç ve üründe neleri yapıp yapamadığı hakkında bilgi edinmesini sağlar. Öğrenciler eksikliklerini görerek ilerideki görevlerinde bu eksikliklerinin üstesinden gelmeye çalışır. Bu bağlamda ilk olarak öğretmenlerin tamamlayıcı ölçme araçları içerisinde yer alan DPA'ya yönelik öz yeterlik inançlarının ortaya çıkarılması önemlidir. Ancak öğretmenlerin DPA'ya yönelik öz yeterliklerini ölçen geçerli ve güvenilir bir araç literatürde bulunmamaktadır. Bu nedenle araştırma kapsamında öğretmenlerin DPA'ya yönelik öz yeterliklerini ölçen bir araç geliştirmek amaçlanmıştır.

Bandura (1977, 1994) insan davranışlarını etkileyen önemli faktörlerden biri olarak ifade ettiği öz yeterliği, bireyin bir işi başarılı olarak yapıp yapmayacağı konusunda kendisine duyduğu inanç şeklinde tanımlamıştır. Bandura (1994), yeteneklerimize ilişkin var olan inançların özyeterlik üzerinde etkili olduğunu belirtmiştir. Bireyin güçlü ya da zayıf öz yeterlik inancına sahip olması, bireyin davranışı ya da performansı üzerinde etkilidir (Zimmerman, 2000). Güçlü öz yeterlik inancı, bireyin herhangi bir problemle karşılaştığında o problemle başa çıkacağına dair motivasyonunu artıran ve çaba göstermesini sağlayan bir davranıştır. Zayıf özyeterlik inancı ise bireyin bir işi yapması ya da sonuçlandırması için çaba göstermesine engel olmaktadır (Jerusalem, 2002). Güçlü bir öz yeterlik duygusu bireyin kendi yaşadığı deneyimlerden, başkalarının yaşadığı deneyimlerden, bireyin bir işi yapacağına dair motivasyon ifadeleri ile bireyin davranışı sergileyeceği andaki duygusal durumundan etkilenmektedir (Bandura, 1994). Schwarzer (1993) öz yeterliğin eğitim, sosyal, gelişim ve sağlık gibi pek çok özel alanla ilişkili olabileceğini bildirmektedir. Bandura ise (1977), bireylerin farklı alanlarda farklı öz yeterlilik düzeylerine sahip olabileceğini, yani öz yeterliğin alana ve duruma göre değişebildiğini dile getirmiştir. Örneğin bir birey bir alanda yüksek öz yeterliğe, başka bir alanda ise düşük öz yeterliğe sahip olabilir.

Öz yeterlik inancı, öğrenme-öğretme ile ilgili araştırmalarda sıklıkla kullanılmaktadır (Özkan, Tekkaya ve Çakıroğlu, 2002; Riggs ve Enochs, 1990; Tschannen–Moran ve Woolfolk–Hoy, 2001; Elias ve Loomis, 2002). Öğretme ve öğrenme üzerinde en etkili faktörlerden biri olan öğretmenlerin, çeşitli alanlardaki öz yeterlikleri de oldukça önemlidir. Öğretmen öz yeterliği, bir öğretmenin zor ya da motive olabilecek öğrencilerin bile derse katılımını ve öğrenmesini sağlamaya yönelik yeteneklerine olan inancıdır (Bandura, 1977). Literatürde öğretmen öz yeterliği ilgi birçok çalışma yapılmıştır (Caprara, Barbaranelli, Steca ve Malone, 2006; Tschannen-Moran, Woolfolk-Hoy ve Hoy, 1998; Tschannen–Moran ve Woolfolk–Hoy, 2001; Yılmaz, Köseoğlu, Gerçek ve Soran 2004). Bu bağlamda öğretmenlerde tamamlayıcı ölçme değerlendirme araçlarını kullanabileceğine dair var olan inançların incelenmesi önem taşımaktadır. Bu durum öğretmenlerin sınıf içi uygulamalarında ne sıklıkla ve ne kadar doğru olarak bu araçları kullandıkları yönünde ön bilgi verebilir. Öğretmenlerin olumsuz öz yeterlik inançlarının belirlenmesi durumunda, DPA'nın kullanılmasını özendirerek ek bilgilendirme seminerleri düzenlenebilir.

2. YÖNTEM

Bu araştırma, temel tarama modelinde planlanmış bir ölçek geliştirme çalışmasıdır.

2.1. Çalışma Grubu

Bu çalışma, 2016-2017 eğitim öğretim yılında gerçekleşmiştir. Çalışmaya dereceli puanlama anahtarı hakkında bilgisi olan 641 ilköğretim ve ortaöğretim öğretmeni ile gerçekleştirilmiştir. Öğretmenlerin DPA'lar hakkındaki bilgisi kendi algılarına dayalı olarak yoklanmıştır. Araştırmanın birinci adımı kapsamında 216 öğretmenden elde edilen veriler temel bileşenler analizinde geriye kalan 425 öğretmenden elde edilen veriler ise doğrulayıcı faktör analizinde kullanılmıştır. Katılımcıların 327 (%51)'si kadın, 314 (%49)'ü ise

erkeklerden oluşmaktadır. Okul düzeylerine göre ilköğretim düzeyinde örneklem katılan öğretmen sayısı (%73,5), ortaöğretim düzeyindeki öğretmenlere (%26,5) göre daha fazladır. Çalışmanın yaygın etkisini arttırmak için Türkiye'nin yedi bölgesinin 16 farklı illindeki devlet okullarında görev yapan çeşitli branşlarda öğretmenlerden veriler toplanmıştır. Çalışma grubuna ulaşmada uygun örneklem yöntemi kullanılmıştır.

2.2 Verilerin Toplanması

Dereceli puanlama anahtarına ilişkin öz yeterlik ölçeğinin geçerlik ve güvenilirlik çalışmaları belirlenen örneklem üzerindeki pilot uygulama sonucunda elde edilmiştir. *Dereceli puanlama anahtarına ilişkin öz yeterlik ölçeği*: DPA'ya yönelik öz yeterlik ölçeği, Likert (1932) tarafından geliştirilen derecelendirme toplamlarına dayalı ölçekleme yaklaşımının adımlarına benzer olarak geliştirilmiştir.

Ölçek geliştirilirken öncelikle, öz yeterliğe dayalı ilgili alan yazın taraması yapılmıştır. Tarama sonucunda DPA ve öz yeterliğe dayalı ilgili kaynaklara ulaşılmıştır. Literatür incelendiğinde "*DPA'ya yönelik öğretmenlerin öz yeterliklerini*" ortaya koyan Türkçe veya yabancı dilde bir ölçme aracına ulaşılmadığından maddelerin oluşturulmasında destekleyici doğrudan bir kaynaktan yararlanılmamıştır. Literatür taramasının yanı sıra ölçek maddelerinin oluşturulması amacıyla 10 kişilik ilköğretim ve 2 lise düzeyinde öğretmene DPA'ları sınıf içerisinde uygulama, hazırlama ve kullanmalarına yönelik görüşleri ve bununla birlikte DPA'ya yönelik varsa olumlu veya olumsuz deneyimlerini açıklamaları istenmiştir. Elde edilen nitel verilerden yararlanarak öğretmenlerin DPA'ları hazırlaması, kullanımı ve uygulamasına ilişkin öz yeterliklerini ortaya koyabilecekleri 47 madde oluşturulmuştur.

Ölçeğin pilot uygulamaya hazırlanması sürecinde maddelerin incelenmesini iki ölçme ve değerlendirme ve iki Türk dili uzmanı gerçekleştirmiş, alınan görüşlere göre 12 madde ölçülmek istenilen durumu yansıtmaması ve anlatım bozukluklarının olması nedeniyle araştırmacılar tarafından ön uygulama formundan çıkarılmıştır. Diğer maddeler uzmanların görüşlerine göre düzenlenmiştir. Ön deneme uygulamasına hazır hale getirilen ölçekte öğretmenin DPA'ya yönelik öz yeterliğini ortaya koyan 20 olumlu cümle, öz yeterliğindeki düşüklüğü vurgulayan 15 olumsuz cümle yer almaktadır. Öğretmenler ifadelerin her birini katılıp katılmama durumuna göre '5' Tamamen Katılıyorum, '4' Katılıyorum, '3' Kararsızım, '2' Katılmıyorum ve '1' Hiç Katılmıyorum biçiminde derecelendirilmiş seçeneklerden seçmektedir.

2.3. Verilerin Analizi

Ölçek geliştirme aşamasında ilk olarak veri yapısının durumunu ortaya koymak ve faktör indirgemek amacıyla temel bileşenler analizi tekniğinden yararlanılmış, daha sonra yapıyı test etmek için doğrulayıcı faktör analizi uygulanmıştır.

Temel bileşenler analizinden önce veri yapısının analiz için uygunluğu incelenmiştir. Çok değişkenli ve tek değişkenli uç değerler belirlenmiş, buna göre 12 kişi beklenmedik veri yapısına sahip olması nedeniyle analiz dışında bırakılmıştır. Test edilen diğer varsayımlar örneklem büyüklüğüne dayalı olarak veri yapısının faktör analizine uygunluğunu ortaya koyan KMO (Kaiser-Meyer-Olkin) değeri ve verilen çok değişkenli normal dağılım gösterme durumunu hakkında ipucu veren Barlett Testi'dir. Tablo 1'de KMO ve Barlett Küresellik Testine ilişkin istatistikler verilmektedir.

Tablo 1. KMO ve Barlett Testi

KMO		.79
Barlett Testi	Ki-kare	2351.76
	sd	595
	p	.00

Tablo.1 incelendiğinde KMO değerinin 0.79 olduğu gözlenmiştir. Bu değere göre örneklem büyüklüğü faktör analizine devam etmek için iyi düzeydedir. Veri setinin çok değişkenli normallik varsayımını karşılayıp karşılamadığı ipucu ise Barlett Küresellik Testi ile kontrol edilmiştir. Elde edilen değer, veri setinin çok değişkenli normallik varsayımını karşıladığını göstermektedir ($\chi^2= 2351.76$; $p<0.01$). Bu varsayımların yanı sıra değişkenler arasındaki çoklu bağlantı problemi de Pearson Momentler Çarpımı korelasyonu ile incelenmiş ve çoklu bağlantının olmadığı gözlenmiştir.

Temel bileşenler analizi sırasında faktör öz değeri 1'den büyük olan faktörler dikkate alınmış ve faktör yükleri en az 0.32 (Tabachnick ve Fidel, 2001) olan maddeler kabul edilerek asıl ölçek için seçilmiştir. Maddelerin birbirleri ve testle olan korelasyonunu veren iç tutarlılık anlamındaki Cronbach Alfa değeri testin tamamı ve alt boyutlarına göre güvenilirlik için incelenmiştir. Maddelerin ayırt ediciliği için alt-üst %27'lik gruplar karşılaştırılmış ve madde toplam test korelasyonlarına bakılmıştır.

Ölçeğin yapısını doğrulamak için doğrulayıcı faktör analizi (DFA) yöntemi uygulanmıştır. Bu aşama ölçme modelinin test edilmesi sürecini içermektedir. Bu sayede temel bileşenler analizi ile faktörleştirilmiş yapının bir model olarak doğrulanıp doğrulanmadığı incelenmiştir. DFA'ya geçilmeden önce temel bileşenler analizinden farklı 425 kişilik veri yapısı incelenmiş, uç değerler ve kayıp değerlere bakılmıştır. Tek yönlü ve çok yönlü uç değerler bakımından 8 kişi beklenmedik veri yapısına sahip olması nedeniyle analiz dışında tutulmuştur. Verilerde kayıp veri durumu incelendiğinde kayıp veri yapısının %1.25'i kadar olduğu belirlenmiş ve araştırmacılar tarafından ortalamaya dayalı kayıp verinin atanması kararlaştırılmıştır. DFA sırasında verilerin normal dağılım varsayımını karşılamamasından dolayı veriler normalleştirilerek analize devam edilmiştir.

DFA sırasında Temel Bileşenler Analizinin sonuçları dikkate alınarak, her bir gözlenen değişkenin yalnızca kendi altında yer alan bir gizil değişkenle ilişki göstermesine izin verilmiştir.

3. BULGULAR

Çalışmanın bu bölümünde dereceli puanlama anahtarı geliştirmeye ilişkin temel bileşenler analizi, güvenilirlik analizleri ve faktör analizine ilişkin bulgulara ver verilmiştir.

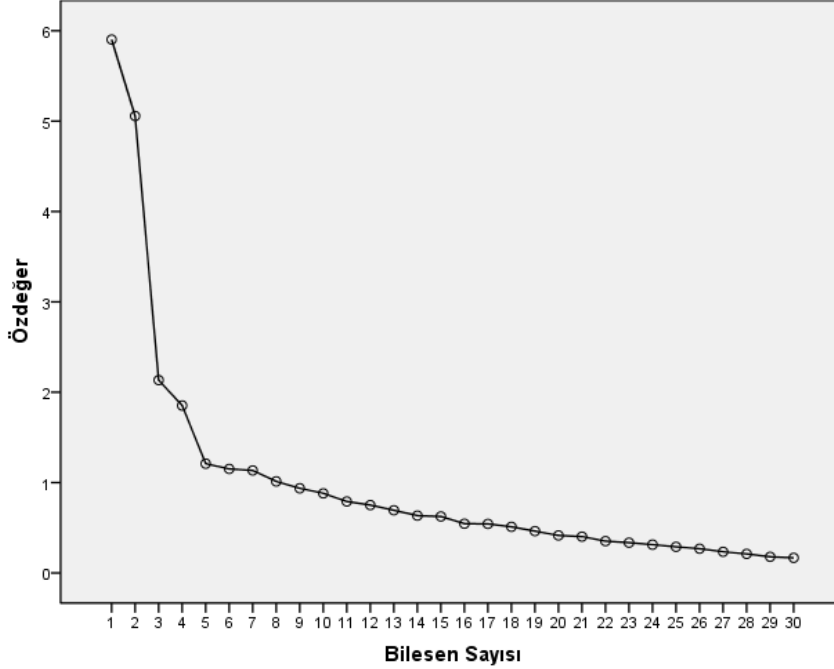
3.1. DPA öz yeterlik ölçeği temel bileşenler analizi

Ölçeğin faktör yapılarını belirlemek amacıyla yapılan temel bileşenler analiz yöntemine göre öz değeri 1.00'da yüksek dokuz faktör elde edilmiştir. Bu dokuz faktör toplam varyansın %65.168'ini yansıtmaktadır. Öz değer ve açıkladıkları varyanslara dayalı bulgular Tablo 2.'de verilmiştir.

Tablo 2. Öz değerler ve açıkladıkları varyanslar

Faktör	Öz değer	Varyans Yüzdesi	Toplam Varyans Yüzdesi
1	6.78	19.37	19.37
2	6.03	17.23	36.60
3	2.22	6.35	42.95
4	1.95	5.56	48.51
5	1.35	3.86	52.38
6	1.22	3.49	55.87
7	1.17	3.34	59.22
8	1.06	3.04	62.26
9	1.02	2.91	65.17

İlk dört faktör toplam varyansın %48.51'ini açıklamaktadır. Bu dört faktörden sonra diğer faktörlerin toplam varyans yüzdesine yaptığı katkı azalmaktadır. Dört faktörlü yapının araştırılan değişken için yeterli varyansı açıkladığı görülmektedir. Bu durum öz değer bileşenlerini gösteren çizgi grafiğinde de (Şekil 1) görülmektedir.



Şekil 1. Öz değer bileşen çizgi grafiği

Çizgi grafiğine göre yüksek ivmeli, hızlı düşüşlerin yaşandığı eğim önemli sayıda faktör sayısına işaret etmektedir. Şekil 1 incelendiğinde dört faktörden sonra eğim sabitlenmekte tek düze bir farklılaşma gözlenmektedir. Tablo 2, Şekil 1 ve madde yazımında dikkate alınan boyutlar bir arada değerlendirilerek faktör sayısının dört olmasına karar verilmiştir.

Bununla birlikte döndürmeden önce faktör yük değerleri incelendiğinde bütün maddelerin faktör yük değerlerinin 0.32'den büyük olduğu en küçük değer 0.486 ve en büyük değer 0.762 olduğu belirlenmiştir.

Faktörleşme sürecinde dikkate alınan adım kullanılan döndürme yönteminin belirlenmesidir. Temel bileşenler analizinde ortaya çıkan faktörler arasında yüksek derecede ilişki olması beklenmediğinden döndürme yöntemi olarak varimax tercih edilmiştir. Döndürme sonucunda elde edilen bulgulara göre 1. faktör altında 11 madde, 2. faktör altında 9 madde ve 3. faktör altında 10 madde ve 4. Faktör altında 5 madde belirlenmiştir. Maddelerin faktör yük değerlerine göre dağılımı incelendiğinde en düşük yük değeri 0.43 en yüksek yük değeri 0.79'tir. Bununla birlikte maddelerin binişiklik durumu incelendiğinde 5 maddenin birden fazla faktör altında toplandığı ve her faktörde de yüksek yük değerine sahip olduğu gözlenmiştir. Faktör yük değerleri arasındaki fark 0.10'dan küçüktür. En yakın yük değerine sahip ilk maddeden başlayarak sırasıyla maddeler ölçekte çıkarılmıştır. Analizler bu sırada tekrarlanmıştır. Ancak maddelerin binişikliklerinin kalkmamasından dolayı belirlenen bu beş madde ölçeye dahil edilmemiştir.

Dört faktör altında toplanan 30 madde toplam varyansın %49.05'ini açıklamaktadır. Faktör analizi sonucunda dört faktör altında toplanan ve ölçekte kalmasına karar verilen maddelere ait faktör yük değerleri aşağıdaki Tablo 3'te gösterilmektedir.

Tablo 3. Faktör analizi sonucunda maddelere ilişkin elde edilen faktör yük değerleri

Madde No	Faktör 1 Yük Değeri	Faktör 2 Yük Değeri	Faktör 3 Yük Değeri	Faktör 4 Yük Değeri
M19	0.75	0.19		-0.11
M8	0.72	0.20		
M9	0.67		0.27	
M20	0.65	0.22		
M13	0.62	0.19		0.11
M7	0.60	-0.12		
M10	0.59	0.22		
M27	0.59	0.28	-0.13	
M2	0.42	0.21	0.14	0.21
M39		0.78		
M41	0.19	0.77		
M33	0.14	0.70	-0.11	-0.22
M40	0.25	0.68	0.16	
M44	0.11	0.65		
M43	0.39	0.60	0.13	
M29	0.34	0.56	-0.11	
M32	0.18	0.41		-0.28
M12		0.16	0.77	
M15		-0.15	0.70	
M11	0.22		0.69	0.11
M6			0.66	0.18
M26			0.64	0.25
M18		-0.21	0.57	0.22
M24	0.14		0.55	0.26
M37			0.47	0.32
M31			0.14	0.82
M30	0.10		0.12	0.81
M28			0.20	0.79
M42			0.17	0.68
M35			0.22	0.66

3.2. Madde ayırt ediciliği ve testin güvenirliğinin incelenmesi

Maddelerin ayırt edicilik düzeylerinin belirlenmesi amacıyla madde toplam test korelasyonu katsayıları ve %27'lik alt-üst gruplar için madde ayırt ediciliği değerleri incelenmiştir. Elde edilen bulgular Tablo 4'te verilmiştir.

Madde ayırt edicilik düzeylerinin belirlenmesinde maddelerden alınan puanlar ile ölçeğin toplam puanı arasındaki ilişkiyi açıklayan madde toplam test korelasyonları (madde ayırt edicilik değerleri) incelendiğinde; 15. ve 18. maddenin toplam ile korelasyonu sırasıyla 0.10 ve 0.19 olarak en düşük korelasyon değerine sahip olduğu gözlenmiştir. Diğer maddelerin madde toplam test korelasyon değerleri 0.65 ve 0.31 arasında değişmektedir. Bununla birlikte alt üst yüzde 27'lik gruplar arası madde puanları ortalamaları arası farka bakıldığında 15. maddenin ayırt edici olmadığı, 18. madde de ise alt ve üst gruplar arası ortalama farkın 0.05 düzeyinde anlamlı olmasına rağmen değer birbirine yakın olduğu gözlenmiştir. İki farklı ayırt edicilik bulgularına dayalı olarak 15. ve 18. maddeler ölçeğin dışında tutulmuştur.

Tablo 4. Madde analizi sonuçları

Madde No	Madde Toplam Test Korelasyonu n=204	T (Alt%27-Üst %27) nalt=nüst=56
M2	0.48	-7.20***
M6	0.31	-3.68***
M7	0.32	-4.88***
M8	0.50	-7.99***
M9	0.46	-8.20***
M10	0.47	-7.19***
M11	0.32	-5.03***
M12	0.39	-5.92***
M13	0.32	-4.91***
M15	0.10	-1.55
M18	0.19	-2.41*
M19	0.42	-6.19***
M20	0.34	-5.433***
M24	0.32	-5.60***
M26	0.33	-5.20***
M27	0.41	-6.59***
M28	0.55	-6.26***
M29	0.57	-7.17***
M30	0.53	-5.72***
M31	0.54	-6.36***
M32	0.38	-3.78***
M33	0.51	-5.78***
M35	0.33	-3.08***
M37	0.40	-4.57***
M39	0.57	-7.31***
M40	0.64	-8.91***
M41	0.61	-8.12***
M42	0.47	-4.20***
M43	0.62	-8.46***
M44	0.47	-5.35***

***p<0.001 *p<0.05

28 maddelik testin güvenilirliği incelendiğinde Cronbach alfa değeri 0.85 olarak belirlenmiştir. Bu durum testin yüksek güvenilirlikte ölçme yaptığını göstermektedir. Güvenirlik kat sayısı ayrıca her faktör için test edilmiştir. 1. Faktörün Cronbach Alfa değeri 0.80, 2. Faktörün Cronbach Alfa değeri 0.89 ve 3. Faktörün Cronbach Alfa değeri 0.70 ve 4. Faktörün Cronbach Alfa değeri 0.83 olarak belirlenmiştir. Cronbach Alfa değerleri incelendiğinde alt boyutlarının yüksek güvenilirliğe (iç tutarlılık) sahip olduğu gözlenmiştir. Üçüncü faktörün alfa değeri diğer faktörlere göre daha düşüktür. Bu faktörün alfa değeri kabul edilebilir düzeydedir.

Temel bileşenler analizi sonucunda 28 maddeden oluşan ölçek dört faktörden altında toplanmaktadır. 1. Faktör altında 9 madde, 2. Faktör altında 8 madde, 3. Faktör altında 6 madde ve 4. Faktör altında 5 madde yer almaktadır. Birinci faktör “*Öğrenci gelişimini izleme yeterliği*”, ikinci faktör *Öğretimi yönetme yeterliği*”, üçüncü faktör “*Zorlukların üstesinden gelme yeterliği*”, dördüncü faktör ise “*DPA oluşturma yeterliği*” şeklinde adlandırılmıştır.

3.3. DPA öz yeterlik ölçeği doğrulayıcı faktör analizi (DFA)

Araştırmanın üçüncü adımında temel bileşenler analizi yardımı ile dört boyuta indirgediğimiz maddelerin yapı geçerliği DFA analiziyle test edilmiştir. Ölçüm modeli sonuçlarına dayalı olarak aşağıdaki Tablo 5 oluşturulmuştur.

Tablo 5. Ölçüm modeli sonuçları

Faktör/Madde	Standartlaştırılmış Yükler	t-değeri	R ²
<i>Faktör1</i>			
M2	0.52	10.61	0.27
M7	0.47	9.54	0.22
M8	0.72	16.04	0.52
M9	0.65	14.04	0.42
M10	0.57	11.97	0.32
M13	0.53	10.93	0.28
M19	0.70	15.47	0.49
M20	0.69	15.24	0.48
M27	0.58	12.25	0.34
<i>Faktör2</i>			
M29	0.58	12.15	0.34
M32	0.36	7.07	0.13
M33	0.56	11.45	0.31
M39	0.70	15.38	0.49
M40	0.67	14.34	0.45
M41	0.69	14.89	0.48
M43	0.63	13.46	0.40
M44	0.61	13.93	0.37
<i>Faktör3</i>			
M6	0.58	12.86	0.34
M11	0.65	12.98	0.42
M12	0.58	11.51	0.34
M24	0.47	8.92	0.22
M26	0.57	11.25	0.32
M37	0.38	7.08	0.14
<i>Faktör4</i>			
M28	0.71	15.47	0.50
M30	0.79	17.89	0.62
M31	0.75	16.73	0.56
M35	0.55	11.34	0.30
M42	0.52	10.49	0.27

Tablo 5'te standartlaştırılmış yükler her bir gözlenen değişken ile ilgili olduğu gizil değişken arasındaki korelasyonlar hakkında bilgi vermektedir. Faktör1 ile en yüksek korelasyonu M8 (0.72), en düşük korelasyonu M7 (0.47) göstermektedir. Faktör1'de değişkenliğin en çok M8 (R²=0.52) değişkeni tarafından açıklandığı görülmektedir. Faktör2 ile en yüksek korelasyon gösteren M41 ve en düşük korelasyon gösteren M32 değişkenidir. Bu nedenle R² katsayısı en yüksek olan M41 (0.48) değişkenidir. Faktör 3 incelendiğinde M11 değişkeninin Faktör 3 ile 0.65 düzeyinde en yüksek korelasyonu gösterdiği. M37'in ise en düşük korelasyon katsayısına (0.38) sahip olduğu belirlenmiştir. Faktör 3'teki değişkenliğin çoğu (R²=0.42) M11 tarafından açıklanmaktadır. Faktör 4 incelendiğinde M30 en çok (0.79) ve M42 (0.52) en az korelasyon katsayısına sahiptir. Bu faktörde en fazla değişim M30 (R²=0.62) değişkeni tarafından açıklanmaktadır. Ölçüm modelinde gözlenen değişkenlerin hata varyansları incelendiğinde hata varyanslarının 0.37 ile 0.87 arasında değiştiği gözlenmiştir.

Bununla birlikte gözlenen t değerlerinin bütün değişkenler için manidar ($p < 0.00$) olduğu hesaplanmıştır. Ölçüm modelinin ilk incelemesinden elde edilen bulgulara göre model uyumunu bozan bir durum yoktur. Ölçüm modelinden elde edilen model uyum indeksleri Tablo 6'da verilmiştir.

Tablo 6. Ölçüm modeli için uyum indeksleri

Uyum ölçüsü	Ölçüt	Değeri	Uyum
X^2/sd	≤ 3	811.04 / 344 = 2.36	İyi
RMSEA	$0.05 \leq RMSEA \leq 0.08$	0.057	İyi
SRMR	$0.05 \leq SRMR \leq 0.10$	0.06	İyi
NFI	≤ 0.90	0.91	İyi
NNFI	≤ 0.90	0.94	İyi
CFI	≤ 0.90	0.94	İyi
GFI	≤ 0.90	0.88	Zayıf
AGFI	≤ 0.90	0.86	Zayıf
CN	≥ 200.00	216.29	Yeterli örneklem

Gözlenen ve beklenen matrisler arasındaki farkı ortaya koyan X^2 değerinin anlamlılığı incelendiğinde bu değer anlamlı ($p < 0.00$) olduğu gözlenmiştir. Bu durum araştırmada ele alınan örneklem büyüklüğünden kaynaklı olabilir. Ancak modelde X^2/sd değerinin istenilen düzeyde olduğu görülmektedir. Bu nedenle model uyum indekslerinin incelenmesine devam edilmiştir. Model uyum indeksleri bakımından GFI ve AGFI dışındaki diğer indeksler iyi uyum değerine sahipken GFI değeri (0.88) ve AGFI değeri (0.86) zayıf bir uyum göstermektedir.

Model uyumlarına ek olarak analiz çıktılarının önerdiği modifikasyonlar da incelenmiştir. Modifikasyonlarda Faktör1'de yer alan M19'un Faktör 3 ve Faktör 4 ile eşleştirilmesi ve M37'in Faktör2 ve Faktör4 ile eşleştirilmesi önerilmektedir. Sırasıyla yapılan düzenlemelerde ilk olarak X^2 değerini en fazla düşüren M37 ve daha sonra M19 maddesinin modelden çıkarılma durumu incelenmiştir. Elde edilen bulgulara göre sadece X^2/sd değerinde belli bir iyileşme gözlenirken diğer uyum iyiliği indekslerinde istenilen iyileşme gözlenmemiş, GFI değeri istenilen uyum düzeyine ulaşmamıştır. Bu nedenle araştırmacılar tarafından iki maddenin de modelde kalmasına karar verilmiştir. DFA sırasında herhangi bir modifikasyon yapılmamıştır. Temel bileşenler analizi ile test edilen dört faktörlü yapı doğrulanmıştır.

4. TARTIŞMA VE SONUÇ

Birçok ülke eğitim-öğretim programlarında yapılandırmacı yaklaşımı temel almakta, kağıt kalem testleri içerisinde yer alan açık uçlu, kısa yanıt ve çoktan seçmeli sorular gibi ürün temelli geleneksel ölçme-değerlendirme araçlarının yanında portfolyo, proje uygulamaları, performans görevi, kavram haritası gibi süreç temelli (durum belirleme odaklı) tamamlayıcı ölçme-değerlendirme araçlarının kullanımını önemli görmektedir. Bu tamamlayıcı ölçme-değerlendirme araçlarının etkili ve verimli bir şekilde kullanılabilmesi için öğretmenlerin bu konudaki bilgilerinin ve yeterliliklerinin tam olması gerekmektedir. Bilgi ve yeterliklerin belirlenmesi için de bu özellikleri ölçen geçerlik ve güvenilirlik kanıtları ortaya konmuş ölçme araçlarına ihtiyaç vardır. Bu sayede daha objektif sonuçlara ulaşılabilir. Bu araştırma kapsamında öğretmenlerin DPA'ya yönelik öz yeterlikler inançlarını ortaya koyan bir ölçek geliştirilmiştir.

Geçerlik ve güvenilirlik analizleri dikkate alınarak geliştirilen araçta ilk olarak kapsam geçerliği uzman görüşlerine dayalı olarak sağlanmıştır. Bir ölçeğin istenilen özelliğe dayalı ölçüm yapıp yapmadığı hakkında bilgi veren geçerlik analizlerinin başında kapsam geçerliği gelmektedir (Cronbach. 1970). Ölçeğin yapısını ortaya koymak amacıyla temel bileşenler

analizi yapılmış ve öz yeterlik yapısının dört faktör altında toplandığı, bu faktörlerin değişkenin yaklaşık yüzde ellisini açıkladığı belirlenmiştir. Çok faktörlü desenlerde açıklanan varyans oranının %40 ve %60 olması beklenir (Scherer, Wiebe, Luther ve Adam, 1998; Akt: Tavşancıl, 2005). Elde edilen bulgu belirtilen kaynakları destekler niteliktedir. Ayrıca çizgi grafiğindeki eğimler incelendiğinde dört faktörlü yapı açıkça görülmektedir.

Geliştirilen ölçeğin iç tutarlık kat sayısı hakkında bilgi veren Cronbach alfa değerleri hem ölçeğin kendisi hem de alt boyutlar için yüksek güvenilirlik düzeyine sahiptir. Bu durum ölçekte yer alan maddelerin birbirleri ile yüksek ilişki gösterdiğini belirtmektedir. Cronbach (1970) ölçekte yer alan alfa düzeyinin 0.70 üstünde bulunması durumunda ölçeğin yüksek iç tutarlığa sahip olduğunu belirtmektedir.

Maddelerin ayırt edicilik düzeylerinin belirlenmesi amacıyla yapılan alt-üst %27'lik gruplara ayırma yöntemine göre de sadece iki maddenin tutumları olumlu olan ve olumsuz olanları ayıramadığı, düşük korelasyon kat sayısı ile belirlenmiş ve ölçeğe alınmamıştır. Ayırt edicilik değerinin 0.19'dan düşük olması durumunda maddeler gözden geçirilmeli ve eğer düzenlenemiyor ise ölçeğe alınmaması önerilmektedir (Kelley, 1939). Maddelerin geçerliği madde toplam test korelasyonu ile de incelenmiştir. Düşük korelasyon değeri maddenin ölçekten çıkarılmasına işaret etmektedir (Cureton, 1966; Guilford, 1953). Bu bulguya göre de aynı iki madde ölçekten çıkarılmıştır.

Ölçeğin yapı geçerliği incelenmiş ve temel bileşenler analizinde ortaya konan yapının tekrar sağlandığı gözlenmiştir. Analiz sonuçları incelendiğinde ki kare değerinin anlamlı çıktığı gözlenmiştir. Ki kare değeri örneklemden etkilenilen bir istatistik olduğundan (Byrne, 2003) model uyum indekslerine bakılmıştır. GFI ve AGFI dışındaki diğer indeksler iyi uyum değerine sahipken GFI değeri ve AGFI değeri zayıf bir uyum göstermektedir. Literatür incelendiğinde AGFI değerinin 0.85'in üstünde olması durumunda "kabul edilebilir uyum" olarak değerlendirilebilir (Raykov ve Marcoulides, 2006; Schermelleh-Engel, Moosbrugger ve Müller, 2003; Vieira, 2011).

DPA'ya yönelik öz yeterlik ölçeği öğretmenlerin DPA'lar hakkındaki öz yeterlik algılarını ortaya koymaktadır. Ölçekten alınan artan puanlar öz yeterlik algılarının yüksekliği, azalan puanlar ise düşüklüğü şeklinde yorumlanabilir. Geliştirilen ölçek, ilgili araştırmalar ve kurumlar tarafından öğretmenlerin öz yeterlik düzeylerini belirlemek ve buna dayalı eğitim ve öğretim etkinliklerine yön vermek amacıyla kullanılabilir. Özellikle tamamlayıcı ölçme ve değerlendirme yöntemleri üzerinde çalışan araştırmacılar öğretmenin DPA'ya yönelik öz yeterliği ve öğrenci performansı arasındaki ilişkileri farklı değişkenleri de dikkate olarak çok boyutlu modellerle inceleyebilir.

Teşekkür

Bu çalışma Aksaray Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından desteklenmiştir. Proje Numarası: 2015-096

5. KAYNAKLAR

- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30
- Andrade, H. G., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3). Retrieved from <http://PAREonline.net/getvn.asp?v=10&n=3>
- Andrade, H., Wang, X., Du, Y., & Akawi, R. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287-302.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84 (2), 191-215.

- Bandura, A. (1994). *Self-efficacy*. In V.S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior*, (4), 71-81. New York: Academic Press.
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.
- Byrne, B. M. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: a study at the school level. *Journal of School Psychology*, 44, 473-490.
- Cronbach, L. J. (1970). *Essentials of psychological testing*. 3rd Edition, Harper & Row.
- Cureton, E. E. (1966). Corrected item-test correlations. *Psychometrika*, 31, 93-96.
- Elias, S., & Loomis, R. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *Journal of Applied Social Psychology*, 32(8), 1687-1702.
- Guilford, J. P. (1953). The correlation of an item with a composite of the remaining items in a test. *Educational and Psychological Measurement*, 13, 87-93
- Jerusalem, M. (2002). Theroretischer Teil - Einleitung I, *Zeitschrift für Pädagogik*, 44, 8-12
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39 (7), 840-852. doi:10.1080/02602938.2013.875117
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17-24.
- Likert, R. (1932). A Techniques for the measurement of attitudes. *Archives of Psychology*, 140, 5-53
- Metin, M. (2013). Öğretmenlerin performans görevlerini hazırlarken ve uygularken karşılaştığı sorunlar. *Kuram ve Uygulamada Eğitim Bilimleri*, 13(3), 1645-1673.
- Metin, M., & Özmen, H. (2010). Fen ve teknoloji öğretmenlerinin performans değerlendirmeye yönelik hizmet içi eğitim (HİE) ihtiyaçlarının belirlenmesi. *Kastamonu Eğitim Dergisi*, 18(3), 819-838.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7(3),1-5
- Özkan, Ö., Tekkaya, C., & Çakıroğlu, J. (2002). Fen bilgisi aday öğretmenlerin fen kavramlarının anlama düzeyleri, fen öğretimine yönelik tutum ve öz-yeterlik inançları. V. Ulusal Fen Bilimleri Eğitimi Kongresi, ODTÜ, Ankara.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. doi:10.1016/j.edurev.2013.01.002.
- Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). *Scaffolding self-regulated learning through selfassessment and peer assessment: Guidelines for classroom implementation*. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311-326). Cham: Springer International Publishing.
- Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, 55(2), 72-75.
- Popham, W. J. (2007). *Classroom Assessment: What Teachers Need to Know*. Pearson Education, 5th Edition, USA.
- Raykov T & Marcoulides G. A. (2006). *Fundamentals of structural equation modeling. A first course in structural equation modeling*. 2nd ed. London: Lawrence Erlbaum Associates; p.1-3, 41-3

- Reddy, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35, 435- 448. doi:10.1080/02602930902862859.
- Reynolds, J., Smith, R., Moskovitz, C., & Sayle, A. (2009). BioTAP: A Systematic Approach to Teaching Scientific Writing and Evaluating Undergraduate Theses. *BioScience*, 59 (10), 896–903. doi:10.1025/bio.2009.59.10.11
- Riggs, I. M. ve Enochs L. G. (1990). Toward the development of an elementary teacher’s science teaching efficacy belief instrument. *Science Education*, 74(6), 625-637.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Schwarzer R. (1993). General perceived self-efficacy in 14 cultures. Retrieved on 5-June 2007, at URL <http://Web.Fu-Berlin.De/Gesund/Publicat/Ehpscd/Health/World14.Htm>
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Fourth Edition. Needham Heights, MA: Allyn & Bacon.
- Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Nobel Yayınları, Ankara
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68, 202-248.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education*, 17, 783-805.
- Venning, J., & F. Buisman-Pijlman (2013). Integrating assessment matrices in feedback loops to promote research skill development in postgraduate research projects. *Assessment and Evaluation in Higher Education*, 38 (5): 567–579.
- Vieira, A. L. (2011). *Preparation of the analysis. Interactive LISREL in practice*. 1st ed. London: Springer.
- Wiggins, G. (1991). Standart, not standardization: Evoking quality student work. *Educational Leadership*. 48(5), 18-25.
- Yılmaz, M., Köseoğlu, P., Gerçek, C., Soran, H. (2004). Yabancı dilde hazırlanan bir öğretmen öz-yeterlik ölçeğinin Türkçeye uyarlanması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 27, 260-267.
- Zimmerman, B.J. (2000). Self-Efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82–91. doi:10.1006/ceps.1999.1016

Tablo Ek-1. Rubriklere yönelik öz yeterlik ölçeği

Rubrik Özyeterlik Ölçeği		Kesinlikle katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle katılıyorum
1	DPA'ları sınıf içerisinde nasıl uygulayacağım konusunda yeterli bilgiye sahip olduğuma inanıyorum					
2	Öğretmenlik deneyimime rağmen. DPA hazırlarken zorlanabilirim					
3	Öğrencilerin DPA'ya karşı olumsuz tutumları olsa bile. DPA uygulamalarını kolayca yapabileceğime inanıyorum					
4	DPA'lar ile öğrencilerin performanslarını kolayca değerlendirebilirim.					
5	DPA hazırlama konusunda yeterli bilgiye sahip olduğuma inanıyorum					
6	Öğrenciler daha önce kullanmamış olsalar bile. DPA'lar ile uygulamalar yapabilirim.					
7	Teorik bilgim olmasına rağmen. DPA'ları hazırlarken zorlanabilirim.					
8	Çok zaman aldığından dolayı DPA uygulamaları yaparken zorlanabilirim.					
9	DPA'yı hazırlarken bir sorunla karşılaşsam üstesinden gelebilirim.					
10	Etkili DPA'lar hazırlayarak öğrenci başarısını artırabilirim.					
11	DPA'ları grup çalışmalarında etkili şekilde kullanabilirim.					
12	DPA'ların kullanım amacını öğrencilere açıklamada zorlanabilirim.					
13	Farklı kaynaklardan bulduğum bir DPA'yı konuya adapte etmekte zorlanabilirim.					
14	DPA uygulamaları ile öğrencilerin konuya yönelik ilgilerini artırabilirim.					
15	Farklı DPA türlerine göre puanlama yapmakta zorlanabilirim.					
16	DPA'lar ile öğrencilerin öğrenme sürecindeki eksikliklerini kolaylıkla belirleyebilirim.					
17	Konuya uygun DPA türünü belirlemede zorlanabilirim.					
18	DPA'yı hazırlarken öğrenme kazanımlarını belirlemede zorlanabilirim.					
19	Sınıf yönetiminde zorlansam bile. DPA uygulamaları yapabilirim.					
20	Yazılı sınavları DPA ile adil şekilde değerlendirebilirim.					
21	DPA'yı hazırlarken ölçülecek davranışlara karar vermekte zorlanabilirim.					
22	Kalabalık sınıflarda DPA uygulamaları yaparken zorlanabilirim.					
23	Farklı kaynaklardan yararlanarak DPA hazırlama konusunda kendimi geliştirebileceğime inanıyorum.					
24	Hazırladığım DPA'ları öğrencilerin kolayca anlayabileceğine inanıyorum.					
25	DPA uygulamaları ile öğrencilerin konuyu öğrenmeye yönelik endişelerini azaltabileceğime inanıyorum:					
26	Uğraşsam bile. DPA'ları hazırlarken öğrenci davranışlarını ölçecek detaylı tanımlar yapmakta zorlanabilirim.					
27	Öğrencilerin nitelikli çalışmalar ortaya çıkarmalarını sağlayacak DPA'lar hazırlayabilirim.					
28	Eğer uğraşırsam. öğrencilere DPA kullanma becerisini kazandırabilirim.					