International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 180–197
https://doi.org/10.21449/ijate.1388256
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

Using self-organizing map and rating scale model in examining internal structure of scales

Sinan M. Bekmezci[1*], **Nuri Doğan**[2]

[1]Manisa Celal Bayar University, Faculty of Education, Mathematics and Science Education, Manisa, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Türkiye

ARTICLE HISTORY
Received: Nov. 9, 2023
Accepted: Nov. 29, 2024

Keywords:
Scale development,
Internal structure,
Self-organizing map,
Rating scale modal,
Exploratory factor analysis.

Abstract: This study compares the psychometric properties of scales developed using Exploratory Factor Analysis (EFA), Self-Organizing Map (SOM), and Andrich's Rating Scale Model (RSM). Data for the research were collected by administering the "Statistical Attitude Scale" trial form, previously used in a separate study, to 808 individuals. First, EFA, SOM and RSM were applied to decide the number of dimensions of the scale, and to select items. Subsequently, Confirmatory Factor Analysis (CFA) was used to the forms obtained from different methods and their CFA fit indices were compared. The analysis revealed variations in the number of dimensions and item distribution across different methods. Results indicated that the form generated using SOM exhibited the highest fit indices. Furthermore, the CFA fit indices of the form created with RSM were found to be satisfactory, offering detailed insights into both items and individuals.

1. INTRODUCTION

Acquiring and measuring knowledge, skills, and affective qualities are extremely important in education. Measurement procedures related to the affective qualities expected to be possessed by the individual find a vast place in academic studies and international exams. For this reason, obtaining statistically significant results from the data collection tools used in the measurement and evaluation processes is important to determine the level of these affective qualities. In this context, the psychometric properties of the scores obtained from the data collection tool, such as validity and reliability, should be aligned with the standards.

There are different definitions and classifications for validity in literature. In one approach, validity is defined as the degree to which a test can measure the desired feature without confusing it with other features (Ebel & Frisbie, 1991). Validity, as defined by the American Psychological Association, refers to the extent to which evidence and theory support the interpretations of test scores for their intended uses (AERA, 2014, p.11). According to AERA (2014), there are four validity evidences: test content, response processes, internal structure, relations to other variables. The internal structure includes the number of dimensions of the test, whether the structure is the same among different groups, which requires statistical calculations in its presentation. Depending on the type of test, the internal structure may be influenced by

*CONTACT: Sinan M. BEKMEZCİ sinan.bekmezci@cbu.edu.tr Manisa Celal Bayar University, Faculty of Education, Mathematics and Science Education, Manisa, Türkiye

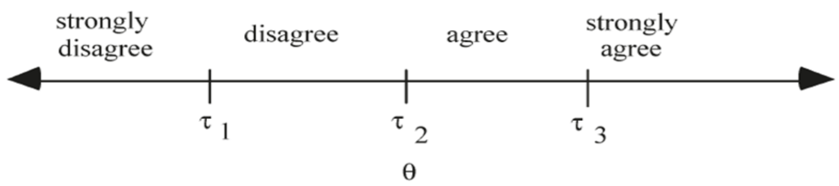
e-ISSN: 2148-7456

differences in item types within the test, type of item response category, unexpected nonlinear relationships between items, non-normal distribution of data, unexpected relationships between factors, and size of the sample. For this reason, when developing a measurement tool, it is thought to be critical to choose a statistical technique appropriate to the nature of the measurement tool to reveal evidence about the internal structure in a meaningful way. Fava and Velicer (1992) and Wood *et al*. (1996) have suggested that failure to accurately determine the number of dimensions may lead to the merging of dimensions with each other and item loading onto incorrect factors. The current study focuses on the internal construct validity evidence of Likert-type scales.

Scales can be employed to measure individuals' abstract features, called 'structure', that cannot be directly observed (Erkuş, 2012; Nunally & Bernstein, 1994). Likert-type scales typically consist of items with ordinal and multi-category responses. To statistically prove the validity of the scores obtained from the Likert-type scales, it is important to determine the number of dimensions of the scales and select the most qualified items. During the development process, various algorithms or models rooted in different theories may be applied for item selection and dimensionality determination. Factor analysis is commonly used in dimension determination and item selection (Anastasi, 1976; Crocker & Algina, 2008). The literature includes various comparative studies on techniques for determining the number of dimensions in Likert-type scales and selecting items. Examples of such studies include Şimşek's (2006) comparison of EFA with cluster analysis and multi-dimensional scaling, Doğan and Başokçu's (2010) comparison of factor analysis with stepwise cluster analysis, and Kozan's (2016) study comparing factor analysis, cluster analysis, and multi-dimensional scaling. In addition to classical methods, Item Response Theory (IRT) models also provide extensive opportunities for examining the psychometric properties of the Likert-type scale.

In this context, since the Likert-type scale typically consist of polytomous item, it would be reasonable to use polytomous IRT models to estimate non-linear relationships between the respondent's propensity level and the likelihood of responding in a specific category (Embretson & Reise, 2000). In IRT, model selection is made considering the dimensionality of the measurement tool (single or multiple), the categorization of item responses (dichotomous or polytomous), and the parameters to be estimated. Graded response model, partial credit model, rating scale model, and generalized partial credit model are among the models that can be used in the analysis of Likert-type ordinal polytomous measurement instruments (De Ayala, 2009; Embretson & Reise, 2000; Ostini & Nering, 2006). Polytomous IRT models reflect two types of conditional probabilities: (a) the probability of responding in a specific category and (b) the probability of responding positively instead of negatively at a particular boundary between two categories (Ostini & Nering, 2006). Current study, RSM (Andrich, 1978a; Andrich, 1978b) was employed. The reason for using RSM in this study is that the category threshold values calculated in the RSM process remain the same for all items in the scale (Bond & Fox, 2015). In addition to the fact that the calculated boundary values in RSM remain the same for all items, a boundary value equal to one less than the number of categories is calculated for all items in the scale. Figure 1 illustrates the representation of an item with four response categories and three thresholds.

Figure 1. *Category thresholds.*



According to Figure 1, an individual can respond positively to the first and second thresholds (leading to Category 3), respond negatively to the first and second thresholds (leading to

Category 1), or respond positively to the first threshold and negatively to the second threshold (leading to Category 2) (De Ayala, 2009). In RSM, each item has its own estimated item difficulty, and all items share a common threshold structure (Bond & Fox, 2015), the fact that all items share a common threshold structure distinguishes RSM from other polytomous IRT models.

In recent years, with the rapid advancement of information technologies, there has been a significant increase in the amount of data that can be collected. Among these large datasets, previously unknown relationships may exist, requiring more advanced analysis to find these relationships. Additionally, sometimes when the assumptions of classical methods cannot be met, the analysis results may not be sufficiently dependable. Considering such reasons, it has been thought that artificial neural network (ANN) algorithms, which are part of machine learning, can also be beneficial. ANN has proven successful in various fields, including object recognition, natural language processing, numerical prediction, and classification (Bramer, 2020), for purposes such as prediction, clustering, and classification (Eğrioğlu *et al*., 2019).

The ability of artificial neural networks to create complex and nonlinear models is also considered an advantage (Çevik & Tabaru-Örnek, 2020). Murtagh and Hernandez-Pajares (1995) call SOM a well-known neural network that is closely related to cluster analysis. Kohonen (1982), who discovered SOM, said that SOM is a network where neurons in a two-dimensional grid are individually adjusted for each input signal, particularly in large datasets and for different input patterns. Kohonen (2014) stated that while producing low-dimensional projections from high-dimensional data, SOM preserves similarity relationships. SOM is a nonlinear statistical technique and can be used to analyze multivariate data for exploration and clustering purposes, working with fewer traditional statistical principles, and requiring less in-depth knowledge than in statistical and multivariate analyses (Ferles *et al*. 2021; Galvan, *et al*., 2020; Galvan *et al*., 2020; Galvan *et al*., 2021; Haykin, 2009). The unsupervised learning model of SOM enables clustering in situations where groups are unknown. Competitive learning rule is used to perform unsupervised learning (Haykin, 2009). SOM selects a subset of "models," each consisting of a data vector and its best-matching model, called the "winner," along with spatial neighbors, and adjusts them for better matching (Kohonen, 2014). Distance measures are used to find the similarity between the input vector and neurons (Miljković, 2017). These distance measures include Euclidean, correlation, cosine similarity, and Manhattan distance (Grajciarova *et al*., 2012).

When reviewing the literature, it is evident that while some studies exclusively employed IRT models, others compared classical models with IRT models. For instance, Krishnan and Idris (2018), Cantó Cerdán *et al*. (2021), İlhan and Güler (2018), Peixoto *et al*. (2021), Takasaki *et al*. (2021) and Karlin and Karlin (2018) utilized various Rasch models in scale development studies. Additionally, Karlin and Karlin (2018) contended that besides assessing test validity, the Rasch model also offers an opportunity to evaluate students. In their studies, Uysal (2015) and İlhan and Güler (2017) compared classical test theory and IRT, with İlhan and Güler (2017) reported a strong correlation and relative agreement between ability measures obtained from Classical Test Theory and Rasch analysis. In a few rare studies, it is seen that SOM is also used. For instance, Kiang, and Kumar (2001) found that SOM outperformed pre-rotation factor analysis, providing more accurate insights into the latent structure, particularly in skewed data. Francis (2001) suggested the use of ANN models for dimension reduction in data with nonlinear relationships, comparing them to factor analysis in their study. Sadesky (2007) aimed to determine test structure using SOM and highlighted its ability to accurately preserve contiguity and proximity relations, suggesting further research applications in educational assessment. Tezbaşaran (2016) and Eriş Hasırcı (2019) employed the SOM method and reported its utility in item selection to determine the dimension-item structure of tests measuring affective features with polytomous items. In this context, it can be observed that IRT models may be effective in revealing the internal structure of Likert-type scales. Additionally, it appears that SOM can be useful in determining the number of dimensions and selecting items for the Likert-type scales.

Despite the advantages identified in the studies, it is thought that there is a significant gap in research on determining the psychometric properties of IRT models and artificial neural networks, especially in the context of Likert-type scales. In addition, it is thought that comparative studies evaluating the effectiveness of these methods and clarifying their advantages and disadvantages are lacking. For this purpose, the current study was designed to compare different techniques that can be used in the test development process. In this context, it is thought that this study can help researchers who want to develop a Likert-type scale in the process of dimension determination and item selection.

For this purpose, the study aims to answer the following question: "What are the confirmatory factor analysis CFA fit indices for test forms developed from items selected based on the results of EFA, SOM and RSM?"

2. METHOD

In the initial step of the study, the pilot version of the Statistical Attitude Scale was administered. In the second step, variations of the Statistical Attitude Scale were created using EFA (1 dimension - 32 item), SOM (1 dimension - 16 item, 2 dimension - 26 item), and RSM (1 dimension - 25 item, 1 dimension - 32 item). The fit indices obtained by separately applying CFA to these 5 models were compared, and the form with the best-fit values was revealed. In this context, the research has descriptive features in terms of revealing scale structures to be obtained from different techniques (Grove *et al*., 2012) and relational features in terms of comparing CFA fit indexes of scale structures obtained from different techniques (Büyüköztürk *et al*., 2017).

2.1. Study Group

The Attitude Scale Towards Statistics trial form used in the study was administered to 808 university students and graduates. While 26% (*N*: 211) of the group is male, 74% (*N*: 597) were female. While 32% (*N*: 257) of the participants stated that they had not taken a statistics course before, 68% (*N*: 551) stated that they had taken a statistics course. "Which of the following best describes your academic field?" It is seen that 36% (*N*: 289) quantitative, 43% (*N*: 347) equal weight (quantitative and qualitative), 18% (*N*: 149) qualitative, and 3% (*N*: 23) foreign language answers were given to the question.

2.2. Data Collection

Data were collected in the fall semester of 2020 after obtaining the approval from the Hacettepe University Ethics Committee. During the research process, the scale was planned to be administered face-to-face. However, when the implementation process was started, the scale trial form was delivered to the participants online due to the national and international COVID-19 pandemic.

2.3. Instruments

There are 40 items in this scale, which were prepared to determine the level of attitude towards statistics. Twenty of the items are positive, and 20 of them are negative statements. Negatively worded statements were reverse scored. The scale trial form and the descriptive statistics of the items in the scale are in Appendix A. Item pairs m9-m40 and m1-m31 in the scale were used as control items. Before the analyses were performed, the polychoric correlation value between m1-m31 (.69) and m9-m40 (.77) item pairs was significant at .01 significance level; it was determined that there is a relationship above the medium level. Therefore, control items (m31 and m40) with weaker psychometric properties were excluded from the scale in the subsequent analyses.

2.4. Data Analysis

Analyzes were performed with MATLAB, Winsteps, JASP, Factor, and R software. To provide cross-validation before the analyses were performed, the data set was randomly divided into three groups of 33%, and EFA was performed with the JASP program over three groups. After these analyses, it was found that there was no significant difference between the factor loadings of the same items in different forms. Afterward, the polychoric correlation matrix was used in the EFA performed with the entire data set. In addition, estimations were made using the unweighted least squares method since the item responses were in the ordering scale and the responses were not normally distributed. Parallel Analysis was also used to decide the number of dimensions. The scale's psychometric properties were determined in line with the variance rates explained because of the analysis, item factor load, and dimensionality convergence index.

SOM analyses were performed in the MATLAB software. SOM works by creating maps consisting of "nxn" neurons, where "n" is the number of neurons, each neuron represents a dimension, and which items are in which dimensions, and this distribution can change at different "n" values. In the study, before the final evaluation according to SOM, the item and size distributions in the maps consisting of 4x4, 3x3, and 2x2 neurons were examined, and it was determined that the model consisting of 3x3 neurons was ideal and the items to be selected for the final test were determined through this model.

RSM analysis was performed using Winsteps software. In the RSM analysis, the items selected for the final form were chosen according to their infit and outfit values. Accordingly, a form was obtained by removing the items with at least one of the infit or outfit values above "1.4" (Bond & Fox, 2015). Then, the items with at least one of the infit or outfit values above "1" were removed, and another form was created.

Since each of the techniques discussed in the research performs its own calculations, instead of directly comparing the results, CFA was applied via the JASP software to the different models established, and each model's fit and error indices were compared. In CFA, diagonally weighted least squares estimation (DWLS), which is the most reliable parameter estimation method for ordinal data and generally when the variables are not normally distributed, was used (Brown, 2006; Flora & Curran, 2004; Li, 2016).

3. FINDINGS

3.1. Findings Related to EFA

The results of KMO and Bartlett Sphericity Test obtained from EFA are given in Table 1. A KMO value above .90 indicates that the sample size is appropriate for factoring, and that the Bartlett test is statistically significant, showing that the inter-item correlation matrix is suitable for factor extraction (Şencan, 2005). As shown in Table 1, data were suitable for EFA.

Table 1. *KMO-Bartlett's sphericity test results.*

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.978
Bartlett's Test of Sphericity	Approx. Chi-Square	27891.0
	df	703
	Sig	.000

Table 2 shows the eigenvalues obtained using the unweighted least squares technique and the explained variance percentages. Çokluk *et al*. (2012) stated that one of the points to be taken into consideration when deciding on the number of factors is the percentage of variance explained and that the number of factors can be determined as "1" if the percentage of variance explained by the eigenvalues after the percentage of variance explained by the first eigenvalue is significantly reduced. As shown in Table 2, the first eigenvalue has a 59% contribution to the explained variance. In comparison, the second eigenvalue has a relatively low contribution of 9%. According to Table 2, the scale is one-dimensional.

Table 2. *Explained variance based on eigenvalues.*

Factor	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
1	22.26430	.58590	.58590
2	3.56160	.09373	.67963
3	1.62018	.04264	.72227
4	0.91090	.02397	

Table 3 shows the matrix of unrotated factor loadings. The reason for using unrotated factor loadings in the analysis process is that there is no similar process in the analyses of SOM and RSM. Since the last item in Table 3 is 39, there seem to be 39 items in the analysis, but there are 38 items in the analysis. This is because the item numbers were kept the same to avoid confusion when the analysis was performed after item 31 and item 40 were subtracted and compared with the items selected by other methods. The same is true for Table 4. Table 3 shows that item 3, item 15, and item 16 communality values are less than 0.50. And item 34 and item 38 give the highest loading values to the third factor. In addition, the PA technique with one-dimensional limitation was applied to support the one-dimensionality assessment based on Table 2 and Table 3.

Table 3. *Unrotated loading matrix.*

Item	F1	F2	F3	Communality	Item	F1	F2	F3	Communality
1	.839	-.227	-.084	.763	20	.840	.232	-.106	.771
2	.819	-.238	-.087	.736	21	.843	.140	-.098	.740
3	.525	.402	.110	**.449**	22	.766	-.360	-.094	.726
4	.757	.229	-.010	.626	23	.789	-.388	-.083	.780
5	.862	-.221	-.020	.792	24	.831	-.190	.035	.727
6	.774	-.344	-.088	.726	25	.841	-.332	-.053	.820
7	.715	.375	.071	.656	26	.776	-.426	-.060	.786
8	.768	.240	-.201	.688	27	.775	.292	.052	.689
9	.739	.276	-.218	.669	28	.728	.305	-.149	.646
10	.760	.260	-.247	.706	29	.831	.204	-.040	.735
11	.770	-.417	-.032	.767	30	.750	-.269	.008	.635
12	.837	-.133	.064	.723	32	.862	.287	-.012	.826
13	.843	.300	-.016	.801	33	.762	.339	.139	.714
14	.812	.304	.052	.755	34	.514	.018	**.621**	.650
15	.538	.351	.036	**.414**	35	.783	.377	-.013	.755
16	.302	.326	.083	**.204**	36	.630	-.170	.444	.623
17	.769	-.317	.005	.691	37	.803	-.389	.005	.796
18	.868	-.186	.008	.788	38	.514	-.042	**.629**	.662
19	.863	.210	-.011	.789	39	.741	-.372	-.026	.688

Table 4 shows the PA factor loadings matrix with a one-dimensional boundary. Table 4 has shown that the communality values of item 3, item 15, item 16, item 34, item 36 and item 38 were less than 0.5. In addition, when the one-dimensional convergence indices obtained as a result of PA are examined, the unidimensionality fit index (UniCo), which is desired to be more than 0.95, was found to be 0.971, the explained common variance index (ECV), which is desired to be more than 0.85, was found to be 0.866, the mean index of the residual absolute loadings of the item, which was required to be less than 0.30, was found to be 0.284. These results also support the interpretation that the scale is one-dimensional.

Table 4. *Unrotated loading matrix according to single factor limitation.*

Item	F1	Communality	Item	F1	Communality
1	.868	.754	20	.843	.710
2	.846	.715	21	.841	.707
3	.538	**.289**	22	.798	.638
4	.749	.561	23	.822	.675
5	.882	.778	24	.830	.690
6	.804	.647	25	.865	.749
7	.728	.530	26	.821	.673
8	.779	.607	27	.789	.623
9	.759	.576	28	.738	.544
10	.780	.609	29	.838	.702
11	.808	.654	30	.760	.578
12	.846	.716	32	.885	.783
13	.854	.730	33	.780	.608
14	.830	.689	**34**	**.514**	**.264**
15	.544	**.295**	35	.806	.650
16	.303	**.092**	36	.630	**.397**
17	.787	.620	37	.824	.679
18	.872	.760	**38**	**.518**	**.268**
19	.869	.754	39	.770	.593

McDonald's Omega reliability coefficient of the form was calculated as .97, and Cronbach's Standardized Alpha reliability coefficient was calculated as .97. These values provide evidence that the results obtained from the measurement tool are reliable.

As a result, it is likely appropriate to exclude item 16, item 34, and item 38 from the test, whose factor loadings in Table 3 are not high enough and/or give high loads to more than one dimension. In addition to these items, item 3, item 15, and item 36, whose common variance values were less than .5 according to the PA results made with one-dimensionality limitation, were not included in the final form. Thus, a one-dimensional and 32-item scale was obtained.

3.2. Findings Related to SOM

Figure 2 shows a SOM graph consisting of 3x3 neurons, n=3. It says which dimension and how many items are in the hexagons. The neuron and hence the size numbering is done in a specific order, with the lower left neuron being the first neuron and the upper right neuron being the ninth neuron. According to Figure 2, there are 16 items in neuron number eight, 10 items in neuron number three, 3 items in neuron number one, 2 items each in neurons 6 and 7, and 1 item in neurons 2, 4, and 5. The fact that the inside of a neuron is fully colored indicates that the neuron is saturated, without the color being important (different programs may have different color representations). We can consider the saturation of the neuron as the network receives as much material as a neuron can receive at a given number of pieces of training. Accordingly, the eighth neuron was fully saturated, the third neuron was nearing saturation, and the remaining neurons were not saturated According to Figure 2, items are collected in 2 main dimensions.

Figure 2. *Graph of hits (item neuron distribution) for SOM analysis.*

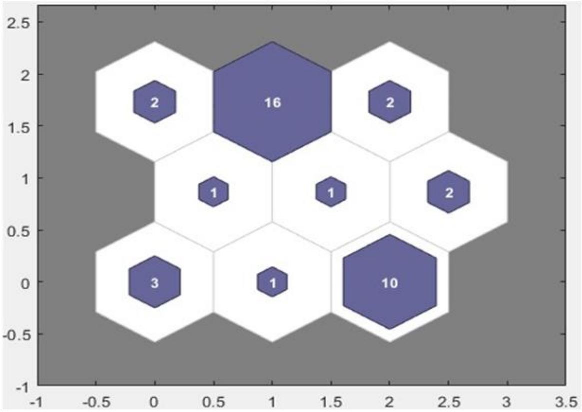


Figure 3 gives the color representation of the distance between neurons. By examining the "Neighboring Weight Distances" between neurons in Figure 3 and the items in the scale, items in close dimensions (neurons) can be considered together. The fact that the colors go from light to dark shows that the distance between the dimensions increases. In this case, the dimensions 3-6 and 5-9 are close to each other and the items in these dimensions can be combined. The researcher can decide by considering the items' distances and the theoretical structure.

Figure 3. *Graph of hits (item neuron distribution) for SOM analysis.*

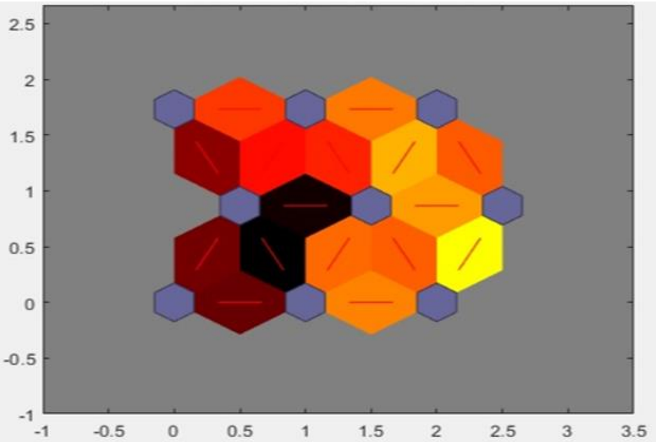


Table 5 shows which item is included in which dimension. When Figure 2 and Figure 3 are evaluated together, a 1- or 2- dimensional structure can be created by considering the theoretical background of the scale. According to Table 5, items 34, 36, and 38 belong to the first dimension; item 24 to the second dimension; item 15 to the fourth dimension; item 5 to the fifth dimension; items 1 and 2 to the sixth dimension; items 3 and 16 to the seventh dimension; and items 12 and 18 to the ninth dimension.

Table 5. *SOM substances neuron distribution.*

Dimensions	Item
1	34, 36, 38
2	24
3	6, 11, 17, 22, 23, 25, 26, 30, 37, 39
4	15
5	5
6	1, 2
7	3, 16
8	4, 7, 8, 9, 10, 13, 14, 19, 20, 21, 27, 28, 29, 32, 33, 35
9	12, 18

According to Figure 1, a one-dimensional form comprising 16 items was created with items 4, 7, 8, 9, 10, 13, 14, 19, 20, 21, 27, 28, 29, 32, 33, and 35 located in the eighth dimension, which is saturated. Another two-dimensional form with 26 items was created by adding items 6, 11, 17, 22, 23, 25, 26, 30, 37 and 39 in the third dimension to the items in the eighth dimension.

3.3. Findings Related to RSM

Rasch analysis examines dimensionality through primary components analysis of Rasch measurement residuals. Since EFA was performed during the research process and RSM dimensionality results were also highly like these results, the dimensionality outputs of RSM were not included in the findings.

Table 6 shows descriptive statistics, and reliability values of 804 people who were determined as non-extreme (outlier) individuals. As shown in Table 6, fit statistics of the person' mean are very close to 1, and the data on the individuals provide the model-data fit. In addition, The Cronbach Alpha reliability coefficient is relatively high as .97. The fact that the separation index of individuals is more than 2 (Linacre, 2012) indicates that students with different attitudes toward statistics are effectively differentiated.

Table 6. *Summary of 804 measured (non-extreme) person.*

	Total Score	Measure	Model S. E.	Infit Mnsq	Infit Zstd	Outfit Mnsq	Outfit Zstd
Mean	116.6	0.0674	0.2039	1.05	-0.3	1.09	-0.2
P. Sd.	33.3	1.2829	0.0762	0.63	2.4	0.79	2.3
S. Sd.	33.4	1.2837	0.0762	0.63	2.4	0.79	2.3
Max.	189.0	5.2893	1.0037	4.08	8.4	9.90	9.0
Min.	39.0	-5.0817	0.1736	0.19	-5.9	0.20	-5.7
Real RMSE .2502 True SD 1.2583				Separation 5.03			Person Reliability .96
Model RMSE .2177 True SD 1.2643				Separation 5.81			Person Reliability .97
S.E. of Person Mean = 0.0453							
Person Raw Score-to-Measure Correlation = .96							
Cronbach Alfa (KR-20) Person Raw Score “Test” Reliability = 0.97 SEM = 5.41							

Table 7. *Item measurement reports.*

	Total Score	Measure	Model S. E.	Infit Mnsq	Infit Zstd	Outfit Mnsq	Outfit Zstd
Mean	2478.0	0.0000	0.0417	1.01	-0.9	1.09	-0.1
P. Sd.	477.4	0.7979	0.0018	0.35	5.2	0.50	4.7
S. Sd.	483.9	0.8086	0.0018	0.36	5.3	0.51	4.8
Max.	3362.0	1.2794	0.0486	2.28	9.9	3.13	9.9
Min.	1715.0	-1.6056	0.0398	0.62	-9.1	0.60	-8.4
Real RMSE 0.0445 True SD 0.7966				Separation 17.92			Item Reliability 1.00
Model RMSE 0.0417 True SD 0.7968				Separation 19.11			Item Reliability 1.00
S.E. of Item Mean = 0.1312							
Item Raw Score-to-Measure Correlation =-1.00							

Table 7 shows descriptive statistics, and reliability values of the items. The items' mean infit and outfit values in Table 7 are very close to 1. Item reliability value is 1. The item reliability index indicates the consistency of item placements when the same items are given to another sample with similar behavior (Bond & Fox, 2015). In addition, since the item reliability level is relatively high, there is no need to look at the item separation index.

Table 8 gives category threshold values, category infit and outfit values, and descriptive statistics of categories. The threshold values in Table 8 and the infit and outfit values of these values show us that the item categorizations work properly. As shown the fit statistics in Table 8 which between .5 and 1.5, indicating that the categories work properly (Linacre, 2014). Therefore, the answer categories and the five-point Likert rating work properly.

Table 8. *Category threshold values.*

Category	%	Observed Average	Sample Expect	Infit	Outfit	Andrich Threshold	Category Measure	Estimate Disc.
1	18	-1.742	-1.71	1.01	1.24	NONE	-2.56	
2	17	-0.7128	-0.746	1.01	1.07	-1.14	-1.10	1.10
3	24	0.0168	0.0149	0.84	0.81	-0.69	-0.05	1.00
4	22	0.8005	0.7728	0.86	0.92	0.48	1.08	1.00
5	19	1.7060	1.738	1.21	1.38	1.35	2.69	0.91

Table 9 gives the item measure values, which can be interpreted as RSM item difficulty and infit-outfit. The measurement value gives information about the item's difficulty level (the level of having the measured feature) (Linacre, 2012). Small measure values in Likert-type scale items indicate that the state of expressing strongly disagree and disagree is high, and the state of agreeing and strongly agree is low. According to Table 9, item 3 has the lowest difficulty, while item 23 has the highest. Table 9 also shows that items 3, 15, 16, 34, 38, and 39 have infit or outfit values of 1.4 or higher, indicating they are not compatible with the model. Consequently, these items were removed, resulting in a one-dimensional form with 32 items.

Table 9. *Rating scale model results.*

Item	Measure	Infit	Outfit	Item	Measure	Infit	Outfit
1	0.7568	0.6355	0.6802	20	-0.2739	0.7668	0.7792
2	0.9528	0.7613	0.928	21	-0.0963	0.8723	0.829
3	-1.6056	1.5467	1.6677	22	0.9562	0.9055	0.8517
4	-0.7262	1.0874	1.056	23	1.2794	0.8178	0.8317
5	0.2884	0.6798	0.6612	24	0.4397	0.6994	0.7171
6	1.1631	0.8623	1.0584	25	0.9339	0.7293	0.8302
7	-1.1194	1.2334	1.1815	26	1.2663	0.8653	0.9299
8	-0.13	0.9958	0.9749	27	-.4788	1.0177	1.1069
9	-0.3889	1.1492	1.3611	28	-.3608	1.102	1.1208
10	-0.0547	0.8923	0.94	29	-0.2398	0.8915	0.8378
11	1.1233	0.8498	0.8813	30	0.6808	0.9141	0.9163
12	-0.0324	0.8049	0.8832	32	-0.5719	0.8127	0.7784
13	-0.846	0.9183	0.8535	33	-0.9942	0.9958	0.9914
14	-0.5277	0.9456	0.8961	34	-1.061	1.7331	1.8537
15	-0.4822	1.8003	2.5511	35	-0.7474	0.9677	0.9364
16	-0.6113	2.2792	3.131	36	-.2903	1.3273	1.3758
17	0.7319	0.7922	0.8461	37	1.0965	0.7163	0.7583
18	0.1916	0.6234	0.6017	38	-1.169	1.5985	1.5629
19	-0.2414	0.8702	0.8517	39	1.1886	0.9938	1.48

In addition, infit and outfit statistics are desired to be 1.00 because it shows perfect match of between data and modal (Güler *et al*., 2018). For this reason, an attempt was made to ensure perfect data-modal compatibility by considering that items with at least one of the infit and outfit values greater than '1' could be deleted. Accordingly, another one-dimensional 25-item form was created by removing items numbered 3, 4, 6, 7, 9, 15, 16, 27, 28, 34, 36, 38, 39, of which at least one of the infit, and outfit values are greater than 1.

3.4. CFA Findings of Different Models

Below is a table of critical values for CFA indexes and tables of CFA indexes of different forms. Table 10 presents the critical values for CFA as suggested by Schermelleh-Engel *et al*. (2003).

Table 10. *Critical values for CFA indices.*

Indices	Good Fit	Acceptable Fit
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 < \chi^2/df \leq 3$
RMSEA	$.00 \leq$ RMSEA $\leq .05$	$.05 <$ RMSEA $\leq .08$
NFI	$.95 \leq$ NFI ≤ 1.00	$.90 \leq$ NFI $< .95$
NNFI	$.97 \leq$ NNFI ≤ 1.00	$.95 \leq$ NNFI $< .97$
CFI	$.97 \leq$ CFI ≤ 1.00	$.95 \leq$ CFI $< .97$
GFI	$.95 \leq$ GFI ≤ 1.00	$.90 \leq$ GFI $< .95$
SRMR	$.00 \leq$ SRMR $\leq .05$	$.05 <$ SRMR $\leq .10$

Table 11 shows the fit indices of different models. Considering the fit indices of the one-dimensional and 32-item model obtained from EFA in Table 11, χ^2/df, and RMSEA values are not within acceptable limits. SRMR value shows acceptable fit. CFI, NFI, NNFI, and GFI values show a good fit. The fit indices of the one-dimensional and 16-item SOM model show good fit. In addition, the fit indices of the two-dimensional and 26-item SOM model also show good fit. As shown in Table 11, there was no difference of more than .01 between the fit indices of the two models generated through SOM. It can be stated that there is no statistically significant difference between the models. This comment was made because, according to Cheung and Rensvold (2002), a difference of less than .01 between the indices indicates measurement invariance. When the items in the two-dimensional model created by SOM were examined, it was seen that there were items consisting of positive expressions in one dimension and items consisting of negative expressions in the other dimension.

Table 11. *Fit indices of different models.*

	EFA	SOM		RSM	
Indices	1 D. 32 item	1 D. 16 item	2 D. 26 item	1 D. 25 item	1 D. 32 item
χ^2/df	7.133	1.493	1.493	6.668	6.627
CFI	.977	.999	.999	.981	.978
NNFI	.975	.998	.999	.979	.977
NFI	.973	.996	.995	.978	.975
RMSEA	.087	.025	.019	.084	.084
GFI	.977	.997	.996	.982	.979
SRMR	.093	.040	.039	.088	.090

The one-dimensional 25-item model obtained with the RSM shows that the χ^2/df value is not within acceptable limits, although the RMSEA and SRMR values are within the limit. Additionally, the CFI, NFI, NNFI, and GFI values indicate a good fit. The one-dimensional 32-item model obtained with the RSM shows that the χ^2/df value is not within acceptable limits, while the RMSEA and SRMR values are within the limit. Additionally, the other indices indicate a good fit. When comparing the fit indices of the two models obtained by RSM, it is observed that there is no difference greater than 0.01 in all values except for χ^2/df. However, both models' χ^2/df values are outside the acceptable range.

According to Table 11, there is a difference of more than 0.01 between the fit indices of both the one-dimensional and two-dimensional models obtained through SOM and the fit indices of the one-dimensional model obtained through EFA. SOM achieved higher fit values than EFA with the additional 16 items it removed, in addition to the 6 items eliminated according to EFA, while maintaining the same number of dimensions. This may indicate that SOM is more sensitive to item selection than EFA. In cases where application time is important, SOM may be more useful to measure with the same sensitivity with a shorter form. In addition, it is seen that some fit indices of the one-dimensional model obtained from SOM are more than .01 different from some of the fit indices of one-dimensional models obtained by RSM. In this context, the model obtained from SOM indicates a significantly better fit than the models obtained with RSM. When comparing the fit indices of the one-dimensional 32-item models obtained from RSM and EFA, we can see that all the fit statistics are almost the same except for χ^2/df. Out of the 6 items removed according to RSM and EFA, only one of them shows a difference. The RSM shows good fit values with one dimension of 25 items.

4. DISCUSSION and CONCLUSION

The fit indices of the CFA applied to the different forms were examined. According to the results, χ^2/df in the fit indices of the CFA applied to the form obtained by EFA, as well as indices other than SRMR, indicate a good fit. Additionally, all the fit indices obtained from CFA applied to two different forms obtained by SOM indicate good fit. Similarly, the fit indices obtained from the CFA applied to two forms obtained with RSM also indicate a good fit.

The high fit indices of the scale construct obtained from SOM indicate that SOM effectively distributes the dimensional items during test development. This finding is consistent with the results of Tezbaşaran (2016) and Eriş Hasırcı (2019), suggesting that SOM can serve as an alternative to EFA. In his study aiming to determine the structure of multiple-choice tests, Sadesky (2007) also stated that SOM could be effectively used to represent test data. Additionally, Stambuk *et al*. (2007), Astel *et al*. (2007), Chattopadhyay *et al*. (2011), and Das *et al*. (2016) made similar comments, highlighting SOM's suitability for large dataset analysis and result visualization compared to cluster analysis and principal component analysis. Sorrosal Forradellasa *et al*. (2012) used SOM to identify the most relevant items in the data and concluded that the methodology they employed in the study was appropriate.

One of the remarkable results here is that although the data did not strictly meet the assumption of normal distribution in the analysis process, the fit index values of the models belonging to SOM are significantly higher than the fit index values of the models obtained from RSM and EFA when estimations in CFA are made using maximum likelihood instead of diagonally weighted least squares. In this context, it can be interpreted that SOM is more effective in explaining skewed data and creating a better model from these data. This aligns with Francis's (2001) findings suggesting that ANN models are suitable for dimension reduction in datasets with nonlinear relationships, and with the observations of Kiang and Kumar (2001), indicating that SOM effectively detects the structure of non-normally distributed data. In addition, when the final scales were examined on an item basis, it was seen that SOM created models with higher CFA fit indices with the items it deleted in addition to the items that RSM and EFA deleted from the test. Accordingly, with this result, it can be said that SOM performs a more sensitive item selection process.

When examining the CFA fit indices, it becomes evident that RSM, and consequently, IRT models, can be effectively employed in the process of test development. Since the infit and outfit values of the items obtained from the RSM exceed 1.4, it is observed that only one item differs when comparing the items removed from the test and those derived from the EFA results.

Çelen (2008) further suggests that IRT can be viewed as an alternative approach to attaining similar outcomes in the test development process. Among the techniques utilized in this study, RSM offers more detailed statistics on both item and individual levels compared to EFA and SOM and model-data inconsistencies also offer detailed insights into relevant items and individuals. Similar to the findings of Karlin and Karlin (2018), the Rasch model not only validates the test but also enhances student evaluation, providing more accurate results than raw scores. Krishnan and Idris (2018) also observed that the Rasch model effectively determines psychometric properties by identifying and correcting items that compromise validity, rather than re-evaluating all test items. This approach, as demonstrated by Cantó-Cerdán *et al*. (2021), Sözer and Kahraman (2021), Peixoto *et al*. (2021), and Takasaki *et al*. (2021), proves beneficial in scale development and adaptation.

When the findings obtained throughout the study are generally evaluated, it can be stated that SOM and RSM are quite useful techniques in revealing the internal structure of the test. Considering the items of the scale used in the research, it has been observed that items with positive and negative expressions tend to go in two separate dimensions. However, when the theoretical foundations of this structure are considered and items are limited to a single dimension, it has been observed that the model belonging to SOM gives higher fit indices than the one-dimensional model obtained from EFA. In this case, studies can be conducted to examine the effect of deleting more items according to SOM on the test content of the scale. Additionally, studies can be conducted to determine whether the tendency of items with positive and negative expressions to go in two separate dimensions is due to the measurement of different sub-factors of the items or is related to the grammatical structure used in expressing the items. A real data set was used in this study. A simulation study can be conducted to test the performance of SOM in case of different levels of deviation from normality in the used data.

Acknowledgments

This research is part of thesis "Examination of the Construct Validity of Likert Type Scales with Data Mining, Artificial Neural Networks, Factor Analysis and Rasch Model".

Limitations

This scale was applied online due to the COVID-19 Pandemic. For this reason, the sample's level of representation of university students and graduates may be weak.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Hacettepe University Ethics Committee, 35853172-300.

Contribution of Authors

Sinan M. Bekmezci: Research Design, Methodology, Data Collection, Data Analysis, and Writing. **Nuri Doğan**: Research Design, Methodology, Data Collection, Supervision, Critical Review.

Orcid

Sinan M. Bekmezci https://orcid.org/0000-0001-5190-1894
Nuri Doğan https://orcid.org/0000-0001-6274-2016

REFERENCES

AERA. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Anastasi, A. (1976). *Psychological testing*. The Macmillan Company, Collier Macmillan Limited.

Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Astel, A., Tsakovski, S., Barbieri, P., & Simeonov, V. (2007). Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*, *41*, 4566-4578. https://doi:10.1016/j.watres.2007.06.030

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Bramer, M. (2020). *Principles of data mining*. Springer.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. The Guildford Press.

Büyüköztürk, Ş., Çakmak, E.K., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri [Scientific research methods].* Pegem Akademi.

Cantó-Cerdán, M., Cacho-Martínez, P., Lara-Lacárcel, F., & García-Muñoz, A. (2021). Rasch analysis for development and reduction of symptom questionnaire for visual dysfunctions (SQVD). *Scientific Reports, 11*(14855), 1-10. https://doi.org/10.1038/s41598-021-94166-9

Chattopadhyay, M.J., Dan, P.K., & Majumdar, S. (2011). Principal Component Analysis and Self Organizing Map for Visual Clustering of Machine-Part Cell Formation in Cellular Manufacturing System. https://arxiv.org/pdf/1201.5524.pdf

Crocker, L.M., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlilik ve güvenilirliğinin karşılaştırılması. [Comparison of validity and reliability of two tests developed by classical test theory and item response theory]. *Elementary Education Online*, *7*(3), 758-768.

Çevik, M., & Tabaru-Örnek, G. (2020). Comparison of MATLAB and SPSS software in the prediction of academic achievement with artificial neural networks: Modeling for elementary school students. *International Online Journal of Education and Teaching*, *7*(4), 1689-1707.

Çokluk, Ö., Şekercioğlu G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları. [SPSS and LISREL applications of multivariate statistics for social sciences].* Pegem Akademi.

Das, G., Chattopadhyay, M., & Gupta, S. (2016). A comparison of self-organizing maps and principal components analysis. *International Journal of Market Research*, *58*(6). 1-20. https://doi.org/10.2501/IJMR-2016-039

De Ayala, R.J. (2009). *The theory and practice of item response theory*. Guilford Press.

Doğan, N., & Başokçu, O. (2010). İstatistik Tutum Ölçeği İçin Uygulanan Faktör Analizi ve Aşamalı Kümeleme Analizi Sonuçlarının Karşılaştırılması. [Comparison of factor analysis and progressive cluster analysis results applied to the statistics attitude scale]. *Journal of Measurement and Evaluation in Education and Psychology*, *1*(2), 65-71.

Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice- Hall International.

Eğrioğlu, E., Yolcu, U., & Baş, E. (2019). *Yapay sinir ağları öngörü ve tahmin uygulamaları* [*Artificial neural networks prediction and prediction applications*]. Nobel yayınları.

Embretson, S.E., & Reise, S. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Eriş Hasırcı, H.M. (2019). *Öz düzenlemeli haritalar yöntemi ile elde edilen yapı geçerliği kanıtlarının faktör analizi ve kümeleme analizi ile karşılaştırılması. [Comparison of* *construct validity proofs obtained from self-organizing maps method with the factor analysis and clustering analysis]* [Unpublished doctoral dissertation]. Ankara University.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-1: Temel kavramlar ve işlemler. (Measurement and Scale Development in Psychology- Part 1: Basic Concepts and Processes)*. Pegem Akademi.

Fava, J.L., & Velicer, W.F. (1992). The effects of over extraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 387-415.

Ferles, C., Papanikolaou, Y., Savaidis, S.P., & Mitilineos, S.A. (2021). Deep learning self-organizing map of convolutional layers. *Computer Science and Information Technology (CS & IT) Computer Science Conference*, 25-32. https://doi.org/10.5121/csit.2021.110303

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Francis, L. (2001). Neural Network Demystified. *2001 Winter Forum Ratemaking Discussion Papers and Data Management/Data Quality/Data Technology Call Papers*. www.cacact.org./pubs/forum/01wforum/01wf253.pdf.

Galvan, D., Cremasco, H., Gomes Mantovani, A.C., Bona, E., Killner, M., & Borsato, D. (2020). Kinetic study of the transesterification reaction by artificial neural networks and parametric particle swarm optimization. *Fuel 267*, *117221*. https://doi.org/10.1016/j.fuel.2020.117221

Galvan, D., Effting, L., Cremasco, H., & Conte-Junior, C.A. (2020). Can socioeconomic, health, and safety data explain the spread of COVID-19 outbreak on Brazilian Federative Units? *International Journal of Environmental Research and Public Health*, *17*(8921), 2-16. https://doi.org/10.3390/ijerph17238921

Galvan, D., Effting, L., Cremasco H., & Conte-Junior, C.A. (2021). The spread of the Covid-19 outbreak in Brazil: An overview by Kohonen self-organizing map networks. *Medicina*, *57*(*235)*, 2-19. https://doi.org/10.3390/medicina57030235

Güler, N., İlhan, M., & Taşdelen Teker, G. (2018). Comparing the scale values obtained from pairwise comparison scaling method and Rasch analysis. *Inonu University Journal of the Faculty of Education*, *19*(1), 31-48. https://doi.org/10.17679/inuefd.400386

Grajciarova, L., Mares, J., Dvorak P., & Prochazka, A. (2012). Biomedical image analysis using self-organizing maps. *Technicka*, *5*, 166.

Grove, S.K., Burns, N., & Gray, J.R. (2012). *The practice of nursing research: Appraisal, synthesis, and generation of evidence* (7th Ed.). Elsevier.

Haykin, S. (2009). *Neural networks and learning machines* (3rd Edition). Pearson Education.

İlhan, M., & Güler, N. (2017). Likert tipi ölçeklerde klasik test kuramı ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki uyumun test edilmesi. [Testing the agreement between ability estimations made through classical test theory and the ones made through Rasch analysis in Likert type scales]. *Ege Journal of Education*, *18*(1). 244-265.

İlhan, M., & Güler, N. (2018). Likert tipi ölçeklerde Rasch modelinin kullanımı: olumsuz değerlendirilme korkusu ölçeği-öğrenci formu (ODKÖ-ÖF) üzerinde bir uygulama. [The use of Rasch model in Likert type scales: An application on the fear of negative evaluation scale-student form (FNE-SF)]. *Trakya Journal of Education*, *8*(4), 756-775.

Karlin, O., & Karlin S. (2018). Making better tests with the Rasch measurement model. *InSight: A Journal of Scholarly Teaching*, *13*, 76-100.

Kiang, M.Y., & Kumar, A. (2001). An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research*, *12*(2), 177-194.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological*

Cybernetics, *43*, 59–69.

Kohonen, T. (2014). *MATLAB Implementations and Applications of the Self-Organizing Map*. Unigrafia Oy.

Kozan, S. (2016). *Madde sınıflamada faktör analizi, çok boyutlu ölçekleme ve kümeleme analizlerinin karşılaştırılması [Comparison of factor analysis, multidimensional scaling and cluster analysis in item classification]* [Unpublished master's thesis]. Mersin University.

Krishnan, S., & Idris, N. (2018). Using partial credit model to improve the quality of an instrument. *International Journal of Evaluation and Research in Education*, *7*(4), 313-316.

Li, C-H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936–949. https://doi.org/10.3758/s13428-015-0619-7

Linacre, J.M. (2012). *Many-facet Rasch measurement: Facets tutorial*. http://www.winsteps.com/a/ftutorial2.pdf

Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. http://www.winsteps.com/a/facets-manual.pdf

Miljković, D. (2017). Brief review of self-organizing maps. *MIPRO/CTS*, 1252-1257

Murtagh, F., & Hernandez-Pajares, M. (1995). The Kohonen self-organizing map method: An assessment. *Journal of Classification*, *12*(2), 165-190

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd. ed.). McGraw-Hill.

Ostini, R., & Nering, M.L. (2006). *Polytomous item response theory models* (3rd. Ed). Sage Publications.

Peixoto, E.M., Zanini, D.S., & Andrade J.M. (2021). Cross-cultural adaptation and psychometric properties of the Kessler distress scale (K10): an application of the rating scale model. *Psicologia: Reflexão e Crítica*, *34*(21), 2-10. https://doi.org/10.1186/s41155-021-00186-9

Sadesky, G.S. (2007). *Determining the structure in test performance* [Unpublished doctoral dissertation]. Alberta University.

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.

Sorrosal Forradellasa, M.T., Barberà-Mariné, M.G., Fernández Bariviera, A., & Garbajosa-Cabello, M.J. (2012). Advantages of using self-organizing maps to analyse student evaluations of teaching. *Fuzzy Economic Review*. https://doi.org/10.25102/fer.2012.01.03

Sözer, E., & Kahraman, N. (2021). Investigation of psychometric properties of Likert items with the same response categories using polytomous item response theory models. *Journal of Measurement and Evaluation in Education and Psychology*, *12*(2), 129-146. https://doi.org/10.21031/epod.819927

Stambuk, A., Stambuk, N., & Konjevoda P. (2007). Application of Kohonen self- organizing maps (SOM) based clustering for the assessment of religious motivation. 29th International Conference on Information Technology Interfaces, Croatia.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik [Reliability and validity in social and behavioral measurements] (1st ed.)*. Seçkin Yayınları.

Şimşek, D. (2006). *Kümeleme analizi, çok boyutlu ölçekleme, doğrulayıcı ve açıklayıcı faktör analizi ile elde edilen yapı geçerliği kanıtlarının karşılaştırılması. [A comprasion of construct validity evidences obtained through cluster analysis, multidimensional scaling, confirmatory, and exploratory factor analysis]* [Unpublished master's thesis]. Hacettepe University.

Takasaki, H., Kawazoe, S., Miki, T., Chiba H., & Godfrey, E. (2021). Development and validity assessment of a Japanese version of the exercise adherence rating scale in participants with musculoskeletal disorders. *Health and Quality Life Outcomes*, *19(169)*, 2-8. https://doi.org/10.1186/s12955-021-01804-x.

Tezbaşaran, E. (2016*). Temel bileşenler analizi ve yapay sinir ağı modellerinin ölçek geliştirme sürecinde kullanılabilirliğinin incelenmesi. [An investigation on usability of principal component analysis and artificial neural network models in the process of scale development]* [Unpublished doctoral dissertation]. Mersin University.

Uysal, M. (2015). *Araştırma özyeterlik ölçeğinin psikometrik özelliklerinin klasik test kuramı ve madde tepki kuramına göre incelenmesi. [An investigation of psychometric properties of research self-efficacy scale according to classical test theory and item response theory]* [Unpublished master's thesis] Gazi University.

Wood, J.M., Tataryn, D.J., & Gorsuch, R.L. (1996). Effects of under- and over extraction on principal axis factor analysis with varimax rotation. *Psyhological Methods*, *1*(4), 354-365.

APPENDIX

Appendix A. Descriptive Statistics for the Statistical Attitude Scale.

Item	Mean	Std. Deviation	Skewness	S.E. of Skewness	Kurtosis	S.E. of Kurtosis	Item	Mean	Std. Deviation	Skewness	S.E. of Skewness	Kurtosis	S.E. of Kurtosis
1	2.49	1.18	.30	.086	-.81	.172	21	3.15	1.38	-.18	.086	-1.18	.172
2	2.35	1.20	.50	.086	-.75	.172	22	2.34	1.17	.46	.086	-.76	.172
3	4.16	1.03	-1.13	.086	.73	.172	23	2.12	1.11	.70	.086	-.35	.172
4	3.62	1.28	-.62	.086	-.67	.172	24	2.73	1.22	.13	.086	-.96	.172
5	2.85	1.28	.04	.086	-1.07	.172	25	2.36	1.22	.51	.086	-.75	.172
6	2.20	1.15	.58	.086	-.64	.172	26	2.13	1.14	.73	.086	-.36	.172
7	3.88	1.25	-.94	.086	-.16	.172	27	3.44	1.32	-.46	.086	-.88	.172
8	3.17	1.33	-.22	.086	-1.08	.172	28	3.35	1.28	-.37	.086	-.91	.172
9	3.37	1.33	-.38	.086	-1.00	.172	29	3.26	1.37	-.29	.086	-1.11	.172
10	3.12	1.24	-.17	.086	-.94	.172	30	2.55	1.18	.027	.086	-.86	.172
11	2.23	1.13	.57	.086	-.58	.172	31	3.59	1.37	-.60	.086	-.90	.172
12	3.10	1.32	-.15	.086	-1.06	.172	32	3.51	1.35	-.54	.086	-.87	.172
13	3.70	1.33	-.74	.086	-.64	.172	33	3.80	1.21	-.80	.086	-.21	.172
14	3.47	1.33	-.53	.086	-.85	.172	34	3.84	1.16	-.89	.086	.09	.172
15	3.44	1.34	-.43	.086	-.97	.172	35	3.63	1.27	-.65	.086	-.61	.172
16	3.53	1.22	-.40	.086	-.80	.172	36	3.30	1.21	-.29	.086	-.77	.172
17	2.51	1.12	.32	.086	-.72	.172	37	2.25	1.12	.54	.086	-.60	.172
18	2.92	1.24	.00	.086	-1.00	.172	38	3.91	1.07	-.94	.086	.39	.172
19	3.26	1.42	-.33	.086	-1.16	.172	39	2.18	1.16	.64	.086	-.56	.172
20	3.28	1.30	-.39	.086	-.91	.172	40	3.50	1.28	-.52	.086	-.76	.172