Investigation of the factor structure of the Turkish version of the State-Trait Anxiety Inventory

Durumluk-Sürekli Kaygı Ölçeği'nin Türkçe versiyonunun faktör yapısının değerlendirilmesi

Abstract

Aim: Confirmatory factor analysis (CFA) and Rasch Analysis are commonly used methods to examine the structure of the psychological scales. In this study, it is aimed to evaluate the factor structure Turkish version of the State-Trait Anxiety Inventory by using statistics based on the Rasch model and CFA.

Methods: The State-Trait Anxiety Inventory (STAI) was used for the analysis. Of the study group, 186 (46.5%) were male and 214 (53.5%) were female. Unidimensionality was investigated using a Rasch-based principal component analysis (PCA) of the residuals, chi-square tests, item fit statistics, and other statistics. CFA has also been applied to test the hypothesis of a one-factor solution.

Results: The item-trait interaction chi-square statistic was 342.344 for the state scale (p<0.001) and 381.247 for the trait scale (p<0.001). For the state scale, 16.00% of the t-tests for the PCA were significant at the 5% level, while 19.50% were significant for the trait scale. The fit residuals of items 4, 8, and 18 on the state scale were over the +2.5 threshold, while the fit residuals of items 23, 24, and 34 on the trait scale were above the +2.5 threshold. Similarly, the scale structure evaluated by CFA was conditioned to be inadequate goodness-of-fit.

Conclusion: This study found that neither the trait nor the state scale of the STAI met the unidimensionality assumption. Consequently, both the Rasch analysis and CFA have been verified as succeeding tools in assessing the scale sub-dimensions and determining whether the response items can be utilized for a total scale score.

Keywords: anxiety; confirmatory factor analysis; Rasch analysis; the State-Trait Anxiety Inventory; unidimensionality

Öz

Amaç: Doğrulayıcı faktör analizi (DFA) ve Rasch Analizi, psikolojik ölçeklerin yapısını incelemek için yaygın olarak kullanılan yöntemlerdir. Bu çalışmada, Durumluk-Sürekli Kaygı Envanteri'nin Türkçe versiyonunun faktör yapısının Rasch modeli ve DFA temelli istatistikler kullanılarak değerlendirilmesi amaçlanmıştır.

Yöntem: Analiz için Durumluk-Sürekli Kaygı Envanteri'nin kullanıldığı çalışmada, grubunun 186'sı (% 46,5) erkek, 214'ü (% 53,5) kadındı. Tek boyutluluk, rezidüellerin Rasch tabanlı temel bileşen analizi (TBA), ki-kare testleri, madde uyumu istatistikleri ve diğer istatistikler kullanılarak araştırılmıştır. Aynı zamanda ölçeğin tek faktörlü yapısına ait hipotezi test etmek için DFA uygulanmıştır.

Bulgular: Madde-özellik interaksiyon ki-kare istatistiği, durumluk kaygı ölçeği için 342.344 (p <0.001) ve sürekli kaygı ölçeği için 381.247 (p <0.001) idi. Durumluk kaygı ölçeği için yanıt kategorileri değerlendirildiğinde 7. ve 18. maddelerin kesim noktalarının düzensiz yerleştiği, sürekli kaygı ölçeği için kesim noktalarının yerleşiminde böyle bir düzensizlik olmadığı saptandı. Durumluk kaygı ölçeğindeki 4, 8 ve 18 numaralı maddelere ait uyum rezidüelleri +2,5 eşiğinin üzerindeyken, sürekli kaygı ölçeğindeki 23, 24 ve 34 numaralı maddelerin uyum rezidüelleri +2,5 eşiğinin üzerindeydi. Benzer şekilde DFA ile değerlendirilen ölçek yapısı da yetersiz uyumu göstermekteydi.

Sonuç: Bu çalışma ile Durumluk-Sürekli Kaygı Envanteri'nin, ne durumluk, ne de sürekli kaygı ölçeklerinin tek boyutluluk varsayımını karşılamadığı saptanmıştır. Sonuç olarak, hem Rasch analizi hem de DFA, ölçek alt boyutlarının değerlendirilmesinde ve yanıt maddelerinin toplam ölçek puanı için kullanılıp kullanılamayacağının belirlenmesinde kullanılabilecek önemli yöntemlerdir.

Anahtar sözcükler: anksiyete; doğrulayıcı faktör analizi; Durumluk-Sürekli Kaygı Envanteri; Rasch analizi; tek boyutluluk

Leman Tomak[1], Mustafa Erhan Sari[2], Sule Cavus[2], Hatice Zehra Bodur[2]

[1] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Ondokuz Mayis University

[2] Department of Pediatric Dentistry, Faculty of Dentistry, Ondokuz Mayis University

Received/*Geliş* : 02.03.2021
Accepted/*Kabul*: 13.08.2021

DOI: 10.21673/anadoluklin.889735

Corresponding author/*Yazışma yazarı*

Leman Tomak

Ondokuz Mayıs Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Samsun, Turkey
E-mail: lemant@omu.edu.tr

ORCID
Leman Tomak: 0000-0002-8561-6706
Mustafa Erhan Sari: 0000-0001-7497-4930
Sule Cavus: 0000-0003-4514-3796
Hatice Zehra Bodur: 0000-0001-7416-3177

INTRODUCTION

Both the classical and modern measurement models that are used as a foundation for the development of measurement tools and the assessment of the validity of those tools generally focus on item analysis instead of what and how the measured characteristic is; they are founded on the basis that the measured characteristic is unidimensional (1).

To determine the construct validity of assessment instruments, confirmatory factor analysis (CFA) is used to establish whether a particular scale is uni- or multidimensional (2-4). CFA tests the fit between the identified factors using explanatory factor analysis (EFA) and the factors that have been determined from the study's hypotheses (5). This method generates latent variables from those variables observed using a previously formed model (6,7).

In addition to establishing the underlying dimensionality of a particular scale, another concern is whether the subdimensions of the scale meet modern psychometric standards (8). In assessing these standards, the one-parameter Rasch model has taken on increasing significance in scale validation studies (9). The Rasch model assesses the probability of answering each item on the scale as a function of the underlying dimension's ordering. If all the items fit the Rasch model, the items can be considered ordered and, thus, can be added along one dimension. This can be interpreted as an actual interval scale (10). The Rasch model presents a valid and objective measurement approach for assessing interval scale measurements and determining whether item responses meet model expectations. When these conditions are met, the process is considered suitable for evaluation (8).

Scales are assessment instruments consisting of items that aim to measure those variables that cannot be directly observed but that theoretically exist (11). The State-Trait Anxiety Inventory (STAI), which is a widespread tool for determining anxiety, was developed by Spielberg et al. in 1970 and adapted into Turkish by Oner and Le Compte (12,13); it is a form of self-assessment that indicates to what extent individuals use cognitive strategies in dealing with stress. It measures two different types of anxiety: the state scale measures how an individual feels at a specific moment and under specific conditions, while the trait scale measures how an individual feels in general, independent of the situation or condition they are in (13).

The current study intends to assess the unidimensionality of the STAI using Rasch analysis and CFA. When using the Rasch model, the current study will illustrate the importance of principal component analysis (PCA) when it comes to the residuals and other Rasch analysis statistics in ensuring the unidimensionality of the STAI scales. Similarly, using CFA, the model fit indexes will be approved, and the factor structure will be calculated.

MATERIAL AND METHODS

Participants

Students from the Faculty of Dentistry at a Turkish university participated in the current study. In total, 186 of the respondents (46.5%) were male, and 214 (53.5%) were female, with an average age of 20.85 ± 1.84 years. The institutional Ethics Committee approved the study, and written informed consent was obtained from all individual participants (decision number 2017/182).

Scale and procedure

The STAI consists of two scales of 20 items each. Items 1–20 measure situational or state anxiety (STAI-S), and items 21–40 measure underlying or trait anxiety (STAI-T).

Analysis of the validity and reliability of the scales was conducted using classical methods. Although the state scale requires an individual to describe how s/he feels at a specific moment and under specific conditions, the trait scale requires an individual to describe how s/he feels in general, independent of the state or condition s/he is in (13).

On the state scale, individuals were asked to choose one option with the options of “not at all,” “somewhat,” “moderately so,” and “very much so,” here following the intensity of this person's feelings, thoughts, or behaviors as expressed by the individual items. On the state scale, the individuals were asked to choose among the options of “almost never,” “sometimes,” “often,” and “almost always.” The scales have both direct and reversed items. When scoring the reversed items (i.e., those that express positive feelings), responses with a value of 1 are transformed into 4, while those

Table 1. Item fit statistics for the state scale

		Item statistics		Fit statistics	
No	**Item**	**Location**	**SE[a]**	**Residual**	**χ^2**
1	**I feel calm**	0.263	0.073	1.341	12.984
2	**I feel secure**	0.029	0.071	1.28	9.553
3	**I am tense**	0.936	0.076	1.754	17.377
4	**I feel strained**	0.77	0.067	4.369	20.049
5	**I feel at ease**	-0.662	0.067	1.645	8.106
6	**I feel upset**	0.16	0.065	-0.809	9.129
7	**I am presently worrying over possible misfortunes**	0.304	0.065	1.363	16.288
8	**I feel satisfied**	-1.469	0.069	2.872	30.551[b]
9	**I feel frightened**	0.446	0.072	-1.486	12.3
10	**I feel comfortable**	-0.621	0.069	-0.644	17.499
11	**I feel self-confident**	0.324	0.072	2.332	20.294
12	**I feel nervous**	0.678	0.066	-2.621	15.503
13	**I am jittery**	1.016	0.071	-1.234	23.287[b]
14	**I feel indecisive**	0.882	0.069	-1.25	10.345
15	**I am relaxed**	-1.257	0.073	-0.713	6.225
16	**I feel content**	-0.858	0.071	-0.301	4.896
17	**I am worried**	0.614	0.072	-0.433	1.224
18	**I feel confused**	1.168	0.074	4.162	67.478[b]
19	**I feel steady**	-1.622	0.075	-0.42	17.08
20	**I feel pleasant**	-1.102	0.072	-2.163	22.175[b]

[a] SE: Standard error

[b] Fit statistics with statistically significant χ^2 value

Table 2. The statistics of model fit

Fit statistics*	**Observed model fit**		**Criteria**
	The trait scale	**The state scale**	
χ^2/df	9.059	6.011	< 3
GFI	0.679	0.788	≥ 0.90
AGFI	0.599	0.731	≥ 0.85
CFI	0.613	0.636	≥ 0.95
RMSEA	0.142	0.112	<0.80

* χ^2/df (Chi-square/df); the Goodness-of-Fit Index (GFI), the Adjusted Goodness-of-Fit Index (AGFI), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA)

with a value of 4 are transformed into 1. There are 10 reversed items on the state scale (items 1, 2, 5, 8, 10, 11, 15, 16, 19, and 20), and seven on the trait scale (items 21, 26, 27, 30, 33, 36, and 39). The total possible score for each scale ranges between 20 and 80, with a higher score expressing a higher level of anxiety (12,13).

Statistical analysis

In the present study, the underlying dimension structure of both the trait and state scale of STAI was evaluated. Based on the fact that both subscales measure a single underlying dimension, this assumption was tested using CFA and Rasch analysis.

CFA is a process that evaluates the latent structure obtained through the model created from the variables observed with EFA (2,3). CFA is factor analysis used to test the compliance of the factors determined by EFA, with the factor structures being determined by the hypothesis. CFA was used to determine whether the variable groups contributing to a specified number of factors were adequately represented by these factors. The model may be determined by the researcher in theory and tested with CFA, or it could be a model obtained as a result of EFA (4,5). The first step in CFA is to determine the model. In the confirmatory factor model, the number of common factors and observed

Table 3. Threshold values for the State-Trait Anxiety Inventory

Items of the state scale	Response category threshold locations			Items of the trait scale	Response category threshold locations		
	0-1	1-2	2-3		0-1	1-2	2-3
Item 1	-1.625	-1.024	2.648	**Item 21**	-2.616	-0.025	2.641
Item 2	-1.721	-0.071	1.792	**Item 22**	-2.169	0.780	1.389
Item 3	-1.551	0.480	1.071	**Item 23**	-0.641	-0.163	0.804
Item 4	-0.771	0.065	0.705	**Item 24**	-1.236	0.246	0.990
Item 5	-1.016	-0.463	1.479	**Item 25**	-2.128	0.624	1.505
Item 6	-1.257	0.348	0.908	**Item 26**	-1.932	0.187	1.745
Item 7	-1.144	0.694	0.450	**Item 27**	-1.650	0.264	1.385
Item 8	-1.220	0.219	1.000	**Item 28**	-2.124	0.911	1.212
Item 9	-1.713	0.433	1.280	**Item 29**	-1.856	0.095	1.761
Item 10	-1.497	-0.190	1.687	**Item 30**	-2.345	0.063	2.282
Item 11	-1.802	-0.208	2.010	**Item 31**	-1.510	0.227	1.283
Item 12	-0.864	0.148	0.716	**Item 32**	-1.272	0.052	1.220
Item 13	-0.403	0.128	0.275	**Item 33**	-2.308	0.021	2.287
Item 14	-0.633	0.256	0.377	**Item 34**	-1.591	-0.039	1.630
Item 15	-1.448	-0.223	1.671	**Item 35**	-2.096	0.773	1.323
Item 16	-1.504	-0.224	1.727	**Item 36**	-2.050	0.187	1.864
Item 17	-1.504	0.508	0.996	**Item 37**	-1.738	0.116	1.622
Item 18	0.036	-0.182	0.146	**Item 38**	-1.343	0.482	0.861
Item 19	-1.639	0.017	1.622	**Item 39**	-1.880	0.271	1.609
Item 20	-1.819	0.055	1.764	**Item 40**	-1.522	0.039	1.483

Table 4. Item fit statistics for the trait scale

No	Item	Item Statistics		Fit Statistics	
		Location	SE [a]	Residual	χ^2
21	I generally feel pleasant	-0.038	0.082	-1.381	20.285
22	I generally get tired easily	-0.48	0.069	0.908	4.568
23	I generally cry easily	0.389	0.06	4.496	77.01 [b]
24	I want to be as happy as others	-0.093	0.062	2.837	25.087 [b]
25	I miss opportunities since I can't decide quickly	0.252	0.075	-0.756	7.986
26	I feel rested	-1.63	0.074	2.123	23.554 [b]
27	I am calm, cool and collected	-0.275	0.066	0.402	10.684
28	I feel that difficulties are piling up so that I cannot overcome them	0.696	0.081	-1.102	13.515
29	I worry too much over something that really doesn't matter	0.109	0.071	-1.807	19.365
30	I am happy	-0.011	0.077	-0.54	14.665
31	I take everything seriously and worry	0.121	0.066	-1.648	19.298
32	I lack self confidence	1.053	0.072	-0.785	12.395
33	I feel secure	-0.434	0.077	0.486	21.206
34	I avoid troublesome and difficult situations	0.102	0.068	3.687	39.201 [b]
35	I feel sad	0.366	0.076	-1.65	26.111 [b]
36	I am content	0.065	0.073	0.969	16.355
37	Some unimportant thoughts run through my mind and bother me	-0.178	0.069	-0.962	8.941
38	I take disappointments so keenly that I can't put them out of my mind	-0.057	0.062	-1.206	6.483
39	I am a steady and decisive person	0.259	0.072	1.05	11.109
40	I feel uneasy as I think over my recent concerns	-0.214	0.066	0.519	3.43

[a] SE: Standard error

[b] Fit statistics with statistically significant χ^2 value

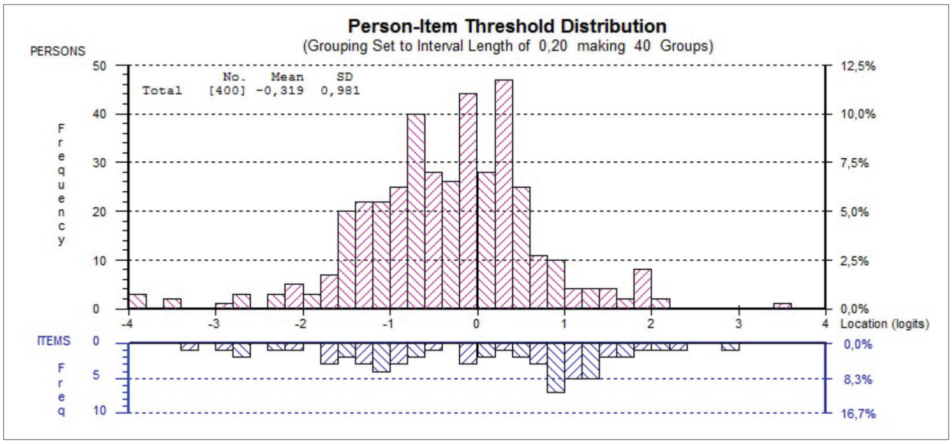

Figure 1. Person item threshold map for the state scale

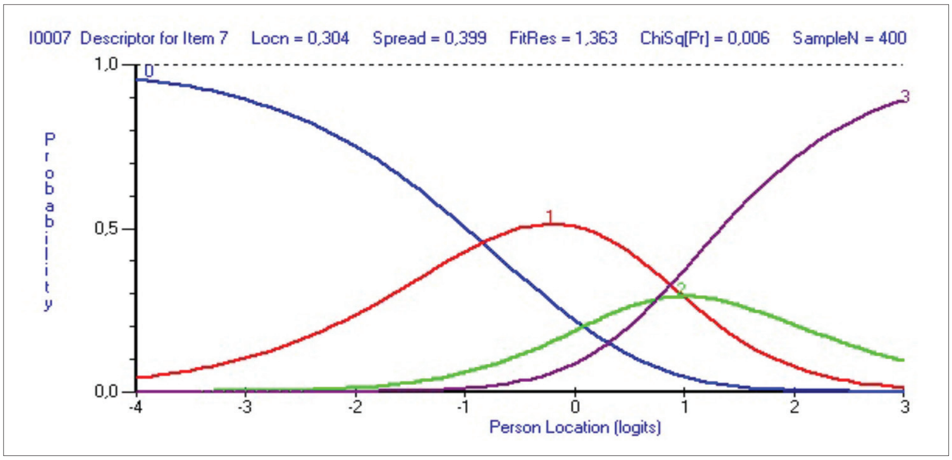

Figure 2. Category probability curve for item 7 with disordered thresholds

variables, the relationship between variance and covariance between specific factors, the relationship between common factors, and the relationships between the observed and common factors should be specified (3,5). There are various statistics for evaluating model fit goodness. The most commonly used statistics are chi-square statistics, the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) (5,6).

The unidimensionality assumption of the STAI scales was evaluated with chi-square statistics and several indices. For an acceptable fit, $\chi 2/df$ should be <3, CFI ≥ 0.95, GFI ≥ 0.90, AGFI ≥ 0.85, and RMSEA <0.80. We also examined factor loadings, and those ≥0.40 were considered acceptable (3-6).

The Rasch model includes only a difficulty parameter (14). In this model, an individual's probability of answering an item correctly is defined as the function of the ratio of that individual's ability level to item difficulty above the many underlying dimensions (15,16). A common variation of the Rasch model is the partial credit model, which is unidimensional and used for Likert scale items; in this model, thresholds differ from item to item (17,18). The model can be summarized using the following equation:

$$\ln\left(\frac{P_{nik}}{1-P_{nik}}\right)=B_n-D_{ik} \quad (1)$$

where ln is the natural log, P_{nik} is the probability of an individual n endorsing item i and threshold k, B_n is the individual's level of ability, and D_{ik} is the level of difficulty, here as expressed as the threshold k of item i (19).

To assess the fit of the proposed structure to the Rasch model, item fit statistics were calculated (9). The fit of the model as a whole was assessed using overall fit statistics. A statistically insignificant p-value for

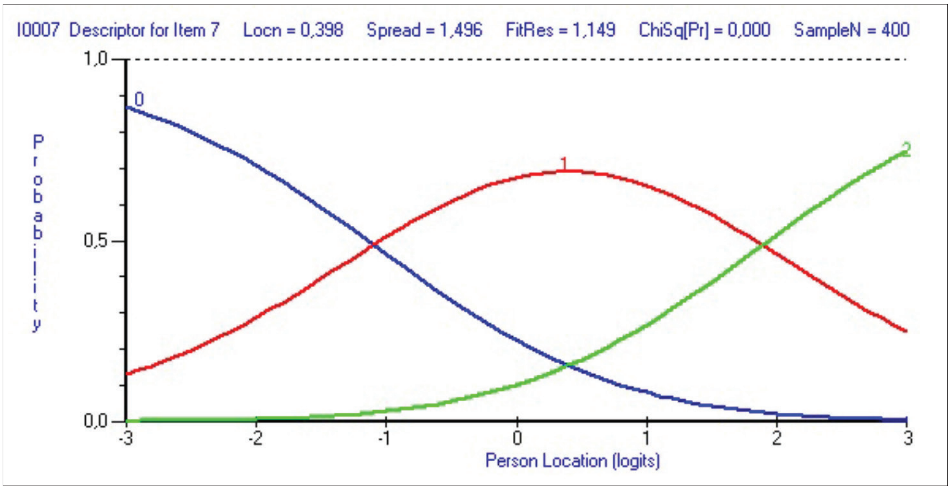

Figure 3. Category probability curve for item 7 with ordered thresholds

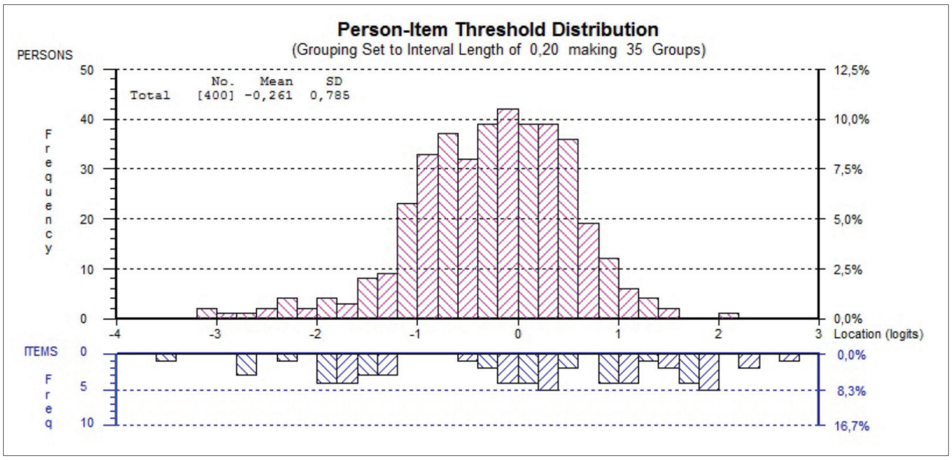

Figure 4. Person item threshold map for the trait scale

chi-square tests was used to indicate the model fit. The person separation index (PSI), which assesses the internal consistency of the item bank and is interpreted in terms of Cronbach's alpha, was calculated (10). Chi-square statistics and standardized residuals were analyzed to assess the fit of each item on the scales. Here, the statistically insignificant chi-square value and the residuals are between ± 2.5 indicates model fit. Positive residuals greater than +2.5 indicate that the unidimensionality assumption is violated, and negative residuals less than -2.5 indicate local dependence (9).

In some situations, fit statistics are not sufficient for determining whether a scale is uni- or multidimensional. Thus, positive and negative loaded residuals were found using PCA, and these two subgroups were compared using independent t-tests. The criterion to decide whether the unidimensionality assumption has been met is for the t-test to have more than 5% of values fall between -1.96 and 1.96. If the two groups are different at a level of 0.05, the unidimensionality assumption is not met (14). To determine whether the factors defined by these scales were unidimensional, the fit of the Rasch model was assessed (10).

In the current study, whether the response levels for each item on the scales were ordered effectively was assessed. A person with a higher level of ability for a particular dimension will have a higher probability of answering an item correctly (14). If all of the items fit the Rasch model, all the items will be ordered along the underlying dimension in a gradually increasing order. If the item responses are disordered, this indicates that there are too many response categories, that the category labels overlap, or that the scale is multidimensional (20).

When individuals in different groups who have the same level of ability have different probabilities of answering an item correctly, it can be said that this item has a difference in functioning (9). Differential item functioning (DIF) is an indicator of the interaction of the items on a scale with the characteristics of particular subgroups. The presence of DIF indicates bias or multidimensionality in a scale (21). To test for this phenomenon, the current study determined whether the responses given to the items differed between female and male students.

RUMM 2030 was used for the validity and reliability analyses of the scales in the STAI (22). CFA was implemented using AMOS Version 25 (23).

RESULTS

The scales on the STAI were analyzed using the Rasch model to assess their dimensionality. For the state scale, the item and person average fit residuals were 0.452 ± 1.999 and 0.349 ± 1.630, respectively. The item–trait interaction chi-square statistic (χ^2, 342.344; p<0.001) indicated a lack of model fit. The PSI was 0.89. The item and fit statistics for the items on the scale are presented in Table 1. Six of the items were found not to fit the model (items 4, 8, 12, 13, 18, and 20). Of these, items 4, 8, and 18 had high positive residuals, indicating that the unidimensionality assumption had been violated. Item 12 has a negative high residual, and this shows local dependency. The chi-square statistics were significant for items 8, 13, 18, and 20.

Of the 400 t-test comparisons, 64 (16.00%) were significant. To test for unidimensionality, the person estimates for the two highest positive-loading items on the first residual component were compared with the person estimates derived from the highest negative-loading items, with both sets calibrated on the same metric. Similarly, the scale structure evaluated by CFA was found to be insufficient for goodness of fit (Table 2).

When comparing the responses by gender, DIF was found for items 10, 11, and 16. A person–item location distribution map was used to assess the targeting of items and persons. In this map, person locations are plotted together with item locations or item threshold locations on the same continuum. The distribution of person and item threshold locations for the state scale are shown in Figure 1.

The thresholds for the response categories for the items on the scale are given as a table (Table 3). When these response categories were examined, disordered thresholds were found for items 7 and 18.

Three categories were obtained by combining the overlapping categories for items 7 and 18 as 012, after which it was found that the thresholds fit the logical order. Probability graphs for item 7, displaying disordered and ordered thresholds, are presented in both figures (Figure 2 and Figure 3). After the thresholds were fixed, the fit statistics for item 18 decreased to 2.761, while no significant improvement was found for the other fit statistics.

For the trait scale, the item and person average fit residuals were 0.282 ± 1.835 and 0.346 ± 1.652, respectively. The item–interaction chi-square statistic (χ^2, 381.247; p<0.001) again indicated that the model did not fit. The PSI was 0.84. Table 4 summarizes the item and fit statistics for the items on the scale. The fit residuals for items 23, 24, and 34 were above the threshold of +2.5 and, thus, the assumption of unidimensionality could not be supported. The chi-square statistics were significant for items 23, 24, 26, 34, and 35.

When positive and negative loaded residuals for the first primary component were found using PCA and these two subgroups were compared with independent t-tests, 78 (19.50%) of the 400 t-test comparisons were significant at the 5% level. Model fit indexes obtained using CFA did not meet the required criteria (Table 2).

In the comparison of responses by gender, DIF was observed for items 3, 7, 8, 10, 11, 12, 13, 15, 16, and 18. The distribution of the person and item threshold locations for the trait scale is shown in Figure 4.

When the item response categories were examined, no disordered thresholds were found (Table 3).

DISCUSSION AND CONCLUSION

The current study assessed whether the two STAI scales (the state scale and trait scale) had a unidimensional structure. The presence of only one latent dimension that accounts for the common variance within the data indicates unidimensionality (24-26). Unidimensionality is important for measurement scales because it is the basic assumption on which the valid calculation of total scores rests (27,28). For unambiguous interpretation to be possible, it is necessary for the total score of a particular scale to represent a single defining characteristic. In other words, a scale used to measure a particular characteristic should not be affected by other characteristics, and scale scores should represent the overall structure underlying the characteristic (29). The scale's unidimensionality theorem was appraised using both CFA and Rasch analysis.

To determine whether the items on the two STAI scales could be combined to generate an overall score, Rasch-based PCA of the residuals was conducted to test for unidimensionality (26,30). In the t-test comparison of the positive and negative loaded residuals generated from the PCA, the ratio of difference was over 5% for both scales. In these t-tests, any result above 5% violates the assumption of unidimensionality (9,24). As a result, neither of the two STAI scales could be regarded as unidimensional. Correspondingly, the model fit statistics evaluated using CFA do not satisfy the obligations of the required criteria; accordingly, the hypothesis of unidimensionality has been violated (3,6).

Fit statistics were also used to analyze the dimensionality of the STAI scales. Average item and person fit residuals were calculated; for an accurate model fit, these residuals should be close to 0, and their standard deviation should be close to 1. Deviation from these expectations indicates a lack of model fit (29,31). The average fit residuals and standard deviations in this study were significantly different from 0 and 1, respectively. The overall chi-square statistic shows the interaction of item characteristics; the individual item is the sum of the chi-square statistics. Insignificant chi-square tests indicate measurement invariance within the item bank with varying levels of measurement structure (31,32). However, the present study found that the chi-square statistics were significant for both scales, thus violating the unidimensionality assumption. Kaipper et al. also assessed the STAI using Rasch analysis, finding that the overall chi-square statistics for both scales were significant (19).

The PSI, which indicates the scale's ability to differentiate between persons, was used to assess the reliability of the scale (14). For both scales, the PSI was over 0.85. Typically, PSI scores over 0.70 and 0.80 represent good reliability (33,34). In Kaipper et al.'s study, the PSI was also over 0.85 for both scales (19).

The internal construct validity of each scale was assessed using fit residuals. In both scales, residuals higher than +2.5 were found. At the same time, there were items in both scales that had significant chi-square values. Both of these results indicate that the unidimensionality assumption was not met (29,35). When developing a scale, items should not be omitted randomly. However, items can be omitted to fit the scale to the Rasch model (14).

The DIF measure represents the difference in the probability of individuals with the same levels of ability but who are in different groups responding correctly to an item. Thus, it assesses the variation between individuals with the same test scores and the same level of ability but who are in different groups in terms of the probability of answering a specific test item (9,28). In the current study, evidence was found for the presence of DIF in terms of gender for both scales. DIF occurs when subgroups give different responses to items, even though the underlying structure is the same, making comparisons between groups impossible. Rasch analysis can be used to evaluate whether items function similarly between subgroups (30,36,37).

The presence of disorder observed among the state scale response category thresholds is an indicator that the response scale did not operate as expected. This may be because of problems in making finely tuned ratings or ambiguous distinctions between categories (37). Disordered thresholds suggest that there is a problem in the interaction between the respondents, items, and response options; hence, the clinical meaning of the response scale is rendered unclear (30,35).

In the current study, disorder occurred at the thresholds of two items on the state scale. Consecutive categories for disordered thresholds can be combined to improve the model fit (9,10). Therefore, in the present study, two neighboring categories were combined for the two affected items on the state scale, and the response levels were renamed. Following this modification to the scale, item fit was reassessed, and insufficient improvement was found but only in the fit statistics of an item. Although problems with rating scale response categories may be associated with multidimensionality, this is probably not the correct explanation here because no improvement was observed in the model fit following the explorative post hoc combination of response categories (29).

It can be concluded that issues with dimensioning occurred because the items did not function together to define the latent variable. In Kaipper et al.'s study, item statistics that did not fit the model and disordered thresholds were also observed; subsequently, seven items were omitted from the state scale, and eight were omitted from the trait scale. When the scales were reanalyzed, it was found that all of the Rasch analysis statistics met the unidimensionality assumption (19,35).

The item–person map of the scales in our study illustrated that although the items were reasonably well distributed, some individuals cannot be measured as reliably as most of the other respondents when using this set of items. The reason for this is that the items were either too intense or not intense enough for these individuals. The item–person map presents the difficulty levels of the STAI items concerning the ability measures of the sample on the same measurement continuum. By analyzing the difference between the average person and item measures, targeting can be assessed (29). A difference of 0 indicates perfect targeting. As the difference between the person and item average measures increases, items become more mistargeted to the sample (28).

In the current study, the structure of the scale was examined by Rasch analysis and CFA, and it was explored whether it provided the one-dimensional conjecture. Rasch-based PCA, item fit analysis, and other analyzes were used to assess the unidimensional structure of the STAI scales. Evidence of unidimensionality is not achieved for the original state and trait scales. To confirm these observations, CFA was conducted by investigating the dimensionality of the STAI, and the results were obtained in the same way. However, researchers and clinicians must be cautious in the use and interpretation of the STAI scales. Further studies should be conducted to determine if a reduction and/or regrouping of items on the two scales can produce more valid and interpretable measures.

Consequently, it is critical to use CFA together with modern psychometric methods such as Rasch analysis in the evaluation of scale structure. Statistics obtained with this method allow the confirmation of the scale structure and the pruning of the scale items without losing data by omitting duplicated items or items that do not fit the underlying structure.

Conflict-of-interest and financial disclosure

The authors declare that they have no conflict of interest to disclose. The authors also declare that they did not receive any financial support for the study.

Editing Service

This manuscript is edited and revised for spelling, grammar, clarity, consistency, and correctness by Scribendi Inc. (ESL Academic Editing Service).

REFERENCES

1. Reckase M. Multidimensional item response theory: Statistics for Social and Behavioral Sciences. New York, US: Springer; 2009.
2. Brown TA. Confirmatory factor analysis for applied research. New York, US: Guilford Publications; 2014.
3. Brown TA, Moore MT. Confirmatory factor analysis. In: Hoyle RH, editor. Handbook of structural equation modeling. New York, US: Guilford Publications, 2012; p. 361-79.
4. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR, eds. Reading and Understanding Multivariate Statistics. Washington DC, US: American Psychological Association, 1995; p. 99-136.
5. Thompson B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington DC, US: American Psychological Association; 2004.
6. Harrington D. Confirmatory factor analysis. New York, US: Oxford University Press; 2009.
7. Medvedev ON, Landhuis CE. Exploring constructs of well-being, happiness and quality of life. PeerJ. 2018;1(6):e4903.
8. Siegert RJ, Jackson DM, Tennant A, Turner-Stokes L. Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research. J Rehabil Med. 2010;42(4):302-9.
9. Forkmann T, Gauggel S, Spangenberg L, Brähler E, Glaesmer H. Dimensional assessment of depressive severity in the elderly general population: psychometric evaluation of the PHQ-9 using Rasch Analysis. J Affect Disord. 2013;148(2):323-30.
10. Wardenaar KJ, van Veen T, Giltay EJ, den Hollander-Gijsman M, Penninx BW, Zitman FG. The structure and

dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorders and healthy controls. J Affect Disord. 2010;125(1):146-54.
11. DeVellis RF. Classical test theory. Med Care. 2006;44(11):50-S9.
12. Spielberger CD. State-Trait anxiety inventory. Wiley Online Library; 2010.
13. Oner N, Le Compte A. Durumluk-Surekli kaygı envanteri el kitabı. Istanbul: Boğaziçi Yayınları; 1985.
14. Franchignoni F, Mora G, Giordano A, Volanti P, Chiò A. Evidence of multidimensionality in the ALSFRS-R Scale: a critical appraisal on its measurement properties using Rasch analysis. J Neurol Neurosurg Psychiatry. 2013;84(12):1340-5.
15. Wilson M. Constructing measures: An item response modeling approach. Mahwah, J: Lawrence Erlbaum Associates; 2005.
16. Brodin U, Fors U, Laksov KB. The application of Item Response Theory on a teaching strategy profile questionnaire. BMC Med Educ. 2010;10(1):14.
17. Zheng X, Rabe-Hesketh S. Estimating parameters of dichotomous and ordinal item response models with gllamm. Stata J. 2007;7(3):313-33.
18. Allen DD. Validity and reliability of the movement ability measure: a self-report instrument proposed for assessing movement across diagnoses and ability levels. Phys Ther. 2007;87(7):899-916.
19. Kaipper MB, Chachamovich E, Hidalgo MPL, da Silva Torres IL, Caumo W. Evaluation of the structure of Brazilian State-Trait Anxiety Inventory using a Rasch psychometric approach. J Psychosom Res. 2010;68(3):223-33.
20. Hagell P, Nilsson MH. The 39-item Parkinson's Disease Questionnaire (PDQ-39): Is it a unidimensional construct? Ther Adv Neurol Disord. 2009;2(4):205-14.
21. Tennant A, Pallant J. Unidimensionality matters!(A tale of two Smiths?). Rasch measurement transactions 2006;20(1):1048-51.
22. Andrich D, Sheridan B, Luo G. RUMM 2030 Version 5.4 for windows. RUMM Laboratory Pty Ltd. 2012:21.
23. IBM SPSS Amos Campus Edition Academic Authorized, Version 25.0., IBM Corp., Armonk, NY; 2012.
24. Ashley L, Smith AB, Keding A, Jones H, Velikova G, Wright P. Psychometric evaluation of the Revised Illness Perception Questionnaire (IPQ-R) in cancer patients: Confirmatory factor analysis and Rasch analysis. J Psychosom Res. 2013;75(6):556-62.
25. Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. Arch Phys Med Rehabil. 2010;91(9):1370-7.
26. GORTON III GE, Stout JL, Bagley AM, Bevans K, Novacheck TF, Tucker CA. Gillette Functional Assessment Questionnaire 22-item skill set: factor and Rasch analyses. Dev Med Child Neurol. 2011;53(3):250-5.
27. Delgado AR., Prieto G., Burin DI. Constructing three emotion knowledge tests from the invariant measurement approach. PeerJ. 2017;13(5):e3755.
28. Aplin T, Chien CW, Gustafsson L. Initial validation of the dimensions of home measure. Aust Occup Ther J. 2016;63(1):47-56.
29. Hendriks J, Fyfe S, Styles I, Skinner SR, Merriman G. Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. Australas Med J. 2012;5(5):251-261.
30. Smith AB, Fallowfield LJ, Stark DP, Velikova G, Jenkins V. A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ)-12. Health Qual Life Outcomes 2010;8(1):45.
31. Lin C-Y, Broström A, Nilsen P, Griffiths MD, Pakpour AH. Psychometric validation of the Bergen Social Media Addiction Scale using classic test theory and Rasch models. Journal of Behavioral Addictions. 2017;6(4):620-9.
32. Smith Jr EV. Understanding Rasch measurement: Detecting and evaluating the impact of multidimenstionality using item fit statistics and principal component analysis of residuals. J Appl Meas. 2002;3(2):205-31.
33. Hsiao Y-Y, Shih C-L, Yu W-H, Hsieh C-H, Hsieh C-L. Examining unidimensionality and improving reliability for the eight subscales of the SF-36 in opioid-dependent patients using Rasch analysis. Qual Life Res. 2015;24(2):279-85.
34. Lin C-Y, Yang S-C, Lai W-W, Su W-C, Wang J-D. Rasch models suggested the satisfactory psychometric properties of the World Health Organization Quality of Life-Brief among lung cancer patients. Journal of Health Psychology 2017:22(4):397-408.
35. Chang K-C, Wang J-D, Tang H-P, Cheng C-M, Lin C-Y. Psychometric evaluation using Rasch analysis of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. Health and Quality of Life Outcomes 2014;3 (12):148.
36. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol. 2007;46(1):1-18.
37. Tomak L, Midik O. Primary principles in developing scale with Rasch analysis: Portfolio anxiety assessment. Niger J Clin Pract. 2018;21(10):1296-303.