

Investigation of Measurement Invariance of State Test Anxiety Scale

Huseyin Selvi ^{1,*}

¹Faculty of Medicine, Department of Medical Education, Mersin University, Mersin, Turkey

ARTICLE HISTORY

Received: Nov. 17, 2020

Revised: May 20, 2021

Accepted: May, 31, 2021

Keywords:

Test Anxiety,
Measurement Invariance,
Multi Group Confirmatory
Factor Analysis,
Test Validity

Abstract: In this study, it was aimed to examine the measurement invariance of State Test Anxiety Scale and its sub-dimensions developed by Şahin (2019) in terms of different variables. For this purpose, data were collected from a total of 956 university students studying in different faculties. The measurement invariance of the scale was examined by multi-group confirmatory factor analysis in terms of gender, faculty and socioeconomic level variables. In the study, the measurement model was established for 22 items and three components of the state anxiety test scale (cognitive, psychosocial and physiological) and tested for configural, metric, scalar and strict equivalence by considering the hierarchical principle in terms of gender, faculty and socioeconomic level variables. The findings showed that Configural equivalence was provided for all dimensions except the cognitive and physiological subscales for the socioeconomic status variable. On the other hand, metric equivalence was achieved in cognitive, psychosocial and physiological dimensions for the gender variable. Metric equivalence was achieved in Cognitive dimension for faculty variable. And for the socioeconomic status variable, it was provided only for the scale as a whole. Scalar and strict equivalence conditions were not met by any of the variables examined in the study.

1. INTRODUCTION

Test anxiety is one of the important variables affecting the academic success of individuals. For this reason, it is one of the subject area that psychology has emphasized since the 1950s. Test anxiety is a special form of anxiety and it can affect individuals of all ages in the society (Sieber, 1980). The exams, which are carried out for different purposes such as selection, placement, diagnosis and guidance, especially the exams with wide participation, affect the lives of individuals significantly today. Considering the meaning attributed to these exams and the potential of these exams to affect the lives of individuals, the dimension of anxiety experienced by individuals and their families can be better understood.

In the literature anxiety is defined as; fear of anticipation of something bad will happen, restlessness and feeling of loss of control (Sapir & Aranson, 1990), as fear and tension felt under threat (Büyüköztürk, 1997), as sadness and distress caused by stressful situations (Özgüven, 2007). As can be understood from the definitions, the concept of "anxiety"; It includes feelings of sadness, distress, fear, failure, helplessness and loss of control (Cüceloğlu,

*CONTACT: Huseyin Selvi ✉ hsyn_selvi@yahoo.com.tr 📍 Faculty of Medicine, Department of Medical Education, Mersin University, Mersin, Turkey

1998). Increasing the level of anxiety reveals the preventive role of anxiety. High test anxiety; it is an important problem that negatively affects the learning process and academic achievement of individuals (Ergene, 1994). As a matter of fact, individuals with high test anxiety may encounter situations like; easily distract, worry about performance, tension, restlessness, sadness, distress, fear, helplessness, loss of control, incompetence, silence, loss of control, less speech, withdrawal, palm sweating, hands trembling, increased heart rate and panic attack (Geen, 1985; Öner, 1990; Zeidner, 2007).

On the other hand, anxiety is generally perceived as a negative emotional state. However, it does not always affect the person negatively. This situation, which is described as the facilitating effect of anxiety, is referred to as "facilitating anxiety" in the literature. Facilitating anxiety; It emerges as a result of the person developing more motivation to cope with this situation and making more effort due to the increase in perception and awareness of the anxiety situation (Albert & Haber, 1960). Studies conducted in the literature on test anxiety show that very low and very high-level test anxiety affects learning negatively and a medium-level test anxiety affects learning positively (Hill & Wingfield, 1984; Bados, 2005; Gençdoğan, 2006). For this reason, keeping test anxiety under control is important for individuals' academic success, self-confidence and motivation.

Another important requirement of keeping test anxiety under control is that the quality of measurement tools used to measure success is affected by this situation. As a matter of fact, the most important requirement of a qualified measurement process is that the tests used in the exams can measure the variable to be measured without mixing it with other variables (Turgut & Baykul, 2012; Alici, 2013). A measurement process that can be performed in a qualified manner independent from the negativity of variables such as test anxiety, test technique and motivation will increase the accuracy of the decisions to be taken (Hill & Wigfield, 1984).

The review of the literature on this issue shows that many studies have been conducted to measure test anxiety and to reveal the reasons by examining it in terms of different variables. Most of these studies consist of scale development / scale adaptation studies for measuring test anxiety and studies that attempt to reveal the causes of anxiety (McDonald, 2001; Driscoll, 2007; Totan & Yavuz, 2009; Akın *et al.*, 2012; Başol, 2017; Aydın & Bulgan, 2017; Bozkurt *et al.*, 2017; Yao-Ting Sung & Tzu-Yang Chao, 2015; Şahin, 2019).

When the scales developed and adapted to measure test anxiety in Turkey are examined, it is seen that the evidence for the reliability and validity of almost all of these scales is collected. Using these scales, many studies have been conducted in the literature to reveal the reasons for test anxiety and to reveal the differences between different subgroups such as gender, education level, socioeconomic status, school type (Totan & Yavuz, 2009; Akın *et al.*, 2012; Başol, 2017; Aydın & Bulgan, 2017; Şahin, 2019). However, it is not correct to explain the differentiation of values obtained by using these scales between groups by only linking the characteristics of individuals. As a matter of fact, the differences between groups can be caused by the measurement tool rather than the individuals. Although there are studies on linguistic validity in scale adaptation studies, the adapted measurement tool may not measure the same structure in the language / culture it was adapted to (Cheug & Rensvold; 2002). The same situation also applies to the implementation to different groups of the scale was developed in Turkey. The developed scale may not measure the same structure for females and males in terms of gender variable and high school and university graduates in terms of education level variable. This is explained by the fact that psychological structures do not exactly overlap in different groups and cultures, in other words, the behaviors related to the structure can be different in different groups and cultures (van de Vijver & Poortinga, 2005).

In order for the scores obtained from the scale to be compared between groups, it must first be shown that the scale measures the same structure in different subgroups, in other words,

measurement invariance is achieved (van de Vijver & Poortinga, 2005). In measurement invariance studies, the items in the measurement tool in different groups; It is examined whether the factor loadings, correlation patterns, error variances are the same. Measurement invariance is a prerequisite for comparison studies between groups (Erkuş & Selvi, 2019; Cheung & Rensvold, 2000). When measurement invariance is not achieved, it means that the measurement tool does not measure the same thing in different subgroups. This causes the comparison studies to lose their meaning. In summary, it is not possible to make a comparison between groups without measurement invariance (van de Vijver & Poortinga, 2005).

Measurement invariance studies are conducted to reveal whether the factor structure obtained for the scale is the same in the sub-groups by using multi-group confirmatory factor analysis (Başusta, 2010). Multi-group confirmatory factor analysis is a method frequently used in structural equating modeling analysis, and it is performed to examine whether the model created by the researcher is the same in more than one group based on the data obtained from the same measurement tool (Tabachnick & Fidel, 2001).

Test anxiety can be compared in terms of variables such as gender, class level, socioeconomic status, age group, school, university, department, education level. But first of all, the measurement tool used must provide measurement invariance. Findings regarding differences between individuals and groups cannot be interpreted without providing measurement invariance (Horn & Mc Ardle, 1992). Measurement invariance studies consist of 4 phased steps, each of which is the prerequisite of the next, such as configural, metric, scalar and strict equivalence (Meredith, 1993; van de Vijver, 1998). Configural equivalence is the most basic level of measurement invariance, and it is the step where it is tested that the factor structure revealed for the measured psychological variable is the same in all groups, in other words, that the free and fixed factor patterns are similar between the groups. As a matter of fact, in order for the groups to be compared, it must be demonstrated that the relevant measurement tool measures the same thing in all groups. If the measured structure is different in the groups to be compared, in other words, if the related measurement tool measures different things in different groups, it is not meaningful to make comparisons between groups. Depending on this condition, it is clear that metric, scalar and strict equivalencies cannot be examined.

After the Configural equivalence is achieved, metric equivalence is examined. Although the measurement tool measures the same structure in different groups, it may not be able to measure individuals in different groups with the same latent structure at the same size. For this reason, the invariance of the units of the measuring tool in metric equivalence between groups examine. In other words, the equality of the units is tested. In order to test the equality of the units, it examines whether the factor loadings obtained for the scale items change or not between groups. When metric equivalence cannot be achieved, a situation arises where scale items are biased and cannot be summed. Therefore, it is not possible to examine scalar and strict invariance in cases where metric invariance cannot be achieved.

For scalar invariance, the equality of the origins of the measuring instrument between groups is tested. In other words, it is tested whether 5 points in one group equal 5 points in the other group. If 5 points in one group equals 7 points in the other group, then the origins are unequal and the measuring tool contains possible bias. For scalar invariance, it is sufficient to demonstrate that group means and factor loadings are equal. In strict equivalence, in addition to configural, metric and scalar invariance, it is tested whether the error and factor variances obtained for the scale items are equal between the groups. If configural, metric, scalar and strict invariance is provided for different groups, it can be interpreted that the scale measures the same structure, the same size and the same precision in these groups. And this makes it possible to compare the scores obtained from these different groups with the same measurement tool.

Regarding test anxiety, the scale used in order to make comparisons between groups should meet the measurement invariance conditions. In this context, the scales developed and adapted to measure test anxiety were examined and summarized in [Table 1](#).

Table 1. Summary of scales developed and adapted in Turkey.

Scale Name	Developed / Adapted	Measurement Invariance
State Test Anxiety Scale	Developed by Şahin (2019)	Unreported
Revised Test Anxiety Scale	Developed by Benson & El-Zahhar (1994), Adapted by Akin <i>et al</i> (2012)	Unreported
IDA Test Anxiety Scale,	Developed by Başol (2017)	Unreported
Children’s Test Anxiety Scale	Developed by Wren & Benson (2004), Adapted by Aydın & Bulgan (2017)	Unreported
Cognitive Test Anxiety Scale-Revised Form	Developed by Cassidy & Johnson (2002), Adapted by Bozkurt <i>et al</i> (2017)	Unreported
Friedben Test Anxiety Scale	Developed by Bados & Sanz (2005), Adapted by Akin <i>et al</i> (2013)	Unreported
Westside Test Anxiety Scale	Developed by Driscoll (2007), Adapted by Totan & Yavuz (2009)	Unreported

When [Table 1](#) is examined, no findings related to measurement invariance have been reported in any of the scales. Evidence regarding measurement invariance was not provided in any of the scales developed or adapted in Turkey to measure test anxiety.

Test anxiety is a variable that negatively affects individuals' learning process, academic achievement, and the quality of the measurement tool used, and it can affect individuals of almost all ages in the society. Therefore, studies conducted to reveal the reasons for test anxiety and to keep it under control are important for the literature.

For this reason in this study, it is aimed to provide a scale to the literature that provides the evidence for measurement invariance. In this study the measurement invariance of the State Test Anxiety Scale developed by Şahin (2019) in terms of variables of gender, socioeconomic level and faculty attended was examined.

2. METHOD

In this study, it was aimed to examine the measurement invariance of the state test anxiety scale. Ethics committee approval was obtained for the study.

2.1. Study Sample

Kline (2005) states that a sample of 200 people is generally sufficient in factor analysis studies. In addition, in different sources, it is recommended to reach 5-20 times the number of items for factor analysis (Alpar, 2016).

In order to reach a sample that can represent the range of the measured latent trait, the sample of the study were determined by purposeful sampling method and the data of the research were obtained from 956 university students. 572 of the students are female (59.8%) and 376 of them are male (39.3%). 8 (0.8%) of the students was not specify their gender. 8 (0.8%) of the students did not specify their gender. 400 of the students stated that they were medical school students (41.8%), 112 were dentistry students (11.7%), and 444 (46.4%) were health college students. 188 of the students (19.7%) are 1; 236 of them (24.7%) are 2; 428 of them (44.8%) are 3 and 96 of them (10%) are 4 grade. 8 (0.8%) of the students did not specify their class. 92 (9.6%) of the students stated that their socioeconomic level was low, 812 (84.9%) were medium, and 52 (5.4%) were high.

2.2. Data Collection Tool

The "State Test Anxiety Scale" was developed by Şahin (2019). The scale is scored in Likert type with 4 degrees and consists of 22 items in total. The scale consists of 3 components as 'cognitive', 'psychosocial' and 'physiological' and these components explain 59.21% of the total variance. This structure, which was revealed by the exploratory factor analysis, was also confirmed by the confirmatory factor analysis. Goodness of fit values were $\chi^2 / df = 1.72$, CFI = 0.96, NNFI = 0.96, IFI = 0.96, RMSEA = 0.05, SRMR = 0.05. The lowest score that can be obtained from the scale is "22" and the highest score is 88. The Cronbach alpha reliability of the scale was calculated as 0.94. Alpha reliabilities for sub-dimensions were calculated as 0.85 for physiological sub-dimension, 0.84 for psychosocial sub-dimension and 0.93 for cognitive sub-dimension. Similarly, the test-retest reliability of the scale, applied 4 weeks apart, was calculated as 0.74 for the physiological sub-dimension, 0.80 for the psychosocial sub-dimension, 0.78 for the cognitive sub-dimension, and 0.81 for the overall scale.

2.3. Data Analysis

The measurement invariance of the state test anxiety scale was analyzed in terms of gender, faculty and socioeconomic level variables using the multi-group confirmatory factor analysis method. In the multi-group confirmatory factor analysis, it is examined whether the model created by the researcher regarding the measurement tool is the same in different groups (Tabachnick & Fidell, 2001). For this purpose, equality limitations are imposed on the model established for subgroups in the multi-group confirmatory factor analysis, and the equivalence of intergroup parameters is examined by following a hierarchical order from the least limited to the most limited model (Başusta, 2010).

In the study, in order to examine the configural equivalence, it was tested whether the structure revealed by the state test anxiety scale measures the same formal structure in subgroups. For this purpose, the similarity of the number of factors and factor loadings of the measurements obtained from the subgroups was investigated. In the examination of metric equivalence, it was examined whether the units (distances between categories) of the conditionality test anxiety scale in different groups were equal. In the examination of scalar equivalence, it was examined whether the constants (origins) of linear regression equations between latent and observed variables change between groups. In the examination of strict equivalence, it was examined whether all the parameters estimated for the model presented were equal among the groups (Erkuş & Selvi, 2019).

In this study, ΔCFI values were used to examine the equivalence of parameters between groups and " $0.01 \geq \Delta CFI \geq -0.01$ " criterion was used for ΔCFI values. ΔCFI values provide information about the relationship between implicit scores and observed scores and are therefore recommended for evaluation of goodness of fit (Amery *et al.*, 2007; Brown, 2006; Vandenberg and Lance, 2000; cited in Uzun, 2010).

For model data fit, values of χ^2 / df , Root Mean Square of Approximate Errors (RMSEA), Comparative Fit Index (CFI), Adjusted Goodness of Fit Index (AGFI) and Root of Residual Means (RMR) were taken into consideration (Kline, 2005). As a criterion, the criteria given in Table 2 were taken into consideration (Çokluk *et al.*, 2010; Şimşek, 2007). Lisrel 8.7 program was used to analyze the data.

In the study, Little's MCAR test was used in missing data analysis. From the findings, it was concluded that the missing data were in random structure ($\chi^2 = 1083.123$, $df = 215$, $p = 0.001$). For this reason, expectation maximization method has been used to eliminate missing data. In addition, the data before the analysis were cleared of outliers and the skewness value calculated for the scale total score was calculated as 0.604 and the kurtosis value as -0.07. Şenocak (2014) states that the distribution can be considered normal if the skewness value is less than 1.00 and

the kurtosis value is less than 2.00. From here, it was accepted that the data of the study were normally distributed.

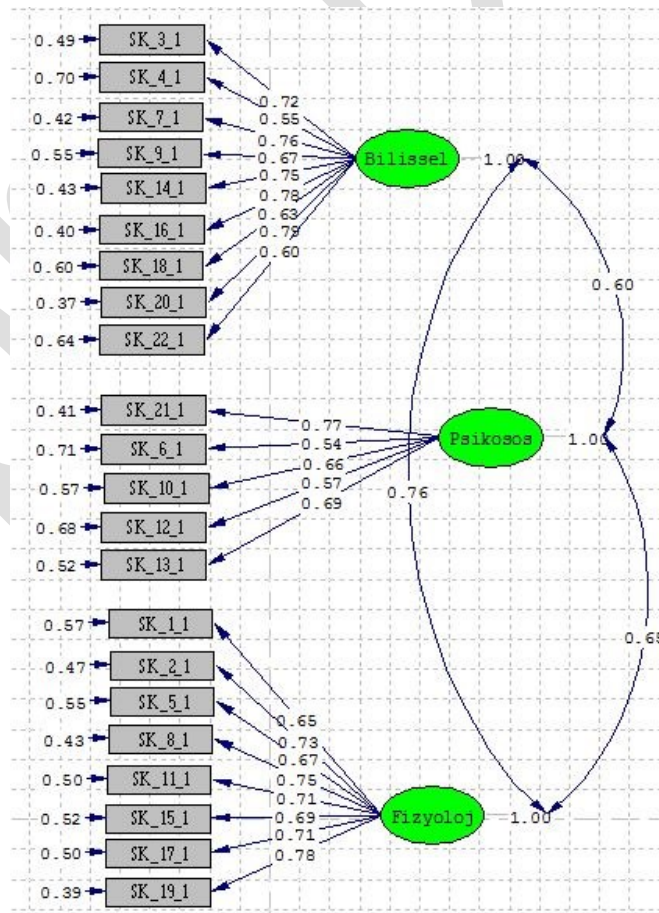
Table 2. Fit indices.

Fit Indices	Good Fit	Acceptable Fit
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 < \chi^2/df \leq 5$
RMSEA	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$
RMR	$0 \leq RMR \leq .05$	$.05 < RMR \leq .08$
SRMR	$0 \leq SRMR \leq .05$	$.05 < SRMR \leq .08$
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI < .95$
NFI	$.95 \leq NFI \leq 1$	$.90 \leq NFI < .95$
NNFI	$.95 \leq NNFI \leq 1$	$.90 \leq NNFI < .95$
GFI	$.95 \leq GFI \leq 1$	$.90 \leq GFI < .95$
AGFI	$.90 \leq AGFI \leq 1$	$.85 \leq AGFI < .90$

3. RESULT / FINDINGS

The fit statistics of the state test anxiety scale were found as $\chi^2/df = 8.95$, RMSEA = 0.08, CFI = 0.94, GFI = 0.90, SRMR = 0.06 and NFI = 0.94. When the obtained values are compared with the criteria in Table 2, it is seen that the fit statistics for the model are within the 'acceptable fit' limits except for the χ^2/df value. It is known that the value of χ^2/df is affected by the sample size and therefore it should not be interpreted alone in measurement invariance studies (Cheung & Rensvold, 2002). For this reason, it was accepted that model data fit was achieved for the model created. The path diagram obtained for the measurement model of the state test anxiety scale is presented in Figure 1.

Figure 1. The path diagram of the measurement model of the state test anxiety scale.



In this study, the measurement invariance of the gender, faculty and socioeconomic level variables of the state test anxiety scale was examined by taking the principle of hierarchy (configural, metric, scalar and strict equivalence) into consideration, and the findings obtained in Table 3 for the gender variable, in Table 4 for faculty variable, in Table 5 for socioeconomic level variable and in Table 6 for summary information on whether the invariance is provided for all of the examined groups.

Table 3. Fit statistics obtained from the measurement invariance study on gender variable.

	Equivalence	χ^2/df	CFI	RMR	GFI	RMSEA	ΔCFI	Invariance	
Gender	Cognitive	Configural	4.85	0.94	0.06	0.92	0.06	-	+
		Metric	8.52	0.94	0.07	0.91	0.07	0.00	+
		Scalar	10.54	0.90	0.08	0.85	0.14	0.03	-
	Psychoso- cial	Configural	8.16	0.96	0.03	0.99	0.07	-	+
		Metric	6.46	0.95	0.07	0.98	0.08	0.01	+
		Scalar	7.15	0.85	0.09	0.96	0.12	0.10	-
	Physiologi- cal	Configural	16.62	0.91	0.04	0.90	0.07	-	+
		Metric	15.6	0.90	0.07	0.90	0.08	0.01	+
		Scalar	13.8	0.85	0.06	0.85	0.16	0.05	-
Entire Scale	Configural	8.49	0.91	0.08	0.90	0.08	-	+	
	Metric	8.30	0.80	0.15	0.60	0.20	0.10	-	

When Table 4 is examined, it is seen that the configural equivalence is provided for the cognitive, psychosocial, physiological dimensions and the entire scale for the groups related to the faculty variable, and the metric invariance is provided for only the cognitive dimension. In addition, metric invariance could not be provided for the psychosocial, physiological dimensions and entire scale.

Scalar and strict invariance could not be provided for sub-dimensions and the entire scale.

Table 4. Fit statistics obtained from the measurement invariance study for faculty variable.

	Equivalence	χ^2/df	CFI	RMR	GFI	RMSEA	ΔCFI	Invariance	
Faculty	Cognitive	Configural	8.07	0.93	0.06	0.91	0.07	-	+
		Metric	7.5	0.92	0.08	0.90	0.08	0.01	+
		Scalar	10.35	0.86	0.19	0.79	0.17	0.06	-
	Psychosocial	Configural	10.01	0.92	0.03	0.94	0.07	-	+
		Metric	9.9	0.87	0.09	0.90	0.16	0.05	-
	Physiologi- cal	Configural	13.4	0.92	0.05	0.90	0.08	-	+
		Metric	12.2	0.88	0.09	0.82	0.18	0.04	-
	Entire Scale	Configural	7.1	0.91	0.07	0.90	0.08	-	+
		Metric	6.9	0.83	0.09	0.72	0.13	0.08	-

When Table 4 is examined, it is seen that the configural equivalence is provided for the cognitive, psychosocial, physiological dimensions and the whole scale for the groups related to the faculty variable, and the metric equivalence is provided for only the cognitive dimension.

In addition, metric equivalence could not be provided for the entire scale, for psychosocial and physiological dimensions. Scalar and strict equivalence could not be provided for sub-dimensions and the entire scale.

Table 5. Fit statistics obtained from the measurement invariance study for socioeconomic level variable.

		Equivalence	χ^2/df	CFI	RMR	GFI	RMSEA	ΔCFI	Invariance
Socioeconomic Level	Cognitive	Configural	8.02	0.89	0.38	0.45	0.15	-	+
	Psychosocial	Configural	6.9	0.94	0.09	0.92	0.08	-	+
		Metric	7.2	0.90	0.23	0.62	0.14	0.04	-
	Physiological	Configural	12	0.78	0.12	-0.98	0.18	-	+
	Entire Scale	Configural	8.5	0.91	0.05	0.90	0.11	-	+
		Metric	8.1	0.91	0.05	0.89	0.11	0.00	+
Scalar		8.5	0.80	0.09	0.78	0.11	0.11	-	

When Table 5 is examined, it is seen that for groups related to the socioeconomic level variable, Configural equivalence is provided only for the psychosocial dimension and the entire scale, and metric equivalence is provided only for the entire scale.

In addition, metric equivalence was not provided for subdimensions. Scalar and strict equivalence could not be provided for sub-dimensions and the entire scale.

Table 6. The results of the measurement invariance study regarding the variables of gender, faculty and socioeconomic level.

Variable	Sub-Dimensions	Configural Equivalence	Metric Equivalence	Scalar Equivalence	Strict Equivalency
Gender	Cognitive	+	+	-	-
	Psychosocial	+	+	-	-
	Physiological	+	+	-	-
	Entire Scale	+	-	-	-
Faculty	Cognitive	+	+	-	-
	Psychosocial	+	-	-	-
	Physiological	+	-	-	-
	Entire Scale	+	-	-	-
Socioeconomic Level	Cognitive	-	-	-	-
	Psychosocial	+	-	-	-
	Physiological	-	-	-	-
	Entire Scale	+	+	-	-

When Table 6 is examined, configural equivalence was provided for the socioeconomic status variable, except for cognitive and physiological dimensions. Metric equivalence was provided for the cognitive, psychosocial and physiological dimensions for the gender variable, for the cognitive dimension for the faculty variable, and for the socioeconomic status variable only for the whole scale. Scalar and strict invariance, on the other hand, could not be provided for any of the variables examined in the study.

4. DISCUSSION and CONCLUSION

Test anxiety can affect individuals of all ages and their families in society (Sieber, 1980). Test anxiety is a variable that disrupts the qualifications of measurement tools, especially their validity, and therefore, it is frequently studied in the literature to measure and determine its reasons.

In the literature, it is seen that almost all studies on test anxiety focus on scale development, scale adaptation, comparison and compilation studies (McDonald, 2001; Pekrun, 2004; Driscoll, 2007; Totan & Yavuz, 2009; Akin *et al.*, 2012; Başol, 2017; Aydın & Bulgan, 2017;

Bozkurt *et al.*, 2017; Bados & Sanz, 2005; Yao-Ting Sung & Tzu-Yang Chao, 2015; Şahin, 2019).

In comparison studies conducted in terms of different variables and groups, it is known that the obtained findings are accepted as the "real" difference between the groups in terms of the measured feature and comparison, interpretation and generalization are made in this direction (Mark & Wan, 2005). It should be kept in mind that these comparative studies and generalizations made without providing evidence regarding measurement invariance may produce erroneous results.

In this study, it was aimed to provide a test anxiety scale that provides evidence for measurement invariance to the literature. For this reason, the measurement invariance of the state test anxiety scale developed by Şahin (2019) was examined in terms of variables of gender, faculty and socioeconomic level.

The findings obtained showed that the state test anxiety scale provided Configural equivalence in groups related to the gender variable. In other words, the factor number and factor loadings pattern related to the measurements obtained from the groups of the gender variable with the state test anxiety scale are equivalent. This is an indication that the items of the measurement tool reveal the same formal structure in the groups examined (Sireci, Patsula & Hambleton, 2005). From this, it can be concluded that the state test anxiety scale measures the same formal structure in the groups of the gender variable.

Metric equivalence was provided within acceptable limits for the gender variable, but this is not valid for the whole scale. If metric equivalence is not achieved, the summability of the scale items is violated and it is emphasized that there may be bias in the items (Erkuş & Selvi, 2019). From here, it can be concluded that the items of the state test anxiety scale produce biased results in men and women. For this reason, it is recommended to examine biases on the scale and items by considering the gender variable and to edit the items found to be biased.

The hypotheses created in measurement invariance studies are examined by comparing the adaptation levels of the preceding model due to the principle of hierarchy (Başusta & Gelbal, 2015). For this reason, Scalar equivalence could not be examined because the metric invariance conditions of the scale were not met. In sub-dimensions of the scale provided metric equivalence, but was not provided scalar equivalence conditions. This shows that the constants used in linear regression equations between latent variables and observed variables are not equivalent between groups (Erkuş & Selvi, 2019). From here, it can be concluded that the scale and its sub-dimensions do not have the same origins for women and men, therefore, comparisons between groups on the basis of the gender variable cannot be made using the state test anxiety scale. On the other hand, strict invariance could not be studied since scalar invariance could not be achieved. Strict equivalence is based on the demonstration that all the predicted parameters are equivalent between groups, but it is known that this condition is often not met in social sciences (Erkuş & Selvi, 2019).

The findings obtained from the measurement invariance study conducted for the faculty variable show that the configural equivalence is provided for the cognitive, psychosocial, physiological sub-dimensions and the whole scale. This is an indicator that the state test anxiety scale reveals the same formal structure in the subgroups of the faculty variable. From here, it can be concluded that the state test anxiety scale measures the same formal structure in different subgroups of the faculty variable. Metric invariance was provided only for the cognitive sub-dimension. This shows that for the subgroups of the faculty variable, only the units of the cognitive sub-dimension of the scale are equivalent within acceptable limits. Units are not equivalent for other dimensions and for the whole scale. For this reason, it is recommended to conduct a bias study on the scale and items considering the faculty variable, and to edit the items found to be biased. Scalar and strict invariance could not be provided for the sub-

dimensions for the faculty variable and for the whole scale. From here, it can be concluded that the scale and its sub-dimensions do not have the same origins between faculties, so the comparison between groups on the basis of the faculty variable cannot be made using the state test anxiety scale. From here, it can be concluded that, apart from the cognitive sub-dimension, the conditionality test anxiety scale and its sub-dimensions produce biased results according to the variable of the faculty attended. This may be due to the different intensity and difficulties of the education programs of the faculties for which data were collected. As a matter of fact, in the faculties of medicine and dentistry, it is possible for students to repeat a year if they fail. But in the health college, students to repeat a course if they fail a course. This may be the reason why students' test anxiety cannot be explained in terms of equivalent origin and equivalent units. Findings on this subject also coincide with the findings of Erözkan (2004).

From the measurement invariance study conducted for the socioeconomic status variable, it is seen that the Configural equivalence was provided only for the psychosocial sub-dimension and the whole scale. In other words, in subgroups related to the socioeconomic status variable, the factor number and factor load of the state test anxiety scale are equivalent for the psychosocial sub-dimension and the whole scale. This is not the case in other sub-dimensions. From this, it can be concluded that the state test anxiety scale measures only the psychosocial sub-dimension in the subgroups of the socioeconomic status variable and the same formal structure in the whole scale. Metric invariance was not provided for sub-dimensions except for the entire scale. This situation shows that the units of the sub-dimensions are not equivalent in the subgroups related to the socioeconomic status variable. Considering the whole scale, the units for the relevant variable are equivalent. However, the scalar and strict invariance conditions for the socioeconomic level variable were not met. Therefore, it appears that the comparison between groups on the basis of the socioeconomic level variable cannot be made using the state test anxiety scale. This may be due to the students with low socioeconomic status not meeting their needs adequately. As a matter of fact, almost all of the situations such as suitable working environment, adequate nutrition, meeting the needs properly, future expectation depend on the socioeconomic level. This may be the reason why test anxieties of students with different socioeconomic levels cannot be explained with equivalent origin and equivalent units. Findings on this subject are in parallel with the findings of Softa, Ulaş Karahmetoğlu & Çabuk (2014).

From the findings obtained, it can be concluded that the state test anxiety scale and its sub-dimensions produce biased results in subgroups regarding the variables of gender, faculty and socioeconomic level, and considering these variables. So comparison between in this groups cannot be made and the findings cannot be generalized. Vanderberg & Lance (2000) state that testing the invariance of only a few parameters in measurement invariance studies may not produce sufficient results. For this reason, it may be suggested to carry out the "partial measurement invariance" study in terms of the variables examined within the scope of this study. On the other hand, this study was carried out on the students of Mersin University on the basis of gender, faculty and socio-economic status variables. It may be suggested to repeat the study on different variables, different universities and different faculties.

Acknowledgments

In this study, the measurement invariance of the state test anxiety scale developed by Şahin (2019) was investigated. I would like to thank Dr. Alper Şahin for allowing us to use this scale in the study.

Declaration of Conflicting Interests and Ethics

The author 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.

ORCIDHuseyin Selvi  <https://orcid.org/0000-0002-3513-0003>**5. REFERENCES**

- Akın, A., Demirci, İ., & Arslan, S. (2012). Revize edilmiş sınav kaygısı ölçeği: geçerlik ve güvenilirlik çalışması [Revised Test Anxiety Scale: Validity and Reliability study]. *Eğitim Bilimleri ve Uygulama*, 11(21), 103-118.
- Akın, A., Akın, U., Sarıçam, H., Aşut, S., Arslan, S., Demirci, İ., Toprak, H., & Çardak, M. (2013). *Friedben sınav kaygısı ölçeği Türkçe formu'nun geçerlik ve güvenilirliği* [The validity and reliability of the Turkish form of the friedben test anxiety scale]. Paper presented at the international conference on innovation and challenges in education. Dumlupınar University.
- Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *The Journal of Abnormal and Social Psychology*, 61(2), 207-215. <https://doi.org/10.1037/h0045464>
- Alpar, R. (2016). *Spor, sağlık ve eğitim bilimlerinde örneklerle uygulamalı istatistik, geçerlik ve güvenilirlik* [Applied statistics validity and reliability with examples in sports, health and education sciences]. DetayYayıncılık.
- Alıcı, D. (2013). Okula yönelik tutum ölçeği'nin geliştirilmesi: güvenilirlik ve geçerlik çalışması [Development of attitude scale towards school: reliability and validity study]. *Eğitim ve Bilim*, 38(168).
- Aydın, U., & Bulgan, G. (2017). Çocuklarda sınav kaygısı ölçeği'nin Türkçe uyarlaması [Turkish adaptation of the test anxiety scale in children]. *İlköğretim Online*, 16(2), 860-899. <https://doi.org/10.17051/ilkonline.2017.304742>
- Bados, A. L., & Sanz, P. (2005). Validation of the revised test anxiety scale and the friedben test anxiety scale in a spanish sample. *Ansiedad y estrés*, 11(2), 163-174.
- Başol, G. (2017). Ayda sınav kaygısı ölçeği: geçerlik ve güvenilirlik çalışması [IDA test anxiety scale: validity and reliability study]. *Uluslararası Eğitim Bilimleri Dergisi*, 4(13), 173-193. <https://doi.org/10.16991/INESJOURNAL.1506>
- Başusta, B. (2010). Ölçme eşdeğerliği [Measurement invariance]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 58-64.
- Başusta, N.B., & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: Pisa öğrenci anketi örneği [Testing measurement invariance in comparisons between groups: Pisa student survey sample]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Benson, J., & El-Zahhar, N. (1994). Further refinement and validation of the revised test anxiety scale. *Structural Equation Modelling*, 1(3), 203-221. <https://doi.org/10.1080/10705519409539975>
- Bozkurt, S., Ekitli, G. B., Thomas, C. L., & Cassady, J. C. (2017). Validation of the Turkish version of the cognitive test anxiety scale-revised. *Sage Open*, 1(1), 1-9. <https://doi.org/10.1177/2158244016669549>
- Büyüköztürk, Ş. (1997) Araştırmaya yönelik kaygı ölçeğinin geliştirilmesi [Development of an anxiety scale for research]. *Eğitim Yönetimi*, 3(1), 453-464.
- Cassady, J.C., & Johnson, R. E. (2002). Cognitive test anxiety and academic procrastination. *Contemporary Educational Psychology*, 27(1), 270-295. <https://doi.org/10.1006/ceps.2001.1094>
- Cheung, G.W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212. <https://doi.org/10.1177/0022022100031002003>

- Cheung, G., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance structural equation modeling. *A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cüceloğlu, D. (1998). *İnsan ve davranışı [Human and his behaviour]*. Remzi Kitabevi.
- Erkuş, A., & Selvi, H. (2019). *Psikolojide ölçme ve ölçek geliştirme III: ölçek uyarlama ve norm geliştirme [Measurement and scale development in psychology III: scale adaptation and norm development]*. Pegem Akademi.
- Ergene, T. (1994). Sınav kaygısı ile başa çıkma programının etkinliği [Effectiveness of coping program with test anxiety]. *Psikiyatri, Psikoloji ve Psikofarmakoloji (3P) Dergisi*, 2(1), 9-16.
- Erözkan, A. (2004). Üniversite öğrencilerinin sınav kaygısı ve başa çıkma davranışları [Test anxiety and coping behaviours of university students]. *Muğla Üniversitesi SBE Dergisi*, 12(1), 13-38.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik (spss ve lisrel uygulamaları) [Multivariate statistics for social sciences]*. Pegem Akademi.
- Driscoll, R. (2007). Westside test anxiety scale validation. *Psychology*, 8(14), 1-6.
- Geen, R.G. (1985). Test anxiety and visual vigilance. *Journal of Personality and Social Psychology*, 49(4), 963-970. <https://doi.org/10.1037/0022-3514.49.4.963>
- Gençdoğan, B. (2006). Lise öğrencilerinin sınav kaygısı ile boyun eğicilik düzeyleri ve sosyal destek algısı arasındaki ilişkiler [Relationships between high school students' test anxiety and their level of submission and social support perception]. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 7(1), 153-164.
- Hill, K.T., & Wigfield, A. (1984). Test anxiety: a major educational problem and what can be done about it. *The Elementary School Journal*, 85(1), 105-126. <https://doi.org/10.1086/461395>
- Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(1), 117-144. <https://doi.org/10.1080/03610739208253916>
- Kline, R.B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.
- Mark, B. A. & Wan, T. T. (2005). Testing measurement equivalence in a patient satisfaction instrument. *Western Journal of Nursing Research*, 27(6), 772-787. <https://doi.org/10.1177/0193945905276336>
- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21(1), 89-101. <https://doi.org/10.1080/01443410020019867>
- Sapir, S. & Aronson, A.E. (1990). The relationship between psychopathology and speech and language disorder in neurological patients. *Journal of Speech Hearing Disorder*, 55(1), 503-509. <https://doi.org/10.1044/jshd.5503.503>
- Sieber, J. E. (1980). *Defining test anxiety: problems and approaches*. Lawrence Erlbaum Associates.
- Sireci, S.G, Patsula, L., & Hambleton, R.K. (2005). *Statistical methods for identifying flaws in the test adaptation process*. Lawrence Erlbaum Associates.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://doi.org/10.1007/BF02294825>
- Pekrun, R. (2004). Test Anxiety and academic achievement, *Encyclopedia of the Social and Behavioral Sciences*. <https://doi.org/10.1016/B0-08-043076-7/02451-7>
- Softa Kaçan, H., Ulaş Karahmetoğlu G., & Çabuk, F. (2014). Lise son sınıf öğrencilerinin sınav kaygısı ve etkileyen faktörlerin incelenmesi [Examining of test anxiety of senior high

- scholl students and the factors affecting them]. *Kastamonu Eğitim Dergisi*, 23(4), 1481-1494.
- Sung, Y.T. & Chao, T.Y. (2015). Construction of the examination stress scale for adolescent students. *Measurement and Evaluation in Counseling and Development*, 48(1), 44-58. <https://doi.org/10.1177/0748175614538062>
- Şimşek, Ö.F. (2007). *Yapısal eşitlik modellemesine giriş (temel ilkeler ve lisrel uygulamaları) [Introduction to structural equating modeling (basic principles and applications of lisrel)]*. Ekinoks.
- Tabachnick, B. & Fidell, L. S. (2001). *Using multivariate statistics*. Allyn & Bacon.
- Şahin, A. (2019). Durumluk sınav kaygısı ölçeği (duskö): geçerlik ve güvenilirlik çalışması [State test anxiety scale: validity and reliability scale]. *Trakya Eğitim Dergisi*, 9(1), 78-90. <https://doi.org/10.24315/tred.450423>
- Şenocak, M. (2014). *Biyoistatistik ve araştırma yöntembilimi [Biostatistics and research methodology]*. İstanbul Tıp Yayınevi.
- Totan, T. & Yavuz, Y. (2009). Westside sınav kaygısı ölçeğinin Türkçe formunun geçerlik ve güvenilirlik çalışması [Validity and reliability study of the Turkish form of the Westside test anxiety scale]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 95-109.
- Turgut, M. F. & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. Pegem Yayıncılık.
- Uzun, B. & Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin tıms-s-r örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi [Evaluation of measurement invariance of some variables related to science achievement by gender in TIMSS_R sample]. *Education and Science*, 35(1), 26-35.
- Öner, N. (1990). *Sınav kaygısı envanteri el kitabı [Test anxiety inventory handbook]*. Yüksek Öğretimde Rehberliği Tanıtma ve Rehber Yetistirme Vakfı Yayınları.
- Özgülven, İ. E. (2007). *Psikolojik testler [Psychological tests]*. PDREM Yayınları.
- Yao-Ting Sung & Tzu-Yang Chao (2015) Construction of the examination stress scale for adolescent students. *Measurement and Evaluation in Counseling and Development*, 48(1), 44-58. <https://doi.org/10.1177/0748175614538062>
- Vanderberg, R.J. & Lance, C. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-69. <https://doi.org/10.1177/109442810031002>
- Van der Vijver, F. J. R. & Poortinga, Y. H. (2005). *Conceptual and methodological issues in adapting tests*. Lawrence Erlbaum Associates.
- Wren, D. G. & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress and Coping*, 17(3), 227-240. <https://doi.org/10.1080/10615800412331292606>
- Zeidner, M. (2007). *Test anxiety in educational contexts: concepts, findings, and future directions*. Elsevier.