Psychological Thought
psyct.swu.bg | 2193-7281

South-West University "Neofit Rilski"

Research Article

Psychometric Properties of the State-Trait Anxiety Inventory-Trait (STAI-T) in Argentinian Adolescents

Sebastian Jesús Garrido*[a,b], Ana Estefania Azpilicueta[c], Valeria Estefania Morán[b,d], Marcos Cupani[a,b]

[a] Psychological Research Institute (IIPsi), Faculty of Psychology, National University of Cordoba, Boulevard de la Reforma esquina Enfermera Gordillo s/n, Córdoba, Argentina.

[b] National Scientific and Technical Research Council (CONICET), Argentina.

[c] 21st Century University, Calle de los Latinos 8555, Córdoba, Argentina.

[d] Research Institute in Social, Territorial and Educational Studies (ISTE), National University of Río Cuarto, Córdoba, Argentina.

Abstract

The aim of the present study was to analyze the psychometric properties of the State-Trait Anxiety Inventory- Trait dimension (STAI-T), Form X. A total of 798 Argentinian adolescents aged between 12 and 18 years (*M* = 14; *SD* = 1.31) participated in this study. The psychometric properties of the instrument were analyzed from the contributions of the Rasch model, which allowed us to assess the measurement model fit, response threshold order, item and person fit, local dependence, differential item functioning (DIF), reliability, targeting, and unidimensionality. The results showed that the STAI-T, anxiety-absent items, and anxiety-present items presented values close to the expected ones for item and person fit, ordered response categories, appropriate reliability levels, and adequate targeting. Only in anxiety-absent items, it was possible to verify compliance with unidimensionality, whereas in anxiety-present items there were items with local dependence or showing DIF. The present study provides empirical evidence based on a rigorous measurement model that allows corroborating the validity of the scale to assess anxiety in Argentinian adolescents.

Keywords: Anxiety; STAI-T; Validation; Psychometric properties; Rasch analysis.

Table of Contents

Method
Results
Discussion
References

Psychological Thought, 2022, Vol. 15(2), 148-167, https://doi.org/10.37708/psyct.v15i2.725
Received: 2022-03-07. Accepted: 2022-06-07. Published (VoR): 2022-10-31.
Handling Editor: Natasha Angelova, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria.
*Corresponding author at: Psychological Research Institute, National University of Cordoba, National Scientific and Technical Research Council (CONICET), Argentina. E-mail: sebajgarrido@unc.edu.ar

Spielberger et al. (1970) developed the STAI by distinguishing between state and trait anxiety. *State anxiety* refers to the transitory emotional state, which is characterized by tension and an apprehension feeling as well as increased activity of the autonomic nervous system. On the other hand, *trait anxiety* refers to the stable tendency of personality involved in how an individual may experience and respond to different situations perceived as threatening. In addition, Spielberger et al. (1980) mentioned that negatively (non-reversed) and positively (reversed) worded items refer to the presence or absence of anxiety, respectively. The author added that while anxiety-absent items are more sensitive to measure low levels of anxiety, anxiety-present items are more sensitive to measure high levels of anxiety (Spielberger et al., 1980).

In general, many researchers have used the classical test theory (CTT) to study the psychometric properties of the STAI. Specifically, this theoretical approach allows analyzing the STAI reliability and validity properties, but does not examine; for example, item-level psychometric properties or the appropriateness of response scale categories. The item response theory (IRT) approach fills this gap, since it allows obtaining invariant measures, regardless of the scale and the sample assessed (Engelhard Jr., 2013). This characteristic makes IRT a useful tool for cross cultural validation studies, as has been proved in previous studies (Cupani et al., 2020).

According to what has been previously established, it is observed that, although the STAI psychometric properties have been already evaluated from the CTT, there is no rigorous evaluation of the instrument that would allow its use in a cultural context different from the one in which it was created. For this reason, the main objective of the present study is to examine the State dimension psychometric properties (present- and absent-anxiety items) by means of the Rasch model (Rasch, 1960) using a sample of Argentinian adolescents.

Method

Participants

Participants were high school students attending years 1-5 in public and private schools from Córdoba, Argentina. Córdoba is the second most populous Argentine city with 1 370 585 inhabitants [National Institute of Statistics and Censuses (INDEC), 2015]. Data from 798 participants (*M* = 14; *SD* = 1.31; age range = 12-18 years; 494 females, 304 males) were included in the analyses. Participants belonged to different socioeconomic status (medium-high and low), and the schools chosen were located in central and peripheral areas of the city.

Procedure

Prior consent was obtained from students' parents and the director of each school. Participants were explained the main propositions of the study during class time. The study was carried out in accordance with the ethical recommendations of the Argentine law (Law N° 25 326, 2000) and the American Psychological Association (APA, 2017). Based on these recommendations, participants were informed about the confidentiality of their responses, the anonymous nature of their participation, and their right to leave the study at any time without negative consequences for them. The instrument was administered on a scheduled day and time during class. Overall, participants responded the test within 20 min.

Measure

State-Trait Anxiety Inventory (STAI; Spielberger et al., 1970)

The STAI is composed by the State Anxiety Scale (STAI-S) and the Trait Anxiety Scale (STAI-T). The STAI-S assesses the current state of anxiety through responses on apprehension, nervous tension, worry, and activation/arousal of the autonomic nervous system. The STAI-T assesses relatively stable aspects of anxiety proneness through responses on general states of calmness, confidence, and security. Both scales have 20 items, and the items are rated on a 4-point frequency scale ranging from 1 (*Almost never*) to 4 (*Almost always*). In the present study, it was used the STAI-T, which consists of 13 items that assess anxiety presence and 7 items that assess anxiety absence and include the statements *Getting tired quickly* or *Feeling secure*, respectively. Concerning the STAI-T internal consistency, Cronbach's Alpha coefficients ranged from .86 in a sample of high school student to .95 in a sample of military personnel. Likewise, the criterion validity analysis in the Spanish adapted version (Leibovich de Figueroa, 1991) used in the present study showed a moderate Pearson correlation with personality trait measures from inventories such as the Minnesota Multiphasic Personality Inventory.

Data analysis

Rasch analysis was run with RUMM2030 software (Andrich et al., 2010), based on the partial credit parameterization. The choice of the partial credit model (PCM) was determined through a significant result (p < .001) in the log-likelihood ratio test. Measurement quality was assessed by the following indicators.

Statistic of item-trait interaction

The chi-square (χ^2) indicator has been commonly used to assess the item-trait interaction, although sample size can influence its estimations. For this purpose, we assessed the fit between the data and the model through χ^2 and the root mean square error of approximation (RMSEA). We assumed a good model fit when the item-trait interaction showed a low χ^2 value with p > .05 (Cavanagh & Waugh, 2011) and the RMSEA had a value lower than 0.02 (Tennant & Pallant, 2012).

Thresholds

The order of thresholds was examined by considering that the measured trait level is consistent with the participants´ response choice (Pallant & Tennant, 2007). Disordered response categories were corrected by combining adjacent response categories. Analyses were then replicated with collapsed categories to determine whether the modifications improved the model (Parkitny et al., 2012).

Item and person fit residual

It was examined the differences between the model observed and expected residual values for persons and items. To consider a good model fit, we expected a mean (*M*) and standard deviation (*SD*) close to 0 for person and item residual values. In addition, we considered the residuals as indicators of a model misfit if their values were above ± 2.5, had a significant chi-square, and adequate Bonferroni fit (p < .05). Considering that fit indices can be modified in cases with anomalous response patterns, the item or person misfit was analyzed and eliminated if necessary (Tennant & Conaghan, 2007).

Local independence

Local dependence related to response dependency was determined by identifying those items with residual correlations above .20 in the person–item residual correlation matrix (Andrich & Marais, 2019). Local independence was addressed by grouping local dependent items into testlets (Nilsson & Tennant, 2011).

Differential Item Functioning (DIF)

It was analyzed DIF to assess measurement invariance. There are two types of DIF: uniform DIF, in which the group shows a systematic difference in the responses given to an item across a measurement range of the attribute, and non-uniform DIF, in which the differences between groups are not uniform or systematic. To identify both types of DIF, we used a two-way ANOVA and Bonferroni correction < .05 for each item. To improve the model fit, uniform and non-uniform DIF received different treatments. Uniform DIF was solved by splitting items between groups (females and males), whereas items with non-uniform DIF were eliminated (Pallant & Tennant, 2007; Tennant & Pallant. 2006).

Reliability index

We examined the scale ability to separate persons along the latent trait using the Person Separation Index (PSI), the number of statistically Distinct Levels of Performance Ability (DLPA), and the Distribution-Independent Person Separation Index (DI-PSI). PSI and DI-PSI values ranging from .70 to .84 were considered sufficient for group level, and a value equal or greater than .85 was adequate for individual-level measurements (Tennant & Conaghan, 2007).

Targeting

Targeting was examined to determine how well the scale range measurement matched the ability distribution of the sample. We assessed the target scale through the Targeting Index (TI) and the person-item location threshold distribution map. The TI was interpreted as a good value when ranging from -1 to 1 and fair when ranging from -2 to 2 (Fisher, 2007). In the case of the person-item threshold distribution map, attention was focused on person locations against item-threshold locations, where mean person location values around zero indicated that the scale was well targeted. Inadequate targeting could be related to lower reliability, thereby the scale could present difficulties to assess persons according to their abilities (Hagquist et al., 2009).

Unidimensionality

Unidimensionality was determined by assessing the clustering of items in the first residual factor. This result was obtained using the principal component analysis (PCA) of the residuals. The PCA delimited two subsets by grouping items into positive and negative loads on the first principal component (± .30). In addition, a paired t-test analysis proposed by Smith (2002) was performed on the subsets to assess whether the cases differed significantly at the 0.05- level. Consequently, unidimensionality was achieved when proportions outside the range of ± 1.96 did not exceed 5% (Pallant & Tennant, 2007; Tennant & Pallant, 2006).

Results

STAI-T

The 20 items of the STAI-T presented an item fit of *M* = 0.18 and *SD* = 5.20 and a person fit of *M* = −0.24 and *SD* = 1.49. A significant chi-square ($\chi^2(60) = 792.11$; $p < .01$; RMSEA = .12) was obtained, which could indicate a poor fit of some items (Table 1). Disordered thresholds were observed in items 3, 4, and 14. The item-level analysis showed residual values of ± 2.5 in 11 items and significant χ2 and F values (Bonferroni at α ≤ .05) in 10 items (Table 2). The person-level analysis did not reveal extreme values, but 8.77% of the cases had residual values (greater than ± 2.5). The residual correlation matrix showed that eight pairs of items had values greater than .20. A uniform DIF was identified in items 3, 5, and 14. The reliability assessed with PSI and DI-PSI showed values above .80. As seen in Table 1, unacceptable values were observed in relation to the unidimensional assumption (per C < 5% = 16.92%). Considering these results, we assessed whether item modifications allow improving the model fit. The categories of items 3, 4, and 14 were ordered by collapsing categories 1 and 2 (0112). Next, item pairs with correlation values higher than .20 were combined, but the change did not improve the model fit. After that, items with uniform DIF were analyzed and only item 11 displayed adequate properties after being split. In addition, items with non-uniform DIF were deleted. As a result of the modifications, a final version was obtained with *M* = 0.36 and *SD* = 1.37 for items and *M* = −0.30 and *SD* = 1.26 for persons. The chi-square statistic showed a better fit than the initial model ($\chi^2(33) = 52.19$; $p = .018$), and RMSEA was 0.03. There was no disorder threshold, and dependence was observed between items 10 and 13 (.268). Furthermore, in the final version, item 6 presented a residual value of 2.59 and showed neither uniform

https://doi.org/10.37708/psyct.v15i2.725

nor non-uniform DIF. Values of PSI and DI-PSI above .80 indicated measurement accuracy at the individual level. Concerning targeting, the TI (Table 1) and the person-item threshold (Fig. 1) show that the scale was on target with a sample distribution from −4.03 to +3.75 logits. The assumption of local independence and unidimensionality was not demonstrated because item 11 was split (Table 1).

Table 1.

Summary of the fit indexes of the STAI-T, anxiety-present and anxiety-absent items: initial and final versions.

	Fit to the Rasch model									Targeting					
	Item fit residual		Person fit residual		Item-trait interaction			Person location				Separation reliability			Unidimensionality
	M	*SD*	*M*	*SD*	X^2 (*df*)	*p*	*RMSEA*	*M*	*SD*	SEM	TI	PSI	DLPA	DI-PSI	Per C
STAI-T															
Initial	0.176	5.202	-0.243	1.488	792.113(60)	.000	.123	-0.483	0.88	0.30	-1.62	.88	7	.98	16.92%
Final	0.356	1.365	- 0.300	1.255	52.186(33)	.018	.027	- 0.376	1.05	0.47	-0.81	.80	‡	‡	‡
Anxiety-present items															
Initial	0.112	4.805	-0.259	1.377	459.574(39)	.000	.116	-0.343	1.01	0.39	-0.89	.85	5	.96	4.51%
Final	0,179	1,376	-0,335	1,267	61,654(33)	.002	.033	-0.501	1.32	0.54	-0.93	.83	‡	‡	‡
Anxiety-absent items															
Initial	-0.365	5.425	-0.362	1.152	312.825(21)	.000	.132	0.981	1.40	0.63	1.55	.80	4	.94	5.76%
Final	0.107	2.405	-0.364	0.964	74.646(12)	.000	.080	1.225	1.83	0.92	1.33	.75	3	.90	1.88%
Recommended value	0	*≤1.4*	0	*≤1.4*	-	> 0.05	<0.02	-	-	-	*[-2,2]*	>.70		*≥0.85*	< 5%

p = Bonferroni-corrected probability value for chi-square; RMSEA = root mean square error of approximation; SEM = standard error of measurement; T.I = targeting index; PSI = person separation index; DLPA = distinct levels of performance ability; DI-PSI = distribution-independent person separation index (based on DLPA); per C < 5% = proportion of t-tests that were significant at the level of significance of .05.

‡ = Analyses were not performed because items were split for uniform DIF; thus, the software could not calculate them.

Table 2.
Fit index of items from the initial analysis of the STAI-T.

Item statistics	Location (logits)	*SE*	Fit residual	χ^2	χ^2 (*p*)	*F*	*F* (*p*)
1*	0.90	0.05	-4.61	44.09	**.000**	21.22	**.000**
2	-0.28	0.05	1.34	3.22	.359	1.12	.341
3	0.25	0.05	-3.69	28.27	**.000**	12.36	**.000**
4	0.00	0.04	-0.15	4.01	.261	1.30	.273
5	-0.07	0.05	**4.75**	41.96	**.000**	12.38	**.000**
6*	-0.23	0.04	0.27	4.69	.196	1.66	.175
7*	0.15	0.04	**3.66**	10.26	.016	3.31	.020
8	-0.16	0.04	-1.35	12.00	.007	4.57	.004
9	-0.69	0.04	1.39	4.15	.246	1.37	.250
10*	0.79	0.05	-2.79	28.97	**.000**	12.42	**.000**
11	-0.30	0.05	1.23	4.60	.204	1.58	.193
12	-0.43	0.04	-3.60	28.34	**.000**	13.51	**.000**
13*	0.00	0.04	-1.66	8.94	.030	3.65	.012
14	-0.53	0.04	**16.00**	324.35	**.000**	70.20	**.000**
15	0.19	0.05	-5.71	62.28	**.000**	33.83	**.000**
16*	0.75	0.05	-3.74	28.90	**.000**	12.57	**.000**
17	-0.09	0.04	-1.88	11.75	.008	4.77	.003
18	0.17	0.04	-4.00	24.90	**.000**	11.98	**.000**
19*	-0.23	0.04	**9.46**	106.87	**.000**	29.50	**.000**
20	-0.17	0.04	-1.38	9.59	.022	3.83	.010

The items that did not meet the criteria established in some indices are shown in bold. * = items are rescored

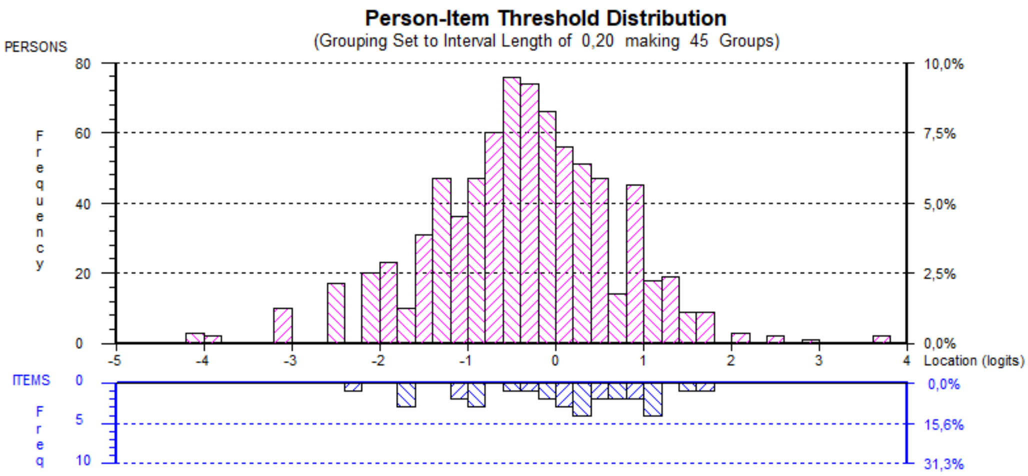

Figure1. *Person-item threshold distribution for the STAI-T*

Person-item distribution graphs for the STAI-T in the upper part of the graph plots the distribution of person ability on a logit scale, and the lower part plots the item difficulty threshold.

Anxiety-present items

The subscale composed of 13 items presented an item fit of M = 0.11 and SD = 4.80, and the person fit was M = − 0.26 and SD = 1.38. A significant chi-square ($\chi^2(39)$ = 459.57; p < .01) and an inadequate fit value in RMSEA = 0.12 were obtained, which could indicate a poor fit of some items. Disordered thresholds were observed in items 3, 4, and 15. As shown in Table 3, item-level analyses showed residual values of ± 2.5 in six items. In addition, seven items presented significant χ^2, and eight items had F values (Bonferroni at α ≤ .05). In the person-level analysis five extreme cases were found, and 9.27% of persons had residual values (greater than ± 2.5). The residual correlation matrix showed values greater than .20 between items 3 and 15 (.24). A uniform DIF was identified in items 3, 5, and 14. With respect to the link between person location and item location, the PSI index and TI presented acceptable values (Table 1). Furthermore, the analysis of unidimensionality assumption showed an acceptable value (per C < 5% = 4.51%).

Table 3.
Fit Index of anxiety-present items: initial analysis

Item statistics	Location (logits)	*SE*	Fit residual	χ^2	χ^2 (*p*)	*F*	*F* (*p*)
(2)	-0.12	0.05	2.12	5.85	.119	2.29	.077
(3)	0.41	0.05	-0.97	12.58	.006	4.04	.007
(4)	0.16	0.04	1.82	4.63	.201	1.13	.335
(5)	0.10	0.05	**4.71**	25.37	**.000**	8.29	**.000**
(8)	0.01	0.04	-2.05	11.47	.009	5.23	**.001**
(9)	-0.57	0.04	-0.70	1.29	.733	0.87	.457
(11)	-0.13	0.05	0.53	4.71	.194	1.61	.186
(12)	-0.27	0.04	-3.49	22.30	**.000**	10.90	**.000**
(14)	-0.40	0.04	**13.32**	260.15	**.000**	59.09	**.000**
(15)	0.36	0.05	-3.38	32.85	**.000**	15.19	**.000**
(17)	0.09	0.05	-3.44	18.42	**.000**	9.32	**.000**
(18)	0.36	0.04	-4.92	41.96	**.000**	23.44	**.000**
(20)	0.00	0.04	-2.10	18.00	**.000**	7.30	**.000**

Items that did not meet the criteria established in some indices are in bold.

Based on previous results, modifications were included in the original subscale to obtain appropriate values of psychometric properties. For this reason, items 3, 4, and 15 were ordered by collapsing categories 1 and 2 (0112). To deal with local dependence, a testlet combining items 3 and 15 was created, but it did not improve the model fit; thus, the testlet was not kept in the final version. Items 2, 5, 14, and 18 were deleted considering non-uniform DIF and residual values above ± 2.5. Next, we replicated the analyses finding that items 3 and 11 showed a uniform DIF; thereby, we conducted a uniform DIF analysis for both items and obtained an adequate model fit. Because of these modifications, we obtained an M = 0.18 and SD = 1.38 for items and an *M* = - 0.34 and *SD* = 1.27 for persons. The chi-square statistic (χ^2 (33) = 61. 65; *p* <.01) and RMSEA (0.03) showed a model misfit. In addition, item 11 for females showed a residual value outside the range from - 2.5 to 2.5. Response dependence was found between item 15 and item 3 for males with a residual correlation of .25, but no disorder threshold was found in the final version. The PSI presented a value of .83, indicating an acceptable reliability. With respect to the sample distribution, the person location ranged from −4.13 to +3.92 logits and showed a TI greater to −1 indicating a good fit in the relationship of item locations and person locations (Fig. 2). The assumption of local independence and unidimensionality could not be proved because some items were split (per C < 5% = 3.63%).

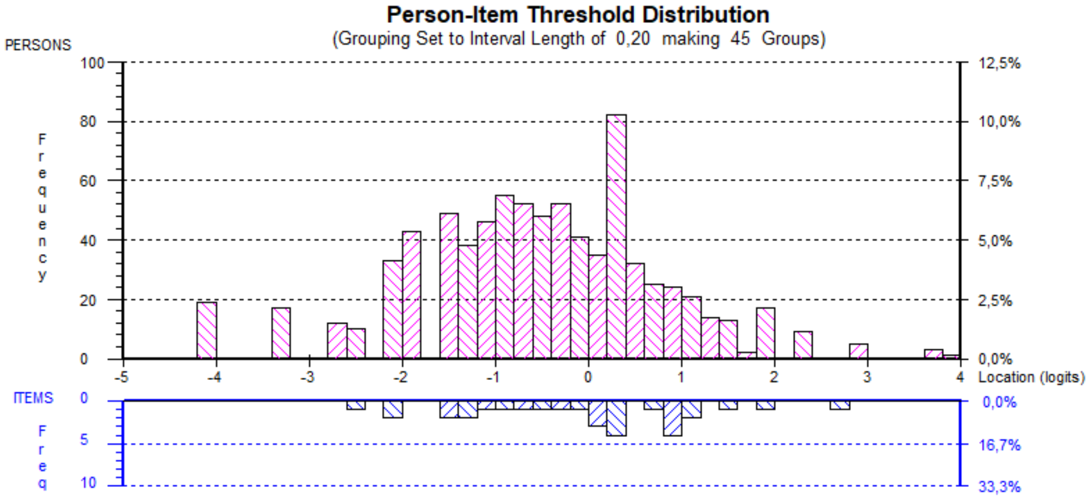

Figure 2. *Person-item threshold distribution for anxiety-present items*

The upper part of the graph plots the distribution of person ability on a logit scale, and the lower part plots the item difficulty threshold.

Anxiety-absent items

The subscale anxiety absence, composed of 7 items, presented an item level fit of *M* = − 0.37 and *SD* = 5.43, whereas the person fit was *M* = −0.36 and *SD* = 1.15. The significant chi-square (χ^2 (21) = 312.83; $p < .01$; RMSEA = 0.13) indicated a poor fit of some items (Table 1). As shown in Table 4, no disordered thresholds were observed in the seven items evaluated. Concerning misfit items, two items showed residual values greater than ± 2.5, and six items presented significant χ^2 and *F* values (Bonferroni at α ≤ .05). In a person-level analysis, 33 cases (4.13%) had extreme values and 24 persons (3.00%) had residual values (higher than ± 2.5). The residual correlation matrix showed correlation values greater than .20 between items 1 and 10 (.23), 1 and 16 (.26), and 10 and 16 (.34). Item 19 presented uniform DIF. PSI and DI-PSI values were acceptable according to reliability criteria. Likewise, the TI presented acceptable values. The analysis of unidimensional assumption showed acceptable values (per C < 5% = 5.76%; Table 1).

https://doi.org/10.37708/psyct.v15i2.725

Table 4.
Fit index of anxiety-absent items: initial analysis.

Item	Location (logits)	*SE*	Fit residual	χ^2	χ^2 (*p*)	*F*	*F* (*p*)
1)	-0.79	0.06	-4.74	57.26	**.000**	32.56	**.000**
(6)	0.71	0.05	-0.53	4.50	.212	1.95	.120
(7)	0.23	0.05	**4.76**	20.02	**.000**	6.80	**.000**
(10)	-0.66	0.06	-4.27	45.26	**.000**	26.39	**.000**
(13)	0.42	0.05	-1.12	16.52	**.001**	6.82	**.000**
(16)	-0.59	0.06	-5.65	54.34	**.000**	32.64	**.000**
(19)	0.70	0.05	**8.99**	114.93	**.000**	31.32	**.000**

Items that did not meet the criteria established in some indices are in bold.

The results of the initial analysis were considered to improve the anxiety-absent item fit. Contrary to expectations, the uniform DIF analysis of item 19 did not improve the model fit. This was observed in the misfit of the item characteristic curves and the residual values above ± 2.5 for item 19 for males and females. We deleted item 19 because it showed misfit in its original and split forms. It was analyzed the model fit by examining separately the testlet between items 1 and 10, 1 and 16, and 10 and 16; without these changes, the model showed a satisfactory fit. We deleted items 7 and 10 because they presented residual values outside ± 2.5 and non-uniform DIF. After these modifications, the model showed an *M* = 0.11 and *SD* = 2.41 for items and an *M* = -0.36 and *SD* = 0.96 for persons. The chi-square statistic was significant, which did not imply a better model fit ($\chi^2(12) = 74.646$; $p < .01$), and RMSEA was .08. The residual matrix correlation did not show items with local dependence. Item 16 presented a residual value equal to – 2.60. There were neither uniform and non-uniform DIF, nor threshold disorder in the items. The reliability index was acceptable for PSI (≥.70) and DI-PSI (.90), suggesting measurement accuracy at the individual level. Regarding the ability distribution in the sample, the person location analysis showed a range of abilities between −4.47 and +3.94 logits. The TI was 1.33, indicating a fair target in the sample, as was also observed by visual inspection of the targeting graph (Fig. 3). The assumption of local independence and unidimensionality was proved (per C < 5% = 1.88%; Table 1).

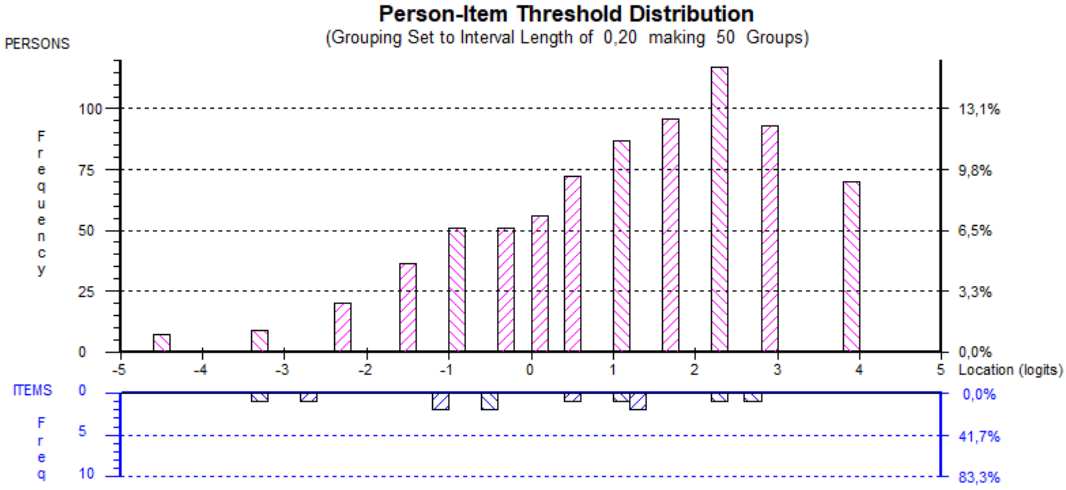

Figure 3. *Person-item threshold distribution for anxiety-absent items*

The upper part of the graph plots the distribution of person ability on a logit scale, and the lower part plots the item difficulty threshold.

Discussion

The present study aimed to assess the psychometric properties of the items composing the STAI-T. For this purpose, we conducted a Rasch analysis following the technical recommendation for psychometric studies (Cupani et al., 2020). The initial analysis showed that the STAI-T anxiety-present and anxiety-absent items had inadequate fit. In addition, there were disordered thresholds on the STAI-T anxiety-present items. One solution might be the combination of thresholds to assess which categorization might work best. To accomplish this, we collapsed categories 1 and 2 and obtained a better performance of categories in both factorial structures (Hobart & Cano, 2009; Tennant & Conaghan, 2007).

With respect to residual fit, in the final version of the STAI-T, there were no under- or overdiscriminating items, which might be because these items were either deleted or showed a better residual fit after the modifications made. However, item 16 of the anxiety-present items and item 11 (for females) of the anxiety-absent items were over-discriminating items. In summary, the lack of under- or overdiscriminating items in the final unidimensional scale might be due to the elimination of multidimensionality (Andrich & Marais, 2019), whereas the overdiscriminating items between anxiety-present and anxiety-absent items might be due to the presence of redundancy and local dependence between items (Andrich & Marais, 2019).

Testlets were created for items whose residual correlations showed values above .20; however, in most cases this strategy did not improve the fit of the data to the Rasch model. Moreover, although the fit of the final model was better than the one of the initial models, the last version of the STAI-T showed local dependence between item10 (Happy) and 13 (Feeling safe), and between item 15 (Feeling blue) and 3 (Crying) for males on the anxiety-present items. Furthermore, in the final version of the anxiety-absent items, there was not local dependence because item 10 was eliminated. In this case, when item 10 was eliminated, the other items did not show negative changes. However, when item 10 was deleted in the STAI-T, this affected item 6, which showed a high misfit, and therefore item 10 was kept.

In the initial analysis, the PCA showed a lack of compliance with the unidimensional assumption for the STAI-T and anxiety-absent items, which might be because the instrument was built from the perspective of the classical theory (Lundgren-Nilsson et al., 2019). This theoretical perspective does not focus on local dependence between items. This could be considered a limitation of this theoretical perspective because the residual correlation between items may result in an inadequate interpretation of the instrument. In addition, the presence of local dependence may hide the multidimensionality of the scale and items with local dependence may inflate some indicators such as the classical reliability indices (Nilsson & Tennant, 2011)

Unlike other studies that demonstrated configural invariance, full metric invariance, and scalar invariance across gender (Cooper et al., 2007), in the present study, the final version of the STAI-T and anxiety-present items had uniform DIF. In item 3 (Crying), females showed higher scores than males (the item is easier for females than for males); this result is similar to the one reported in a sample of Brazilian college students (Andrade et al., 2001). However, in item 11 (Taking things hard) males showed higher scores than females, which is contrary to previous evidence (Andrade et al., 2001). In the present study, we decided to split these items by gender and not to remove them because of the differences in the response patterns caused by gender differences. Regarding item 3, the differences in the response patterns for males and females could be because females are more likely to show their feelings than males (Chaplin & Aldao, 2013). On the other hand, the differences found in the response pattern of item 11 could be due to explanatory style, which tends to be more optimistic for females than for males; thus, males could interpret certain facts as more difficult than females (Martínez & Sewell, 2000).

The TI and the targeting graphs show an adequate match between the measurement range of the scale and the distribution of the calibrating sample, from which it can be interpreted that STAI-T anxiety-present and anxiety-absent items had good potential for accurate measurement (Hobart & Cano, 2009). In addition, the reliability analysis showed that the PSI value for three subscales was in the range considered acceptable for measuring ability at the group level (from .74 to .84). The low PSI values may be due to the skewed distribution of person response. Furthermore, the strategy to deal with this limitation was to calculate DI-PSI because it is not influenced by sample distribution (Pellicciari et al., 2020). For this reason, we calculated the DI-PSI for the anxiety-absence subscale and obtained a reliability value equal to .90, indicating that this subscale can measure ability at the individual level.

Limitations and future research suggestions

There are limitations to the study that influence the generalizability of our findings. One of the limitations we can mention is that, due to the origin of the sample, it is not advisable to use these results to carry out interventions with adolescents in the clinical setting. In addition, we observed non-compliance with the assumption of local independence and the presence of differential functioning of some STAI-T and anxiety-present items, which could affect their potential for accurate measurements. Another limitation is that the mismatch of some items might be because they do not adequately represent the anxiety construct, they are difficult to understand for the target population, and the translation of some items has no equivalence in Spanish (e.g., item 15). In view of these limitations, future research could evaluate the content and wording of the items through a review by expert judges; in addition, the item quality items should be assessed in a pilot test with adolescents with and without anxiety symptoms who may come from clinical and non-clinical samples. This is a methodological strategy that would allow obtaining items that represent the trait anxiety construct and ensuring that the items are comprehensible to adolescents.

Conclusion

This study uses a rigorous psychometric model such as the Rasch analysis model to assess the psychometric properties of the STAI-T (form X), which is used in most studies in Spanish. Also, our Rasch analysis complements previous findings based on the CTT and provide psychometrical evidences based on advantages of this type of IRT-derived models, such as, reliability and validity are not data-dependent, provides adequate person statistics, and measures person and item estimates on the same metric scale.

Funding/Financial Support

The authors have no funding to report

Other Support/Acknowledgement

The authors have no support to report.

Competing Interests

The authors have declared that no competing interests exist.

References

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. Retrieved from https://www.apa.org/ethics/code/ethics-code-2017.pdf

Andrade, L., Gorenstein, C., Vieira Filho, A. H., Tung, T. C., & Artes, R. (2001). Psychometric properties of the Portuguese version of the State-Trait Anxiety Inventory applied to college students: factor analysis and relation to the Beck Depression Inventory. *Brazilian Journal of Medical and Biological Research*, *34*(3), 367-374. https://doi.org/10.1590/s0100-879x2001000300011

Andrich, D., Sheridan, B. & Luo, G. (2010*). RUMM2030: Rasch models for measurement*. RUMM Laboratory.

Andrich D., & Marais, I. A. (2019). *Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Springer.

Cavanagh, R. F., & Waugh, R. F. (Eds.). (2011). *Applications of Rasch measurement in learning environments research* (Vol. 2). Springer Science & Business Media. https://doi.org/10.1007/978-94-6091-493-5.

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychological bulletin*, *139*(4), 735–765. https://doi.org/10.1037/a0030737

Cooper, A. J., Perkins, A. M., & Corr, P. J. (2007). A confirmatory factor analytic study of anxiety, fear, and behavioral inhibition system measures. *Journal of Individual Differences, 28*(4), 179–187. https://doi.org/10.1027/1614-0001.28.4.179

Cupani, M., Morán, V. E., Ghío, F. B., Azpilicueta, A. E., & Garrido, S. J. (2020). Psychometric Evaluation of the Big Five Questionnaire for Children (BFQ-C): A Rasch Model Approach. *Journal of Child and Family Studies, 29*(9), 2472-2486. https://doi.org/10.1007/s10826-020-01752-y

Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences.* Routledge/Taylor & Francis Group.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095.

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: an introduction and illustrative example. *International journal of nursing studies*, *46*(3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health technology assessment*, *13*(12). https://doi.org/10.3310/hta13120

Leibovich de Figueroa, N. (1991). Ansiedad. Algunas concepciones teóricas y su evaluación [Anxiety: Some theoretical concepts and their evaluation]. En M. M. Casullo, N. Leibovich de Figueroa & M. Aszkenazyi (Eds.), *Teoría y Técnicas de Evaluación Psicológica* (pp. 123-155). Psicoteca.

Lundgren-Nilsson, Å., Tennant, A., Jakobsson, S., Simrén, M., Taft, C., & Dencker, A. (2019). Validation of Fatigue Impact Scale with various item sets–a Rasch analysis. *Disability and rehabilitation*, *41*(7), 840-846. https://doi.org/10.1080/09638288.2017.1411983

Martínez, R., & Sewell, K.W. (2000). Explanatory style in college students: Gender differences and disability status. *College Student Journal, 34*(1), 72-78.

National Institute of Statistics and Censuses. (2015). Estimaciones de población por sexo, departamento y año calendario 2010 – 2025 [Population estimates by sex, department and calendar year 2010-2025]. Accessed June 30, 2021. https://www.indec.gob.ar/ftp/cuadros/poblacion/proyeccion_departamentos_10_25.pdf

Nilsson, L. A., & Tennant, A. (2011). Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. Journal of Rehabilitation Medicine, 43(10), 884–891. https://doi.org/10.2340/16501977-0871.

Pallant, J., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*, 1–18. https://doi.org/10.1348/014466506X96931.

Parkitny, L., McAuley, J. H., Walton, D., Costa, L. O. P., Refshauge, K. M., Wand, B. M., Di Pietro, F., & Moseley, G. L. (2012). Rasch analysis supports the use of the depression, anxiety, and stress scales to measure mood in groups but not in individuals with chronic low back pain. *Journal of Clinical Epidemiology, 65*(2), 189–198. https://doi.org/10.1016/j. jclinepi.2011.05.010.

Pellicciari, L., Piscitelli, D., Basagni, B., De Tanti, A., Algeri, L., Caselli, S., Ciurli, M. P., Conforti, J., Estraneo, A., Moretta, P., Gambini, M. G., Inzaghi, M. G., Lamberti, G., Mancuso, M., Rinaldesi, M. L., Sozzi, M., Abbruzzese, L., Zettin, M., & La Porta, F. (2020). 'Less is more': validation with Rasch analysis of five short-forms for the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs). *Brain injury*, *34*(13-14), 1741–1755. https://doi.org/10.1080/02699052.2020.1836402

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research.

Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*, 205–231.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Spielberger, C. D., Vagg, P. R., Barker, L. R., Donham, G. W., Westberry, L. G. (1980). The factor structure of the State-Trait Anxiety Inventory. In Sarason, I. G., Spielberger, C. D. (Eds.)*, Stress and anxiety (Vol. 7, pp. 95-109)*. Hemisphere.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis Rheum, 57*(8), 1358–1362. https://doi.org/10.1002/art.23108

Tennant, A., & Pallant, J. F. (2006) Unidimensionality Matters. *Rasch Measurement Transactions, 20*, 1048-1051.

Tennant, A., & Pallant, J. F. (2012). The root mean square error of approximation (RMSEA) as a supplementary statistic to determine fit to the Rasch model with large sample sizes. *Rasch Measurement Transactions, 25*, 1348–1349.

About the Authors

Sebastián J. Garrido. Psychologist, PhD student (National University of Cordoba, Argentina), is a researcher of the Institute of Psychological Research with a doctoral scholarship of the National Scientific and Technical Research Council (CONICET). His field of research is the effect of psychosocial risk in externalizing and internalizing maladaptive behavior, personality, mental health, the study of psychometric properties of psychological measures, and developmental of Computerized Adaptative Testing (CAT).

Ana E. Azpilicueta. PhD (University of Palermo, Buenos Aires, Argentina). Her current field of research is the longitudinal evaluation of the career self management model of social cognitive career theory. Her research interests are related to self-efficacy beliefs, contextual factors, and how they interact with the cognitive variables that mediate the career decision-making process.

Valeria E. Morán. PhD (National University of Cordoba, Argentina) is a researcher of the National Scientific and Technical Research Council (CONICET). She is a full professor of statistics in psychology and social sciences research methodology at the University Siglo XXI (Argentina). Her scientific career has focused mainly on the construction of psychometric tests that can be used in the field of educational and clinical psychology.

Marcos Cupani. PhD (National University of Cordoba, Argentina) is a full professor (tenure) of psychometric techniques at the National University of Córdoba (UNC) and a researcher of the National Scientific and Technical Research Council (CONICET). His scientific career has focused mainly on identifying individual differences (personality traits, self-efficacy beliefs, aptitudes, among others) that allow predicting academic performance and other variables related to vocational behavior. He is also interested in the construction of psychometric tests that can be used in the field of educational psychology.

Corresponding Author's Contact Address [TOP]

Instituto de Investigaciones Psicológicas,
IIPSI, Unidad Ejecutora CONICET.
Córdoba, Argentina.
Email: sebajgarrido@unc.edu.ar