Article

Measuring State and Trait Anxiety: An Application of Multidimensional Item Response Theory

Leonardo Carlucci [1,*] , **Marco Innamorati** [2,*] , **Melissa Ree** [3] , **Giorgia D'Ignazio** [4] **and Michela Balsamo** [4]

1   Learning Science Hub, University of Foggia, 71121 Foggia, Italy
2   Department of Human Sciences, Università Europea di Roma, 00163 Rome, Italy
3   School of Psychological Science, University of Western Australia, Perth 6009, Australia
4   Department of Psychological, Health and Territorial Sciences, University of "G. d'Annunzio" Chieti-Pescara, 66100 Chieti, Italy
*   Correspondence: leonardo.carlucci@unifg.it (L.C.); marco.innamorati@unier.it (M.I.)

Abstract: The State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) is a widely used measure of state and trait anxiety. Within the Classical Testing Theory model, consistent findings provide support for its multidimensional factor structure, discriminant, convergent, and nomological validity, as well as age and gender invariance, across healthy and clinical samples. Nevertheless, some issues regarding STICSA dimensionality and item-scale composition remain unresolved (e.g., both bifactor and two-factor models were found to fit data equally well). The goal of this study was to investigate the STICSA's dimensionality within the Item Response Theory, and to assess the tenability of the bifactor model as a plausible model over the multidimensional model. The sample consisted of 3338 Italian participants (58.21% females; 41.79% males) with an average age of 35.65 years (range: 18–99; SD = 20.25). Both bifactor and two-correlated dimensions of the STICSA scales were confirmed to fit data by applying the multidimensional Item Response Theory (mIRT). While the bifactor model showed better fit indices, the multidimensional model was more accurate and precise (0.86–0.88) in estimating state and trait latent anxiety. A further comparison between multidimensional item parameters revealed that the multidimensional and bifactor models were equivalent. Findings showed that the STICSA is an accurate and precise instrument for measuring somatic and cognitive symptomatology dimensions within state and trait anxiety. The use of the state/trait total score requires special attention from the clinicians and researchers to avoid bias in the psychodiagnostic assessment.

Keywords: anxiety; psychometrics; somatic; cognitive; assessment; bifactor

1. Introduction

Anxiety has a hypothetical multidimensional nature and multiple symptomatic manifestations [1,2]. Anxiety is multidimensional as it can be divided into different categories, including trait and state anxiety and both cognitive and somatic components (e.g., Martens, Vealey [3] proposed the multidimensional anxiety theory). This is a challenge to researchers and clinicians interested in the assessment of this complex phenomenon [4]. The main concern is the dimensionality of anxiety [1,2]. Similar to other psychological states, anxiety is considered to be composed of state and trait dimensions. It is possible to differentiate between transient emotion that varies in duration and is characterized by observable symptoms, like worry, tension, nervousness, and arousal of the autonomic nervous system (i.e., state anxiety), as well as a more enduring unobservable tendency to typically respond with high levels of anxious apprehension to perceived threats (i.e., trait anxiety) [5,6].

State anxiety has been conceptualized as an emotional response experienced in a limited period, depending on the anticipation of a (real or perceived) threatening stimulus, and is variable in duration [7]. On the other hand, trait anxiety has been viewed as an individual difference regarding an emotional response [8]. It is relatively stable in time

and corresponds to neuroticism [9] and negative or low emotionality [10], is a risk factor for the development of anxiety disorder [11], and is comparable to anxiety sensitivity [12]. Individuals with high state anxiety may experience increased levels of apprehension, fear, and physiological arousal [13]. Conversely, individuals high in trait anxiety become more state anxious (in terms of duration, frequency, and intensity of the episodes of state anxiety) in critical contexts than those low in trait anxiety [7].

Within the assessment of self-reported anxiety, the state–trait distinction has been captured by distinct instructions for the state and trait scale (e.g., "generally" vs "right now") by many authors [14,15]. Both state and trait anxiety have somatic and cognitive elements, creating four distinct components of anxiety: somatic state anxiety, somatic trait anxiety, cognitive state anxiety, and cognitive trait anxiety [13]. Clinicians and researchers have devoted much attention to the differentiation between somatic and cognitive anxiety [16]. From the clinical point of view, the symptoms of anxiety involve a broad range of cognitive, physical, and emotional aspects (e.g., "numbness," "unsteadiness," and "feeling hot") that may include worry, intrusive thoughts, and lack of concentration [17]. This somatic/cognitive distinction might allow for a more fine-grained assessment of anxiety and facilitating clinical treatments to be tailored to the specific and predominant manifestation of anxiety symptomatology (e.g., physiologically oriented relaxation vs cognitively oriented) [16].

A controversial issue in assessing anxiety concerns its overlap with depression [18–20]. In line with the tripartite model of anxiety and depression [21], a negative effect (e.g., fear, anger, guilt) is associated with both anxiety and depression; the lack of positive effect (e.g., feeling tired) is related to depression, whereas physiological hyperarousal (e.g., trembling, dizziness, shaking) is associated with anxiety [22]. Thus, the overlap between anxiety and depression may lead to misdiagnosis, a fairly frequent problem in clinical settings [23]. To this end, the State–Trait Inventory for Cognitive and Somatic Anxiety (STICSA) [17] represents a multidimensional measure of state/trait and somatic/cognitive anxiety, which holds strong discriminant power with depression [24–26].

Previous research supported its multidimensional factor structure (of both state and trait STICSA scales, with each including somatic and cognitive dimensions) [27]. Additionally, it has displayed gender and age invariance, as well as good convergent validity with concurrent anxiety measures and sound internal discriminant validity as determined in large non-clinical samples [27–32]. Whilst the STICSA state and trait scales have been recently reaffirmed as independent measures each composed of somatic and cognitive dimensions [27], the STICSA's dimensionality is still an open issue [33,34]. In detail, fitting both confirmatory (CFA) and exploratory structural equation models (ESEM) to STICSA scores, Styck and colleagues [33] showed that the oblique two-factor and bifactor models fit their data equally well for the separate state and trait forms of the STICSA. They also concluded that cognitive and somatic factors were not equally robust and that STICSA items appear to measure a mixture of both latent somatic and cognitive anxiety. Styck, Rodriguez [33] highlighted that "at least three items appear to meaningfully load onto both somatic-cognitive factors", which complicates STICSA score interpretation (p. 21). The items "feel agonized over problems" (#3), "trouble remembering things" (#11), and "feel trembly and shaky" (#8) were found to cross-load onto both somatic and cognitive dimensions. However, this datum is trivial since the magnitude of these cross-loadings was below the recommended level of = 0.30 [35] on the non-target dimension [33]. In addition, the differences among the primary and alternative factor loadings were found to be above the recommended cut-off of 0.20 [36], except for item #11.

Taken together, these results were not surprising, since under certain statistical conditions, the correlated factors and bifactor models were covariance matrix-nested [37], and resulted in being statistically indistinguishable [38]. In this case, the implied covariance matrix of the STICSA-State/Trait cognitive-somatic correlated factors model can be perfectly reproduced by the bifactor model. In addition, it is well established that the bifactor model has a high fitting propensity (the "probifactor bias" [39]). Equally, cross-loadings

might arise because of the similar content of indicators when measuring closely related but not identical latent variables [40]. Furthermore, cross-loadings' magnitude is strictly related to the sample size and the type of the rotation applied in factor analysis [41].

Additionally, Styck, Rodriguez and Yi [33] affirmed that "STICSA state-trait cognitive and somatic anxiety composite scores do not purely measure latent somatic and cognitive anxiety" both in past and recent studies [17,27,34,42], and "that researchers must make a choice between competing alternatives that is appropriate for their particular study aims" (p.22). As a result, the authors did not provide clear guidance about the dimensionality of the STICSA, suggesting to directly model data complexity through a more flexible and exploratory approach, such as the ESEM. Indeed, the main core of their study shifted from comparing the competing models (bifactorial vs two-factor model) to the impossibility of distinguishing somatic from cognitive items domain in each state and trait scale. Despite the complexity of the data and the use of multiple techniques that fall within the classical testing theory (CTT) framework (i.e., the use in tandem of CFA and ESEM [33]), some issues regarding STICSA dimensionality and item-scale composition remain unresolved. Firstly, if STICSA items measure a nonnegligible mixture of both latent somatic and cognitive anxiety, each state and trait scale should be best represented by a single latent dimension or by the global dimension that arises from the bifactor model. Secondly, the computation of a total state/trait scale score would result in an accurate and precise estimation of anxiety. In addition, if fit indices cannot be used to select championed competing models [33], how can we assess the tenability of a somatic and cognitive specific dimension and its items in a specific model compared to others? A different psychometric approach to the STICSA dimensionality assessment would lead to interesting results.

Hence, a step forward in the analysis of the STICSA state and trait scale could be represented by the application of modern psychometrics approaches, i.e., the Item Response Theory (IRT). One of the most important differences between CTT and IRT is that in CTT, one uses a common estimate of the measurement precision that is assumed to be equal for all individuals. Differently, in IRT, the measurement precision depends on the latent-attribute value (θ) [43]. The IRT was also found to be superior to CTT in individual change detection (especially when tests included at least 20 items) [43]. Next, assessing dimensionality, addressing items' local independency issue, and estimating item parameters and information (e.g., difficulty and discrimination) under IRT assumptions could lead to important benefits in health/clinical assessments [44,45]. The IRT approach is particularly useful for the development and refinement of clinical measures since it allows the researcher to place the abilities of both respondents and items along a single continuum and measure them with common metrics. For example, IRT can capture meaningful nuances in clinical test accuracy since clinical measures are not equally reliable for all intervals of latent-attribute distributions. Again, IRT could allow clinical psychologists to replace total scores with more accurate IRT-based latent-attribute estimates (test scoring) [45].

However, no research has been conducted to assess STICSA dimensionality and properties using IRT models. The IRT models have always been conceived and estimated as unidimensional (or one-dimensional) models (e.g., the Rasch model) due to their attractiveness and measurement properties. However, unidimensional models are often simple and fail to fit the data when more items measure multiple latent traits simultaneously [46]. In addition, clinical instruments were found to provide unusual results about item parameters in clinical applications of IRT models, resulting in violations of the unidimensional and invariance IRT assumptions (for analogy, see the debate in intelligence using Jensen's method; i.e., [47]). In the latter case, the estimation of a multidimensional IRT (mIRT) model is more appropriate [48]. Far from unidimensional IRT models, the estimation of multidimensional IRT models (e.g., bifactor mIRT; two- or three- dimensional mIRT models) may help to address the complex items issue, which involve multiple latent traits [48].

For example, estimating the STICSA bifactor mIRT model could allow for the measurement of a global anxiety dimension (technically labeled as latent-trait or latent-attribute [49]), along with the specific dimensions (e.g., somatic and cognitive), each representing distinct

underlying concepts. Compared to the bifactor model estimated within a CTT framework, the bifactor mIRT model could provide full-information item analysis estimating item parameters. Thus, the mIRT models may represent a feasible methodological approach that allows to evaluate complex psychological constructs that might be understood as a combination of nested sub-scale components (compensatory factors) or a more general construct (non-compensatory factors) [48]. On the other hand, using the general dimension slope parameter to interpret the relationship between the item and its dimension may be confusing and misleading, since the probability of response in one dimension is conditioned by the other dimensions [50].

Based on the above, our main goal was to assess the tenability of the bifactor model as a plausible model that may explain each STICSA state/trait scale over the two-factor model [33]. Next, a comparison at the item parameter level was made by estimating the discrimination and difficulty indices. We examined the dimensionality of the STICSA via IRT since recent advancements, such as the multidimensional IRT, may aid researchers in disentangling the debate on its structure and items functioning, as well as to provide suggestions for improving the psychometric qualities of this measure.

2. Materials and Method

2.1. Participants

The sample was recruited through advertisements in many Italian cities; the majority of the sample was taken from a published database [27] to which new datasets from unpublished studies have been added (N = 256). The STICSA was administered into the Italian language by licensed psychologists. Socio-demographic variables (such as age, gender, and education) were also collected. Inclusion criteria were ages from 18 to 99 years and the ability to complete self-administered questionnaires. Signed informed consent was obtained before the administration. The ethics committee of the Department of Psychological Sciences, Health and Territory, University of Chieti, Italy, approved the study.

2.2. Measures

The State–Trait Inventory for Cognitive and Somatic Anxiety (STICSA; [17]) (for the Italian version, see [27,31,32]) is a 21-item measure designed to assess cognitive and somatic symptoms, both on state and trait. In the state anxiety scale, they rate how they feel at the moment of assessment (from 1- "not at all" to 4- "very much"), whereas in the trait anxiety scale, participants rate how often a statement is true in general (from 1- "almost never at all" to 4- "almost always").

2.3. Statistical Analysis

Preliminarily, IRT assumptions of uni-dimensionality of the STICSA state and trait were assessed using non-parametric IRT Mokken analysis [51,52]. All analyses were conducted using the mIRT package in R [53]. The Automated Item Selection Procedure (AISP) algorithm implemented in the Mokken package of R, using the recommended range of values of c = 0.3–0.5, has been used to partition a set of items (or a set of unscalable items) into Mokken scales [54,55]. The Mokken scale is defined by a set of dichotomously (e.g., yes/no) or polytomous (e.g., 1–4 Likert type response scale) scored items for which all inter-item covariances are positive and scalability coefficient (Hi for a single item/H for the total scale) values < 0.4 identify weak scalability; values between 0.4 and 0.5 are evaluated as moderate, and values > 0.5 as strong scalability [54]. Next, local independence to test the assumption of residual relationship amount item responses, that are not accounted for by the unidimensional model, was assessed by the standardized LD-$\chi 2$ statistic [56]. Large values of standardized LD residuals (equal to |10| or greater) reflected LI issues. LD-$\chi 2$ statistics were calculated using the residuals function in the mIRT package of R. A violation of both uni-dimensionality and local independence assumptions determined the non-appropriateness of a single factor model and the need for a hierarchical model (and thus a multidimensional approach).

According to the STICSA polytomous response format, the graded response model (GRM) and its multidimensional logistic model (mGRM) [48,57] was applied to evaluate the monotonicity of item response function (e.g., estimating difficulty and discrimination in the item parameters) on model(s) selected. In order to test the adequacy of the models, we computed the C2* fit statistic [58] available using the M2 function implemented in the mIRT package. The C2* is a limited-information goodness of fit test statistic for ordinal IRT models. Following the debate in literature on the STICSA, a series of competing models were also addressed (described below). The goodness of fit of the GRM models was also evaluated based on several fit statistics, such as the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR), that are strictly related to those in confirmatory factor analysis [59]. The −2*Likelihood (−2LL; which is distributed as Chi-square with degree of freedom) with the provided Aikake Information Criterion (AIC) [60] and Bayesian Information Criterion (BIC) [61] were used to select the most parsimonious model among the competing ones. The model with the lower −2LL and AIC/BIC score is expected to strike a superior balance between its ability to fit the data set and its ability to avoid over-fitting the data set.

The item diagnostic was assessed using the signed $\chi2$ test (S-$\chi2$) [62]. A significance test ($p < 0.05$) was used under the null hypothesis to select misfitting items. Given that large samples and a large family of tests tend to yield significant $\chi2$ values, the Benjamini and Hochberg [63] procedure was applied to adjust the obtained p-values for the number of tests in the family and to control for experimenter-wise error.

For the unidimensional model, we estimated single slope (discrimination) parameters and three intercept (c) parameters for each 4-category STICSA item. Discrimination parameter values theoretically range from $\pm\infty$ (0.5 to 3.0 is a reasonable range). Intercept parameter (b) values ranging from −3 to + 3 are in the typical range [64]. For hierarchical and multidimensional IRT models, Multidimensional Item Discrimination (MDISC) and Multidimensional Difficulty Index (MDIFF) were also computed, transforming the estimated slope (item discrimination) and intercept parameters (category threshold). Values of MDISC >0.65 indicate discriminative items, and high positive values of MDIFF (>0.5) indicate difficult items [49,65]. Test information function (TIF) and related standard errors of measurement (SE) indicating the precision of the whole test were also estimated to determine at what level of the STICSA state and trait provide the most information.

Models Tested

First, a unidimensional structure (Model 1-uni-GR) was examined with all 21 items defining one latent dimension for the separate trait and state forms of the STICSA. The single factor model was evaluated to assess if a unified concept underlying the data exists, as well as to evaluate the suitability of the multidimensional IRT assumption.

Next, Model 2 (multi-GR) tested the original conceptualization of the STICSA (i.e., the oblique somatic-cognitive dimensions). This model is in line with recent studies suggesting that each STICSA form is well-represented by two correlated factors—cognitive and somatic anxiety [17,26,27,33,42,66].

Model 3 tested a bifactor mIRT model (bifac-GR) for the separate state and trait forms of the STICSA. In the bifactor model, each STICSA state/trait item was restricted to define both a global anxiety factor (i.e., global state anxiety; global trait anxiety) as well as a domain-specific factor (somatic or cognitive) [33,67]. In the bifactor mIRT model, somatic and cognitive domains for each state and trait scale were treated as orthogonal (uncorrelated), and item discrimination was estimated only in the dimension they belonged to, with the discrimination parameters linking items to the remaining dimensions being constrained to zero.

Besides the above, a restricted bifactor IRT measurement model was modeled to address IRT assumption violations (i.e., local independency), which can turn up in a biased estimation of item parameters and/or an overestimation of test reliability [68]. A trace

lines transformation (also named "marginal slopes") [50,69] was applied to the bifactor mIRT model response function weighting by the normal distribution and integrating out the specific dimension and other general dimensions.

Next, we applied a pairwise comparison strategy of item fit and parameter estimates and their total information/precision across (a) the uni-GR vs bifac-GR item; and (b) the multi-GR vs bifac-GR. The first comparison was done in order to assess the usefulness of a general bifactor model, as well as the presence of LD issues that reflected the presence of a multidimensional structure of the STICSA state/trait scales.

Next, given the high degree of overlap among nested models, we compared multidimensional item parameters of multi-GR and bifac-GR models in order to assess the presence of meaningful differences among the models' multidimensional parameters and the compensatory/non-compensatory nature of the STICSA state and trait scales.

3. Results

3.1. Sample Characteristics

Recruited participants were 3338 Italian adults ranging from 18 to 99 years old, including 1943 females (58.21%) and 1395 males (41.79%), most of whom were undergraduate students. The mean age for the entire sample was 35.65 (SD = 20.25) years, compared to increasing Italian national average age (46 years) (ISTAT, 2021). The mean age for men was 37.78 (SD = 20.68) years, and 34.10 (SD = 19.79) years for females. Among these, 248 (7.4%) respondents held a primary school diploma, 461 (13.8%) a lower secondary school diploma, and 1966 (64.9%) a high school diploma, while 247 (10.4%) held a higher education (bachelor/master/doctorate) and the remaining (10.9%) declared no education or preferred not to answer. Concerning marital status, 2178 (65.2%) participants were unmarried/single, 801 (24%) were married, 108 (3.6%) were divorced/separated, 76 (2.3%) were cohabiting, 123 (3.7%) were widowed, and 52 (1.6%) did not answer.

3.2. Item Response Theory Assumptions

Preliminarily, we examined the frequency distribution of all item responses to identify items with high percentages at the extreme ends of the response scale. Such items could affect the subsequent stages of the analysis, especially the analyses conducted within the IRT framework [70]. The results showed that no items had a mean <0.5 and >3.5. From the total sample, we removed N = 13 cases with missing completely at random (MCAR) data on all the STICSA state and trait items. Cases with missing completely at random values were also retained in the IRT analyses. The final database was composed of 3325 observations.

The STICSA state and trait scale items were submitted for Mokken analysis to test the uni-dimensionality assumption. The AISP revealed that all the STICSA state and trait items loaded on multiple latent dimensions, respectively. The inter-item covariances were found to be positive, and all the item scalability coefficients (Hi) ranged from 0.311 to 0.428 (weak) for the state, and from 0.295 to 0.404 (weak) for the trait scale; hence, the second criterion of the Mokken scale was not satisfied. The scalability coefficient for the entire state and trait scale, equal to 0.370 and 0.343, showed a weak scalability. As expected, the assumption of uni-dimensionality was not met for the 21 items of the state and trait scales.

Next, we assessed the tenability of local independence assumptions for each item pair of the state and trait scale, taking in to account the three models tested (see Appendix A). Large positive (+) and negative (−) standardized LD values for the models tested (uni-GR, multi-GR, and bifac-GR models) were summarized in Table 1. A careful examination of the LD values exhibited a pattern of 8/5 large positive LD among item pairs within the somatic area and 8/4 large positive LD among item pairs within the cognitive area regarding the state and trait scale, respectively. Far from the positive values, which can inflate the item slope parameter, the negative LD values can be ignored since they have a tendency to underestimate the slope parameter [71]. Negative LD values are usually detected for self-rating scales that do not measure best performance. The unmodeled covariation,

particularly within the cognitive and somatic domains of state anxiety, suggested that a specific number of latent variables explaining the item response patterns or properly accounting for all covariations may not be outlined. This suggested that items for the state and trait scale may be better modelled by models defined as "multidimensional", such as a two-correlated factors model or bifactor model.

Table 1. Results from uni-GR, multi-GR, and bifac-GR models fit to the 21-item STICSA-STATE.

(a)			
	Uni-GR	**Multi-GR**	**Bifac-GR**
# of positive LD pairs flagged	16	8	4
# of negative LD pairs flagged	3	2	2
# of parameters	84	85	105
−2LL	−60,047.66	−58,793.31	−58,477.48
BIC	120,776.5	118,275.9	117,806.4
AIC	120,263.3	117,756.6	117,165
C2 (*df*)	6505.643 (189) ***	2499.159 (188) ***	1858.838 (168) ***
RMSEA	0.100	0.061	0.055
TLI	0.916	0.969	0.974
CFI	0.984	0.972	0.979
SRMSR	0.074	0.047	0.043
Precision (marginal reliability) [θ −3;3]	0.89	0.86–0.88	0.087–0.62–0.42
(b)			
	Uni-GR	**Multi-GR**	**Bifac-GR**
# of positive LD pairs flagged	9	6	3
# of negative LD pairs flagged	1	0	0
# of parameters	84	85	105
−2LL	−69,934.32	−69,002.42	−68,742.61
BIC	140,549.8	138,694.1	138,336.7
AIC	140,036.6	138,174.8	137,695.2
C2 (*df*)	5253.537(189) ***	2673.871(188) ***	2033.147(168) ***
RMSEA	0.090	0.063	0.058
TLI	0.919	0.960	0.966
CFI	0.927	0.964	0.973
SRMR	0.065	0.047	0.041
Precision (marginal reliability) [θ −3;3]	0.90	0.86–0.88	0.86–0.40–0.65

Note. −2LL = −2 log likelihood or deviance statistic; BIC = Bayesian information criterion; AIC = Akaike information criterion; C2 = limited-information goodness-of-fit statistic; RMSEA = root mean square error of approximation based on C2. Number of LD possible = $j(j − 1)/2 = 21(21 − 1)/2 = 210$. Correlation among latent traits in the multi-GR model was 0.76/0.73. *** $p < 0.001$.

Therefore, a fit of the item response data with a multi-GR model and bifac-GR model was revealed. Again, by a detailed examination, unmodeled local dependency was found again in the LD values for the multidimensional model. However, it was isolated to three and four items in the somatic content area with five and two items in the cognitive content area for the state and trait scale, respectively. Fortunately, the adoption of a multi-GR model would appear to lead towards results that facilitate the local independence assumption. Whether or not the bifac-GR model was fitted to the item response data to better account for the unique variance attributed to each content area, the unmodeled LD was significantly reduced. The only LD values included items #1 and #16 for both the state and trait scale, with three other items (items #2, #12, and #10 for the state scale), and item #9 for the trait scale. The bifac-GR model resulted in the enhancement of the modelling of the item response covariations. This result leads to unique variability to be modelled for each content area after accounting for a general latent trait.

3.3. Models-Data Fit and Comparison

Table 1 summarizes the results of the model comparison and displays fit indexes for the several IRT models in the comparison. All indices for determining the best-fitting STICSA state-trait model seemed to agree with the selected model, which was the bifac-GR model. However, the goodness of fit for the multi-GR model indicated an acceptable-to-good fit for the sample, strictly close to the bifactor model (see CFI and TLI indices above the >0.95 threshold and a RMSEA $\leq$0.05 for excellent fit). Nevertheless, the deviance statistics ($-2LL$) and the absolute difference in AIC and BIC ($>|9|$) [72] provided support for the superiority of the bifactor solution over the multi-GR model. The goodness of fit of the uni-GR model confirmed a less good model fit for STICSA state-trait scales (see LD values).

3.4. Comparison of Item Fit and Parameters

The uni-GR model and bifac-GR model parameter estimates for the STICSA state and trait items, respectively, were summarized in Table 2. We examined fit at the item level for the uni- and bifac-GR models controlling for Type I error using a Benjamini-Hochberg adjustment. Concerning the uni-GR and bifac-GR models, none of the state scale items was identified as misfitting. On the contrary, item #14 ($p = 0.004$) for the trait scale was identified as misfitting in both models, using the same criteria.

Table 2. Unidimensional and bifactor model item slope estimates for the 21-item STICSA state and trait.

| Item | Uni-GR a | STICSA—State Bifac-GR Conditional | | | Bifac-GR Marginal | | | Item | Uni-GR a | STICSA—Trait Bifac-GR Conditional | | | Bifac-GR Marginal | | |
		a_g	a_{S1}	a_{S2}	a^*_g	a^*_{S1}	a^*_{S2}			a_g	a_{S1}	a_{S2}	a^*_g	a^*_{S1}	a^*_{S2}
STIC_S_1	1.196	1.017	0.979		0.881	0.840		STIC_T_1	1.020	1.205	−0.392		1.174	−0.320	
STIC_S_2	1.528	1.365	1.415		1.049	1.103		STIC_T_2	1.264	1.402	−0.078		1.401	−0.060	
STIC_S_6	1.649	1.442	1.574		1.058	1.200		STIC_T_6	1.485	1.859	0.608		1.750	0.410	
STIC_S_7	1.799	1.659	1.636		1.195	1.171		STIC_T_7	1.577	2.543	1.501		1.906	0.834	
STIC_S_8	2.516	2.389	1.993		1.550	1.156		STIC_T_8	1.991	2.565	0.746		2.349	0.412	
STIC_S_12	1.339	1.143	1.063		0.969	0.882		STIC_T_12	1.285	1.514	−0.100		1.511	−0.075	
STIC_S_14	2.012	1.910	1.928		1.263	1.282		STIC_T_14	1.801	2.118	0.154		2.109	0.096	
STIC_S_15	1.280	1.074	1.018		0.921	0.861		STIC_T_15	1.406	1.659	−0.162		1.652	−0.116	
STIC_S_18	1.817	1.597	1.177		1.313	0.858		STIC_T_18	1.723	2.025	−0.390		1.974	−0.251	
STIC_S_20	1.055	0.914	0.591		0.863	0.521		STIC_T_20	0.897	1.056	−0.571		1.001	−0.485	
STIC_S_21	0.920	0.736	0.773		0.670	0.709		STIC_T_21	0.922	1.131	−0.528		1.080	−0.440	
STIC_S_3	1.980	2.207		0.148	2.207		0.090	STIC_T_3	1.691	1.443		1.332	1.443		1.266
STIC_S_4	1.715	1.816		−0.146	1.816		−0.100	STIC_T_4	1.409	1.173		1.026	1.173		0.973
STIC_S_5	1.551	1.646		0.314	1.646		0.228	STIC_T_5	1.307	1.066		1.023	1.066		1.007
STIC_S_9	1.897	2.882		−1.023	2.882		−0.546	STIC_T_9	1.669	1.431		1.253	1.431		1.161
STIC_S_10	1.534	1.933		0.601	1.933		0.408	STIC_T_10	1.426	1.182		1.242	1.182		1.274
STIC_S_11	1.298	1.191		0.143	1.191		0.117	STIC_T_11	1.030	0.965		0.335	0.965		0.296
STIC_S_13	2.256	3.932		−1.549	3.932		−0.659	STIC_T_13	1.919	1.671		1.360	1.671		1.181
STIC_S_16	1.643	1.883		0.555	1.883		0.381	STIC_T_16	1.352	1.127		0.914	1.127		0.852
STIC_S_17	2.063	2.909		1.146	2.909		0.615	STIC_T_17	1.900	1.651		1.411	1.651		1.260
STIC_S_19	2.108	2.597		0.820	2.597		0.466	STIC_T_19	1.964	1.722		1.387	1.722		1.189

Note. a = uni-GR slope; ag = conditional slope for STICSA state or trait general trait; aS1 − aS2 = conditional slopes for specific traits somatic/cognitive; a*g = marginal slope for STICSA state or trait general trait; a*S1 − a* − S = marginal slopes for specific traits somatic/cognitive.

Comparing the uni-GR with bifac-GR model solutions revealed a range from 0 to 0.99 with a mean absolute difference of 0.21 (SD = 0.29) for the trait scale, and a range from 0 to 3.04 of the absolute differences in the values of the item intercept parameters with a mean absolute difference of 0.39 (SD = 0.53) for the state scale, as Appendix B shows. In particular, the absolute value of the mean difference was larger for the third intercept at the extreme; that is, 0.36 (SD = 0.39) for the trait scale and 0.62 (SD = 0.70) for the state. The mean absolute difference was lower for the first intercepts, 0.14 (SD = 0.21) and 0.06 (SD = 0.05) for trait and state scales, respectively. On the other hand, the mean average absolute difference was elevated for the second intercept in the state scale [0.41(SD = 0.48)] compared to the trait scale [0.21(SD = 0.23)]. Indeed, by ignoring LD, the loss of precision created led to shrinkage in intercept values for extreme categories, as you can see in Appendix B. Inflating the conditional slopes of the general latent trait in the bifac-GR resulted in the conditional relationships among the specific and general latent traits [50]. Therefore, since a straight comparison among specific and general slopes is not recommended, a marginal

slope [69] was computed by using Stucky and Edelen [50,69] equations and comparing slopes between the two model solutions (uni vs. bifac model).

Concerning the state scale, a comparison of the general state conditional slopes (see Table 2) from the bifac-GR model solution shows 10 of the 21 item conditional slopes were larger than uni-GR model slopes. The average absolute difference was 0.33 (Min = 0.09, Max = 1.67). For the general trait conditional slope, the bifac-GR model solution shows 11 of the 21 item conditional slopes were larger than uni-GR model slopes (see Table 2). The average absolute difference was 0.28 (Min = 0.06, Max = 0.96). An inspection of the marginal slopes in Table 2 shows that the four items which discriminated more across respondents with respect to the general state dimension, in rank order, were items 13 (a13*g = 3.932), 17 (a17*g = 2.909), 9 (a9*g= 2.882), 19 (a19*g = 2.597), and 3 (a3*g = 2.207). Two items discriminate more across respondents with respect to the general trait dimension: items #8 (a8*g = 2.349) and #14 (a14*g = 2.109). A comparison across marginal and conditional slopes of the bifac-GR models showed that there is a slight difference in the magnitude only for those items which have slopes close to 0 on the somatic dimension of the state scale, and negligible difference only for somatic items of the trait scale.

However, the marginal slopes for the general state and trait scales were much smaller in size for items which had a relevant conditional slope on the specific somatic/cognitive dimensions (i.e., state items #8 [$a8_{S1}$ = 1.99], #14 [$a14_{S1}$ = 1.92], and #18 [$a18_{S1}$ = 1.77]; trait items #7 [$a7_{S1}$ = 1.50] and #8 [$a8_{S1}$ = 1.92]).

3.5. Comparison of TIFs and Marginal Reliability

Next, the total information functions were compared across the uni-GR and bifac-GR models for the STICSA state and trait scales. How well the construct is measured across any level of the trait continuum is the main information derived from using the TIFs. A straightforward comparison between the TIF from the unidimensional (uni-GR) model and the conditional bifac-GR (CTIF) model evidenced that the two functions seemed to overlap along all latent STICSA-state values (the conditional bifac-GR model was found to be more informative than uni-GR) (see Figure 1). Concerning the trait scale, the uni-GR was found to be more informative and included the conditional bifac-GR information that was slightly narrowed along the STICSA-trait values (see Figure 1). Based on the MTIF in Figure 1, the measurement precision is roughly constant for values on the state/trait continuum closest to 0, and between −0.5 and 0.5 with total information for the 21-item state/trait of about 15~14 (TIF with prior variance of 1—posterior total [test] information; 41~42 without prior variance of 1). The standard errors (SEEs; Figure 2) of bifac-GR state/trait scores for this interval were close at about 0.25~0.26 ($1/\sqrt{\text{Information}}$). This level of error around the scores converts to an IRT reliability value of about $1-\text{SEE}^2$, which was 0.93−0.94 in this interval, respectively, for the state and trait scales.

Additionally, when the marginal reliability (also named empirically, since we used an EAP estimator) of the unidimensional model (0.89 for state, 0.90 for trait) and the bifactor model (0.87 for state, 0.86 for trait) was contrasted, there was an inflation in score precision when LD was overlooked (see Table 1). This inflated effect was more pronounced in the trait scale. On the other hand, the estimation of the conditional bifac-GR model tended to overestimate state/trait reliability/precision scores along the right segments of the continuum (high state/trait anxiety). Next, the marginal reliability for the two-correlated model (multi-GR, 0.86–0.88 for state and trait, for somatic and cognitive factors), respectively, was found to be higher compared to the bifactor model ones (Bifac-GR, 0.62–0.42 for state, 0.40–0.65 for trait).
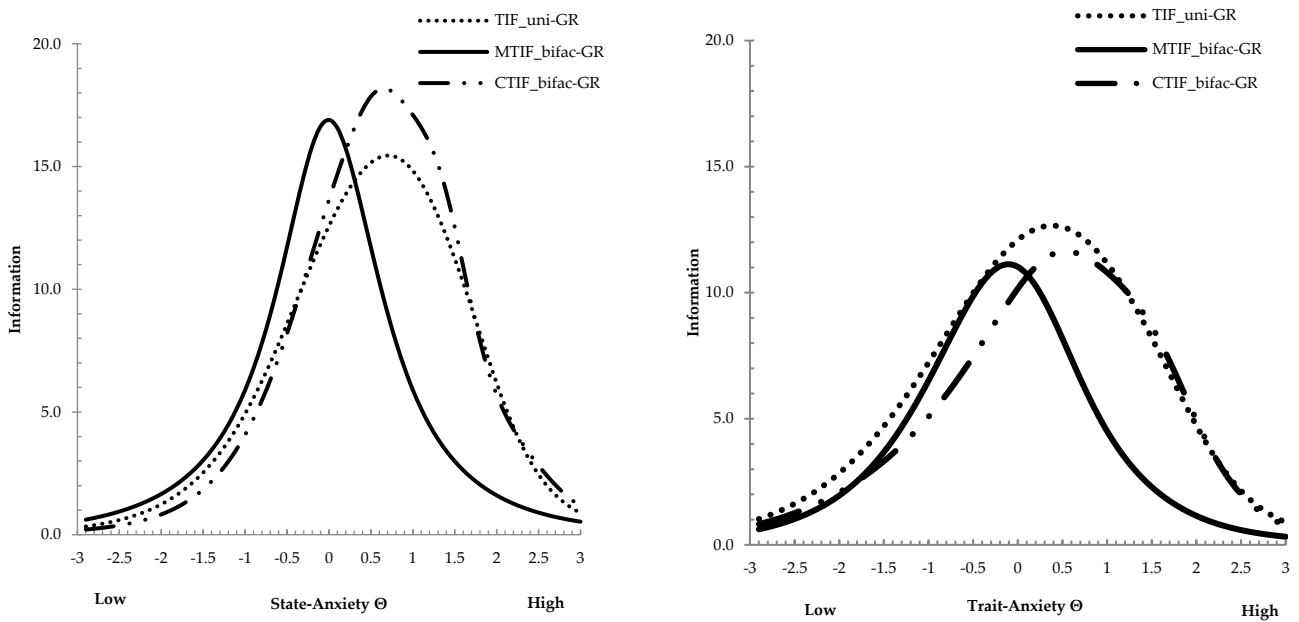
Figure 1. Total information function (TIF) for the state (**left**)/trait (**right**) STICSA. Note: TIF = unidimensional TIF; CTIF = conditional TIF fit by the bifactor GR model based on conditional slopes; MTIF = marginal TIF fit by the bifac−GR model based on marginal slopes. TIF with prior variance of 1 (posterior total [test] information).
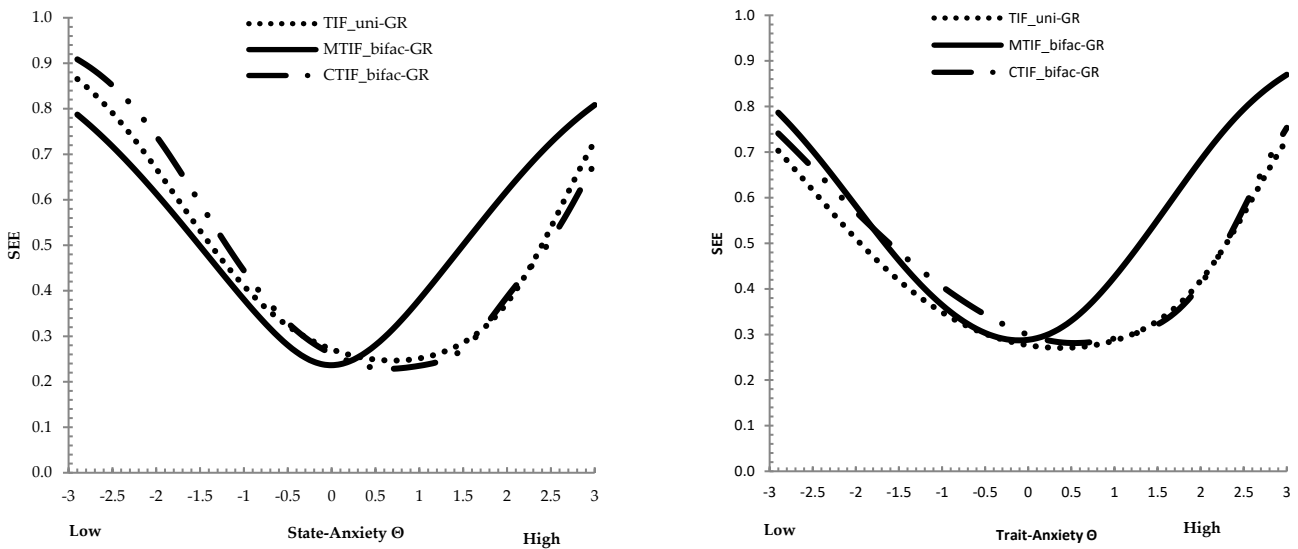
Figure 2. Standard Error of Estimate (SEE) for the state (**left**)/trait (**right**) STICSA. Note: TIF = unidimensional SEE; CTIF = conditional SEE fit by the bifactor GR model based on conditional slopes; MTIF = marginal SEE fit by the bifac−GR model based on marginal slopes.

3.6. Multidimensional Item Diagnostic

The mIRT analyses were conducted separately on the trait and state scales. Table 3 presents the multidimensional parameter estimates of the items, as well as the item diagnostics estimates for both the two-correlated factors (multi-GR) and the bifactor models. As found in the previous pairwise comparison, none of the state scale items were identified as misfitting at $p < 0.05$. On the contrary, item #14 ($p = 0.004$) for the trait scale was confirmed to misfit both multidimensional models.

Table 3. Multi-GR and bifac-GR item fit and multidimensional item parameters for the STICSA state scale.

(a)											
STICSA—State											
Bifac-GR						**Multi-GR**					
	MDISC	MDIFF$_1$	MDIFF$_2$	MDIFF$_3$	Itemfit *p* (fdr)		MDISC	MDIFF$_1$	MDIFF$_2$	MDIFF$_3$	Itemfit *p* (fdr)
STIC_S_1	1.412	0.261	1.396	2.885	0.504	STIC_S_1	1.413	0.268	1.403	2.891	0.573
STIC_S_2	1.966	0.186	1.371	2.311	0.599	STIC_S_2	1.924	0.194	1.390	2.342	0.546
STIC_S_6	2.135	0.831	1.735	2.533	0.614	STIC_S_6	2.084	0.843	1.757	2.567	0.609
STIC_S_7	2.330	0.465	1.405	2.337	0.145	STIC_S_7	2.293	0.474	1.420	2.360	0.128
STIC_S_8	3.112	0.935	1.626	2.293	0.599	STIC_S_8	3.145	0.941	1.632	2.300	0.701
STIC_S_12	1.561	0.649	1.752	2.919	0.467	STIC_S_12	1.565	0.655	1.758	2.924	0.389
STIC_S_14	2.713	0.654	1.537	2.282	0.321	STIC_S_14	2.658	0.663	1.554	2.306	0.128
STIC_S_15	1.480	0.596	1.847	2.875	0.374	STIC_S_15	1.484	0.601	1.852	2.878	0.268
STIC_S_18	1.984	0.813	1.719	2.666	0.609	STIC_S_18	2.008	0.817	1.719	2.662	0.633
STIC_S_20	1.088	1.086	2.082	3.154	0.488	STIC_S_20	1.095	1.089	2.083	3.150	0.526
STIC_S_21	1.067	0.648	1.945	3.091	0.145	STIC_S_21	1.069	0.653	1.948	3.094	0.128
STIC_S_3	2.211	−0.232	0.945	1.718	0.145	STIC_S_3	2.272	−0.222	0.944	1.708	0.144
STIC_S_4	1.822	0.462	1.505	2.446	0.488	STIC_S_4	1.822	0.469	1.510	2.455	0.473
STIC_S_5	1.676	0.052	1.323	2.253	0.145	STIC_S_5	1.696	0.059	1.324	2.248	0.114
STIC_S_9	3.058	0.380	1.206	2.006	0.145	STIC_S_9	2.161	0.421	1.336	2.258	0.131
STIC_S_10	2.024	−0.398	0.631	1.378	0.145	STIC_S_10	1.898	−0.403	0.658	1.426	0.114
STIC_S_11	1.200	0.037	1.611	2.972	0.145	STIC_S_11	1.217	0.044	1.604	2.952	0.114
STIC_S_13	4.226	0.571	1.306	1.931	0.567	STIC_S_13	2.467	0.648	1.471	2.200	0.796
STIC_S_16	1.964	−0.028	0.971	1.806	0.145	STIC_S_16	1.870	−0.024	0.998	1.854	0.114
STIC_S_17	3.127	0.027	0.981	1.761	0.479	STIC_S_17	2.526	0.031	1.049	1.881	0.268
STIC_S_19	2.723	0.124	1.070	1.797	0.437	STIC_S_19	2.456	0.131	1.117	1.872	0.276
(b)											
STICSA—Trait											
Bifac-GR						**Multi-GR**					
	MDISC	MDIFF$_1$	MDIFF$_2$	MDIFF$_3$	itemfit *p* (fdr)		MDISC	MDIFF$_1$	MDIFF$_2$	MDIFF$_3$	itemfit *p* (fdr)
STIC_T_1	1.267	−1.171	0.621	2.577	0.074	STIC_T_1	1.151	−1.239	0.664	2.757	0.055
STIC_T_2	1.405	−0.917	0.781	2.367	0.074	STIC_T_2	1.409	−0.910	0.785	2.369	0.055
STIC_T_6	1.956	0.160	1.270	2.334	0.999	STIC_T_6	1.802	0.174	1.325	2.433	0.968
STIC_T_7	2.953	−0.090	1.059	2.023	0.100	STIC_T_7	1.917	−0.097	1.227	2.365	0.042
STIC_T_8	2.672	0.528	1.450	2.217	0.044	STIC_T_8	2.444	0.550	1.499	2.288	0.016
STIC_T_12	1.517	0.204	1.477	2.735	0.999	STIC_T_12	1.490	0.211	1.498	2.772	1.000
STIC_T_14	2.123	0.501	1.537	2.514	0.004	STIC_T_14	2.154	0.504	1.534	2.508	0.003
STIC_T_15	1.667	0.264	1.619	2.744	0.102	STIC_T_15	1.631	0.271	1.641	2.786	0.117
STIC_T_18	2.062	0.381	1.499	2.506	0.796	STIC_T_18	1.891	0.397	1.558	2.623	0.770
STIC_T_20	1.200	0.138	1.453	2.803	0.044	STIC_T_20	0.950	0.168	1.713	3.332	0.037
STIC_T_21	1.249	0.331	1.695	2.846	0.696	STIC_T_21	1.024	0.380	1.943	3.286	0.630
STIC_T_3	1.964	−0.858	0.556	1.724	0.837	STIC_T_3	1.936	−0.858	0.565	1.741	0.691
STIC_T_4	1.558	−0.004	1.346	2.619	0.587	STIC_T_4	1.552	0.002	1.356	2.633	0.591
STIC_T_5	1.478	−0.322	1.067	2.353	0.308	STIC_T_5	1.464	−0.317	1.080	2.373	0.204
STIC_T_9	1.902	−0.051	1.061	2.138	0.074	STIC_T_9	1.898	−0.046	1.069	2.147	0.063
STIC_T_10	1.714	−1.056	0.235	1.345	0.117	STIC_T_10	1.646	−1.072	0.245	1.376	0.055
STIC_T_11	1.021	−0.482	1.429	3.012	0.187	STIC_T_11	0.974	−0.495	1.482	3.128	0.195
STIC_T_13	2.155	0.315	1.370	2.211	0.074	STIC_T_13	2.170	0.319	1.374	2.212	0.055
STIC_T_16	1.451	−0.526	0.760	2.023	0.074	STIC_T_16	1.448	−0.522	0.767	2.032	0.055
STIC_T_17	2.172	−0.351	0.822	1.914	0.074	STIC_T_17	2.179	−0.345	0.827	1.918	0.055
STIC_T_19	2.211	−0.154	0.914	1.842	0.385	STIC_T_19	2.215	−0.150	0.920	1.849	0.411

Note MDISC = Multidimensional Item Discrimination; MDIFF = Multidimensional Difficulty Index.

Concerning state anxiety, all the multidimensional parameter values for the multi- and bifac-GR models overlapped with a minimal difference in the MDISC average values of 0.179 (SD = 0.255) and negligible differences in the three intercepts (MDIFF$_1$ M $\pm$ SD= -0.011 ± 0.003; MDIFF$_2$ M $\pm$ SD = -0.027 ± 0.013; MDIFF$_3$ M $\pm$ SD = 0.044 ± 0.033). Five items (items #20, #21, #1, #12, #15) for state-somatic and two (items #5, #11) for state-cognitive factors demonstrated "moderate" to "high" multidimensional discrimination, according to Baker [49] and Hasmy [65], ranging from 1.069 (item #21) to 1.696 (item #5). The remaining five items showed "very high" discrimination parameters (MDISCR > 1.7).

Likewise, for the trait scale, a minimal difference was observed in the MDISC average values of 0.112 (SD = 0.057) and negligible differences in the three intercepts (MDIFF$_1$ M $\pm$ SD = -0.004 ± 0.014; MDIFF$_2$ M $\pm$ SD = -0.050 ± 0.035; MDIFF$_3$ M $\pm$ SD = 0.099 ± 0.080). Six items (items #20, 21, 1, 2, 12, 15) for somatic and five (items #4, 5, 10, 11, 16) for cognitive factors demonstrated "moderate" to "high" multidimensional discrimination, ranging from 0.950 (item #20) to 1.646 (item #10). The remaining five items showed "very high" discrimination parameters (MDISCR > 1.7). To sum up, model parameters highlighted the equivalence for both multi- and bifac-GR models into discrimination among high and low levels of state and trait anxiety.

4. Discussion

The STICSA is now a widely accepted measure of state and trait anxiety, although its dimensionality is a matter of interest among researchers [33,34]. Its attractiveness lies in its ability to measure both state and trait anxiety with a single measure, as well as their cognitive and somatic features. Far from the criticized Spielberg state-trait anxiety model [24,73,74], the STICSA displays good psychometric properties, including an excellent divergent validity. This feature represents an advantage for clinicians, enabling them to better discriminate between specific and common anxiety factors that share underlying processes with depression [75,76]. To date, addressing dimensionality issues represents a crucial step to score and interpret STICSA state and trait item values. Studies have relied on different conceptualization of STICSA scores, but a large part of them have supported two-dimension somatic-cognitive correlated factors for each state and trait scale [27,32,34,42].

Unlike past studies, strictly oriented to the classical psychometric approach, the present study aimed to disentangle the STICSA dimensionality issue by applying the multidimensional IRT approach. Multidimensional item response theory (mIRT) represents an extension over the classical IRT approach in which an additional vector of multiple personal attributes is included so that multiple underlying traits can be simultaneously measured [48,77]. Based on Styck, Rodriguez, and Yi [33]'s findings, a bifactor model was tested as a viable and reliable alternative to multidimensional models for assessing the STICSA state and trait scale dimensionality. In detail, information about fitting unidimensional, multi-GR, and bifac-GR models, testing pertinent IRT assumptions, and looking at model-data fit statistics was supplied. Indeed, that a model fits the data well overall compared to other models, but may poorly fit items, is usually observed [78]. Likewise, to interpret the bifac-GR model results, item parameters and TIFs were compared to the several models (following [50]). Parameter estimates and their multidimensional versions were computed to make meaningful comparisons across model parameter estimates (e.g., unidimensional vs. bifactor model; multidimensional vs. bifactor model) and select a championed model.

In line with Styck, Rodriguez, and Yi [33]'s model fit, results suggested that the bifac-GR model fit our data best when compared to the multi- and uni-GR. However, when we compared the IRT conditional/marginal and multidimensional parameters estimated and the information, we observed that the STICSA bifactor model did not offer as many advantages as expected. Yet, in the present study, the restricted bifactor model offered a way to deal with violations of LD (without removing them completely), and to improve the model's fit. It has been widely attested that a bifactor model is more robust compared to minor model misspecifications due more to unmodeled complexity than correlated factor

models. It has increased parametric complexity [79] when compared to correlated factor models. Likewise, larger models tended to fit data sets better than more parsimonious models, though there is a trade-off with generalizability when models overfit a data set [39,80]. In addition, probifactor bias occurs due to the functional form of the parameters estimated in the bifactor model (e.g., [80]; i.e., the "difficulty" of the estimated parameters). Models that estimate the same number of parameters can differ in functional form, and two models with different functional forms, but equivalent free parameters, will exhibit different fitting propensities.

In the present study, the importance of estimating marginal parameters and TIF over the conditional bifactor model was demonstrated. Discrepancy among conditional and marginal slopes emerged since the specific latent-attribute did not affect the probability of response to the item (they are principally reflective of the general trait and secondarily a specific trait) [50,69]. The conditional bifactor model seemed to account for the general state/trait dimension common to all STICSA items, but it failed to address additional specific somatic/cognitive dimensions. In particular, the conditional bifactor model failed to address common variance in cognitive group item contents in the state scale and cognitive group item contents in trait scale. However, this result was already known in the STICSA psychometrics literature [33]. Small factor loadings and zero or negative group factor variances were also commonly identifiable in several studies that aimed to address a 'general psychopathology' reflecting commonality among all forms of psychopathology, along with several narrower psychopathology group factors (most commonly internalizing; see [81]). However, these findings can have limitations from a practical/clinical point of view. For example, the observed total/subscale scores may reflect a mixture of general and group factor variance. Observed correlations might be inflated or attenuated, since they might reflect the influence of the general factor, the group factor, or both. As a results, clinicians might be led to an incorrect assessment and treatment planning during their practices, with harmful consequences [82,83].

Therefore, it becomes mandatory to discern among broad and narrow factors to assess whether factors reflect substantive constructs or artefacts. In line with this recommendation, the usefulness of the STICSA bifactor model in computing a pairwise comparison approach was evaluated. Utilizing the estimated marginal bifactor model and parameters (also restricted), the general dimension was of the most concern, and items were allowed to load in specific factors when necessary (to account for an excess of LD). This model was found to be more precise and informative than the unidimensional and conditional bifactor models. Not surprisingly, unidimensional IRT models show a lack of measurement precision in the estimate dimensionality issue when scales are composed of correlated latent traits [84]. The computation of marginal trace line transformation also emphasized the usefulness of testing IRT model assumptions within the context of classical bifactor modeling.

Next, the suitability of a multidimensional nature of each STICSA state/trait scale was secured via IRT, and it was found that two-correlated factors represent a plausible alternative to the bifactor model, as items fit and multidimensional parameters of both multi-GR and bifac-GR were compared. We tested the possibility that a change in dimensionality of the STICSA model resulted in a different parameterization and meaning of the scale. The results of this second step showed that item #14 ("My arms and legs feel stiff") of the trait scale displayed a poor fit for both compared models. At the parameter level, no significant differences were observed between state and trait items in both models. Compared to the cognitive ones, the somatic items were found to be less discriminative both in trait and state scales. Opposing views exist when explaining the role of either somatic or cognitive patterns of anxiety. For instance, some physiological symptom items in anxiety measures (i.e., in pain scale) have been criticized due to possible overlap with depression symptomatology and have also raised doubts about the uniformity of the anxiety construct [85]. Likewise, somatic and cognitive items were equally endorsed by psychiatric inpatients, and psychological distress was limited to somatic signs of anxiety [86]. As highlighted by different authors, the somatic-cognitive pattern of anxiety gained resonance

among researchers working to ascertain the experiential components of anxiety [74,87]. Concerning the item multidimensional thresholds, in the present study, all the STICSA items encompassed a discrete range of the latent trait continuum, and also showed a clear ordering of the items based on each item's quality and estimated level of multidimensional difficulty.

5. Conclusions

Summing up, our findings supported the equivalence of the bifactor and two-factor correlated models in terms of somatic-cognitive item multidimensional parameters. When only the fit indices of the model were taken into account, the classical bifactor model outperformed our data. Despite this, the bifactor global domain (state or trait) seems to be not as accurate and precise as expected. To overcome this and other methodological issues (e.g., the LD issue within the IRT), other computations might be needed, as in the present investigation (i.e., the marginal trace line transformation). To date, single global STICSA state or trait scores should be used with caution.

Our results showed the viability of measuring trait and state anxiety by the STICSA measure in a community sample with a wide age range, thus demonstrating its versatility. While different confirmatory factor models have been found in literature [33,34], in this study, a two-factor model for both the trait and state scales was confirmed and preferred over the inflated classical bifactor model. Specifically, IRT yielded factors that were consistent with the original domains and showed good fit on a separate set of trait and state datasets [33,34]. The moderately high intercorrelations among somatic and cognitive factors provided support for the compensatory nature of the latent traits [88]. In this way, somatic and cognitive symptoms as measured by STICSA items concurred to improve the measurement precision of anxiety.

The results of the mIRT analysis of the STICSA trait and state scales suggested a more accurate method for evaluating the measure and, ultimately, participants' anxiety. When the discrimination parameter was taken into consideration, items on both the trait and state scales exhibited high multidimensional discrimination values, in line with past studies' results found in the literature on existing clinical instruments [89–91].

Further research is needed to verify this hypothesis in clinical samples. The results of this investigation suggest that somatic and cognitive symptoms are equivalent to pinpointing a stable tendency to respond to threats. The greatest advantage of applying multidimensional item response models to clinical assessment is that the symptoms may be treated as ordered indicators of risk, as well as that specific symptoms and specific factors could be scaled onto a common trait. In clinical assessment, unobservable psychopathological constructs are multidimensional in nature (e.g., mental health or well-being). This is also confirmed in the recent success of transdiagnostic treatments for anxiety and mood disorders with high comorbidity [92].

Other limitations of this study include sample size and characteristics. The sample involved is quite large and heterogeneous, and no data were collected regarding diagnosis. As a consequence, the nature of the response process, as well as differences in the trait manifestation, could differ among patients with anxiety as a primary diagnosis from depressed patients with anxiety in comorbidity. Future research should address the feasibility of the mIRT models in clinical samples. Likewise, from a methodological point of view, there is no consensus among researchers that the accuracy of parameter estimates in mIRT may be biased or inflated by the sample size [93,94].

Author Contributions: All authors contributed to conceptualization; methodology, L.C. and M.B.; validation, L.C.; formal analysis, L.C.; investigation, M.B., M.R. and M.I.; resources, G.D.; data curation, L.C.; writing—original draft preparation, L.C. and M.B.; writing—review and editing, L.C., M.B. and M.R.; visualisation, M.R.; supervision, M.I. All authors have read and agreed to the published version of the manuscript.

Appendix A

Figure A1. LD Indices for the Uni-, Multi-, Bifac−GR Models of the STICSA State and Trait Scales.

Appendix B

Table A1. Unidimensional and Bifactor Graded Response Model Item Parameter Estimates for the 21-Item STICSA State.

Item	Uni-GR				Bifac-GR Conditional						Bifac-GR Marginal					
	c1	c2	c3	a	c1	c2	c3	a_g	a_{S1}	a_{S2}	c1	c2	c3	a^*_g	a^*_{S1}	a^*_{S2}
STIC_S_1	−0.339	−1.839	−3.855	1.196	−0.369	−1.971	−4.072	1.017	0.979		−0.363	−1.938	−4.004	0.881	0.840	
STIC_S_2	−0.314	−2.337	−4.020	1.528	−0.365	−2.695	−4.543	1.365	1.415		−0.267	−1.974	−3.328	1.049	1.103	
STIC_S_6	−1.517	−3.181	−4.715	1.649	−1.774	−3.704	−5.407	1.442	1.574		−1.230	−2.569	−3.750	1.058	1.200	
STIC_S_7	−0.924	−2.791	−4.698	1.799	−1.084	−3.274	−5.445	1.659	1.636		−0.653	−1.973	−3.282	1.195	1.171	
STIC_S_8	−2.497	−4.307	−6.115	2.516	−2.910	−5.060	−7.136	2.389	1.993		−1.218	−2.118	−2.987	1.550	1.156	
STIC_S_12	−0.942	−2.550	−4.291	1.339	−1.013	−2.735	−4.556	1.143	1.063		−0.886	−2.393	−3.986	0.969	0.882	
STIC_S_14	−1.442	−3.405	−5.131	2.012	−1.773	−4.170	−6.191	1.910	1.928		−0.928	−2.183	−3.241	1.263	1.282	
STIC_S_15	−0.824	−2.566	−4.022	1.280	−0.881	−2.732	−4.254	1.074	1.018		−0.820	−2.544	−3.961	0.921	0.861	
STIC_S_18	−1.530	−3.229	−5.039	1.817	−1.612	−3.409	−5.289	1.597	1.177		−1.009	−2.135	−3.312	1.313	0.858	
STIC_S_20	−1.168	−2.239	−3.398	1.055	−1.181	−2.265	−3.432	0.914	0.591		−1.292	−2.478	−3.755	0.863	0.521	
STIC_S_21	−0.656	−1.985	−3.180	0.920	−0.691	−2.075	−3.299	0.736	0.773		−0.939	−2.819	−4.482	0.670	0.709	
STIC_S_3	0.477	−1.958	−3.549	1.980	0.512	−2.091	−3.799	2.207		0.148	0.232	−0.947	−1.721	2.207		0.090
STIC_S_4	−0.815	−2.660	−4.337	1.715	−0.842	−2.741	−4.457	1.816		−0.146	−0.464	−1.509	−2.454	1.816		−0.100
STIC_S_5	−0.093	−2.153	−3.653	1.551	−0.088	−2.217	−3.775	1.646		0.314	−0.053	−1.347	−2.293	1.646		0.228
STIC_S_9	−0.824	−2.660	−4.519	1.897	−1.162	−3.689	−6.136	2.882		−1.023	−0.403	−1.280	−2.129	2.882		−0.546
STIC_S_10	0.701	−1.102	−2.407	1.534	0.806	−1.278	−2.789	1.933		0.601	0.417	−0.661	−1.443	1.933		0.408
STIC_S_11	−0.054	−1.999	−3.664	1.298	−0.045	−1.933	−3.566	1.191		0.143	−0.038	−1.623	−2.994	1.191		0.117
STIC_S_13	−1.492	−3.428	−5.117	2.256	−2.414	−5.521	−8.163	3.932		−1.549	−0.614	−1.404	−2.076	3.932		−0.659
STIC_S_16	0.052	−1.733	−3.247	1.643	0.055	−1.907	−3.546	1.883		0.555	0.029	−1.013	−1.883	1.883		0.381
STIC_S_17	−0.051	−2.320	−4.177	2.063	−0.085	−3.067	−5.507	2.909		1.146	−0.029	−1.054	−1.893	2.909		0.615
STIC_S_19	−0.275	−2.475	−4.163	2.108	−0.337	−2.912	−4.893	2.597		0.820	−0.130	−1.121	−1.884	2.597		0.466

Note. a = uni−GR slope; ag = conditional slope for STICSA State general trait; aS1–aS2 = conditional slopes for specific traits Somatic/Cognitive; a*g = marginal slope for STICSA State general trait; a*S1–a*−S = marginal slopes for specific traits Somatic/Cognitive; c1–c4 = intercepts.

Table A2. Unidimensional and Bifactor Graded Response Model Item Parameter Estimates for the 21-Item STICSA Trait.

Item	Uni-GR				Bifac-GR Conditional						Bifac-GR Marginal					
	c1	c2	c3	a	c1	c2	c3	a_g	a_{S1}	a_{S2}	c1	c2	c3	a^*_g	a^*_{S1}	a^*_{S2}
STIC_T_1	1.377	−0.721	−3.059	1.020	1.483	−0.787	−3.264	1.205	−0.392		1.231	−0.653	−2.709	1.174	−0.320	
STIC_T_2	1.230	−1.044	−3.202	1.264	1.288	−1.096	−3.324	1.402	−0.078		0.919	−0.782	−2.371	1.401	−0.060	
STIC_T_6	−0.267	−2.154	−3.997	1.485	−0.314	−2.483	−4.564	1.859	0.608		−0.169	−1.336	−2.455	1.750	0.410	
STIC_T_7	0.185	−2.107	−4.112	1.577	0.266	−3.127	−5.972	2.543	1.501		0.105	−1.230	−2.348	1.906	0.834	
STIC_T_8	−1.162	−3.211	−4.927	1.991	−1.409	−3.874	−5.923	2.565	0.746		−0.549	−1.510	−2.309	2.349	0.412	
STIC_T_12	−0.281	−2.088	−3.909	1.285	−0.310	−2.241	−4.149	1.514	−0.100		−0.205	−1.480	−2.740	1.511	−0.075	
STIC_T_14	−0.961	−2.969	−4.892	1.801	−1.064	−3.264	−5.338	2.118	0.154		−0.502	−1.541	−2.520	2.109	0.096	
STIC_T_15	−0.399	−2.496	−4.260	1.406	−0.440	−2.699	−4.574	1.659	−0.162		−0.265	−1.627	−2.757	1.652	−0.116	
STIC_T_18	−0.694	−2.780	−4.702	1.723	−0.785	−3.091	−5.169	2.025	−0.390		−0.388	−1.526	−2.553	1.974	−0.251	
STIC_T_20	−0.147	−1.589	−3.112	0.897	−0.165	−1.745	−3.365	1.056	−0.571		−0.156	−1.652	−3.187	1.001	−0.485	
STIC_T_21	−0.367	−1.924	−3.278	0.922	−0.413	−2.116	−3.553	1.131	−0.528		−0.365	−1.871	−3.141	1.080	−0.440	
STIC_T_3	1.553	−1.003	−3.123	1.691	1.685	−1.092	−3.387	1.443		1.332	1.168	−0.757	−2.347	1.443		1.266
STIC_T_4	0.004	−2.006	−3.928	1.409	0.006	−2.097	−4.081	1.173		1.026	0.005	−1.788	−3.479	1.173		0.973
STIC_T_5	0.447	−1.503	−3.320	1.307	0.476	−1.577	−3.478	1.066		1.023	0.447	−1.479	−3.263	1.066		1.007
STIC_T_9	0.094	−1.883	−3.802	1.669	0.097	−2.019	−4.066	1.431		1.253	0.068	−1.411	−2.841	1.431		1.161
STIC_T_10	1.659	−0.365	−2.105	1.426	1.811	−0.403	−2.305	1.182		1.242	1.532	−0.341	−1.950	1.182		1.274
STIC_T_11	0.492	−1.463	−3.082	1.030	0.493	−1.459	−3.076	0.965		0.335	0.511	−1.512	−3.188	0.965		0.296
STIC_T_13	−0.630	−2.761	−4.466	1.919	−0.678	−2.953	−4.764	1.671		1.360	−0.406	−1.767	−2.851	1.671		1.181
STIC_T_16	0.740	−1.064	−2.849	1.352	0.763	−1.103	−2.935	1.127		0.914	0.677	−0.979	−2.604	1.127		0.852
STIC_T_17	0.705	−1.650	−3.856	1.900	0.762	−1.785	−4.157	1.651		1.411	0.462	−1.081	−2.518	1.651		1.260
STIC_T_19	0.314	−1.886	−3.806	1.964	0.341	−2.022	−4.074	1.722		1.387	0.198	−1.174	−2.366	1.722		1.189

Note. a = uni-GR slope; a_g = conditional slope for STICSA Trait general trait; a_{S1}–a_{S2} = conditional slopes for specific traits Somatic/Cognitive; a^*_g = marginal slope for STICSA Trait general trait; a^*_{S1}–a^*_{S} = marginal slopes for specific traits Somatic/Cognitive; c1–c4 = intercepts.

References

1. Shear, M.K.; Bjelland, I.; Beesdo, K.; Gloster, A.T.; Wittchen, H.U. Supplementary dimensional assessment in anxiety disorders. *Int. J. Methods Psychiatr. Res.* **2007**, *16*, S52–S64. [CrossRef]
2. Lebeau, R.T.; Glenn, D.E.; Hanover, L.N.; Beesdo-Baum, K.; Wittchen, H.U.; Craske, M.G. A dimensional approach to measuring anxiety for DSM-5. *Int. J. Methods Psychiatr. Res.* **2012**, *21*, 258–272. [CrossRef]
3. Martens, R.; Vealey, R.S.; Burton, D. *Competitive Anxiety in Sport*; Human Kinetics: Champaign, IL, USA, 1990.
4. Stein, D.J.; Craske, M.G.; Rothbaum, B.O.; Chamberlain, S.R.; Fineberg, N.A.; Choi, K.W.; Jonge, P.; Baldwin, D.S.; Maj, M. The clinical characterization of the adult patient with an anxiety or related disorder aimed at personalization of management. *World Psychiatry* **2021**, *20*, 336–356. [CrossRef]
5. Pacheco-Unguetti, A.P.; Acosta, A.; Callejas, A.; Lupiáñez, J. Attention and Anxiety. *Psychol. Sci.* **2010**, *21*, 298–304. [CrossRef]
6. Spielberg, C.D. Assessment of State and Trait Anxiety: Conceptual and Methodological Issues. *South. Psychol.* **1985**, *2*, 6–16.
7. Heeren, A.; Jones, P.J.; McNally, R.J. Mapping network connectivity among symptoms of social anxiety and comorbid depression in people with social anxiety disorder. *J. Affect. Disord.* **2018**, *228*, 75–82. [CrossRef] [PubMed]
8. Tellegen, A. Structures of Mood and Personality and Their Relevance to Assessing Anxiety, with an Emphasis on Self-Report. In *Anxiety and the Anxiety Disorders*; Routledge: London, UK, 1985; pp. 681–706.
9. Ashton, M.C.; Lee, K.; Goldberg, L.R. A Hierarchical Analysis of 1,710 English Personality-Descriptive Adjectives. *J. Personal. Soc. Psychol.* **2004**, *87*, 707–721. [CrossRef]
10. Goldberg, L.R. The development of markers for the Big-Five factor structure. *Psychol. Assess.* **1992**, *4*, 26–42. [CrossRef]
11. Weems, C.F.; Costa, N.M.; Watts, S.E.; Taylor, L.K.; Cannon, M.F. Cognitive Errors, Anxiety Sensitivity, and Anxiety Control Beliefs. *Behav. Modif.* **2007**, *31*, 174–201. [CrossRef] [PubMed]
12. Lilienfeld, S.O.; Turner, S.M.; Jacob, R.G. Anxiety sensitivity: An examination of theoretical and methodological issues. *Adv. Behav. Res. Ther.* **1993**, *15*, 147–183. [CrossRef]
13. Cox, R. *Sports Psychology: Concepts and Applications*, 4th ed.; McGraw-Hill: New York, NY, USA, 2007; Volume 45, pp. 124–154.
14. Allen, B.P.; Potkay, C.R. On the arbitrary distinction between states and traits. *J. Personal. Soc. Psychol.* **1981**, *41*, 916–928. [CrossRef]
15. Luthans, F.; Avolio, B.J.; Avey, J.B.; Norman, S.M. Positive Psychological Capital: Measurement and Relationship with Performance and Satisfaction. *Pers. Psychol.* **2007**, *60*, 541–572. [CrossRef]
16. Steptoe, A.; Kearsley, N. Cognitive and somatic anxiety. *Behav. Res. Ther.* **1990**, *28*, 75–81. [CrossRef] [PubMed]
17. Ree, M.J.; French, D.; MacLeod, C.; Locke, V. Distinguishing Cognitive and Somatic Dimensions of State and Trait Anxiety: Development and Validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behav. Cogn. Psychother.* **2008**, *36*, 313–332. [CrossRef]
18. Bryant, C. Anxiety and depression in old age: Challenges in recognition and diagnosis. *Int. Psychogeriatr.* **2010**, *22*, 511–513. [CrossRef]
19. Flint, A.J. Generalised Anxiety Disorder in Elderly Patients. *Drugs Aging* **2005**, *22*, 101–114. [CrossRef]
20. Wetherell, J.L.; Gatz, M.; Pedersen, N.L. A longitudinal analysis of anxiety and depressive symptoms. *Psychol. Aging* **2001**, *16*, 187–195. [CrossRef]

21. Clark, L.A.; Watson, D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* **1991**, *100*, 316–336. [CrossRef]

22. Watson, D.; Clark, L.A.; Weber, K.; Assenheimer, J.S.; Strauss, M.E.; McCormick, R.A. Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *J. Abnorm. Psychol.* **1995**, *104*, 15–25. [CrossRef]

23. Therrien, Z.; Hunsley, J. Assessment of anxiety in older adults: A systematic review of commonly used measures. *Aging Ment. Health* **2012**, *16*, 1–16. [CrossRef] [PubMed]

24. Balsamo, M.; Romanelli, R.; Innamorati, M.; Ciccarese, G.; Carlucci, L.; Saggino, A. The State-Trait Anxiety Inventory: Shadows and Lights on its Construct Validity. *J. Psychopathol. Behav. Assess.* **2013**, *35*, 475–486. [CrossRef]

25. Grös, D.F.; Antony, M.M.; Simms, L.J.; McCabe, R.E. Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychol. Assess.* **2007**, *19*, 369–381. [CrossRef]

26. Roberts, K.E.; Hart, T.A.; Eastwood, J.D. Factor structure and validity of the State-Trait Inventory for Cognitive and Somatic Anxiety. *Psychol. Assess.* **2016**, *28*, 134–146. [CrossRef] [PubMed]

27. Carlucci, L.; Watkins, M.W.; Sergi, M.R.; Cataldi, F.; Saggino, A.; Balsamo, M. Dimensions of Anxiety, Age, and Gender: Assessing Dimensionality and Measurement Invariance of the State-Trait for Cognitive and Somatic Anxiety (STICSA) in an Italian Sample. *Front. Psychol.* **2018**, *9*, 2345. [CrossRef] [PubMed]

28. Saggino, A.; Balsamo, M.; Carlucci, L.; Cavalletti, V.; Sergi, M.R.; da Fermo, G.; Dèttore, D.; Marsigli, N.; Petruccelli, I.; Pizzo, S.; et al. Psychometric Properties of the Italian Version of the Young Schema Questionnaire L-3: Preliminary Results. *Front. Psychol.* **2018**, *9*, 312. [CrossRef] [PubMed]

29. Saggino, A.; Carlucci, L.; Sergi, M.R.; D'Ambrosio, I.; Fairfield, B.; Cera, N.; Balsamo, M. A Validation Study of the Psychometric Properties of the Other As Shamer Scale–2. *SAGE Open* **2017**, *7*, 215824401770424. [CrossRef]

30. Carlucci, L.; D'Ambrosio, I.; Innamorati, M.; Saggino, A.; Balsamo, M. Co-rumination, anxiety, and maladaptive cognitive schemas: When friendship can hurt. *Psychol. Res. Behav. Manag.* **2018**, *11*, 133–144. [CrossRef]

31. Balsamo, M.; Carlucci, L.; Sergi, M.R.; Romanelli, R.; D'Ambrosio, I.; Fairfield, B.; Ree, M.J.; Saggino, A. A new measure for trait and state anxiety: The State Trait Inventory of Cognitive and Somatic Anxiety (STICSA). Standardization in an Italian population. *Psicoter. Cogn. Comport.* **2016**, *22*, 229–232.

32. Balsamo, M.; Innamorati, M.; Van Dam, N.T.; Carlucci, L.; Saggino, A. Measuring anxiety in the elderly: Psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* **2015**, *27*, 999–1008. [CrossRef]

33. Styck, K.M.; Rodriguez, M.C.; Yi, E.H. Dimensionality of the State–Trait Inventory of Cognitive and Somatic Anxiety. *Assessment* **2022**, *29*, 103–127. [CrossRef]

34. Barros, F.; Figueiredo, C.; Brás, S.; Carvalho, J.M.; Soares, S.C. Multidimensional assessment of anxiety through the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): From dimensionality to response prediction across emotional contexts. *PLoS ONE* **2022**, *17*, e0262960. [CrossRef]

35. Kline, R.B. *Principles and Practice of Structural Equation Modeling*; Guilford Publications: New York, NY, USA, 2023.

36. Hinkin, T.R. A brief tutorial on the development of measures for use in survey questionnaires. *Organ. Res. Methods* **1998**, *1*, 104–121. [CrossRef]

37. Bentler, P.M.; Bonett, D.G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* **1980**, *88*, 588–606. [CrossRef]

38. Bentler, P.M.; Satorra, A. Testing model nesting and equivalence. *Psychol. Methods* **2010**, *15*, 111–123. [CrossRef]

39. Preacher, K.J. Quantifying Parsimony in Structural Equation Modeling. *Multivar. Behav. Res.* **2006**, *41*, 227–259. [CrossRef] [PubMed]

40. Howard, M.C. A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *Int. J. Hum.-Comput. Interact.* **2016**, *32*, 51–62. [CrossRef]

41. Schmitt, T.A. Current methodological considerations in exploratory and confirmatory factor analysis. *J. Psychoeduc. Assess.* **2011**, *29*, 304–321. [CrossRef]

42. Tindall, I.K.; Curtis, G.J.; Locke, V. Dimensionality and Measurement Invariance of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) and Validity Comparison with Measures of Negative Emotionality. *Front. Psychol.* **2021**, *12*, 644889. [CrossRef]

43. Jabrayilov, R.; Emons, W.H.; Sijtsma, K. Comparison of classical test theory and item response theory in individual change assessment. *Appl. Psychol. Meas.* **2016**, *40*, 559–572. [CrossRef]

44. Horton, M.C.; Marais, I.; Christensen, K.B. Dimensionality. In *Rasch Models in Health*, 1st ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2013; pp. 137–162.

45. Thomas, M.L. The Value of Item Response Theory in Clinical Assessment: A Review. *Assessment* **2011**, *18*, 291–307. [CrossRef]

46. van der Linden, W.J.; Hambleton, R.K. Item Response Theory: Brief History, Common Models, and Extensions. In *Handbook of Modern Item Response Theory*; Springer: New York, NY, USA, 1997; pp. 1–28. [CrossRef]

47. Wicherts, J.M. Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence* **2017**, *60*, 26–38. [CrossRef]

48. Reckase, M.D. Multidimensional Item Response Theory Models. In *Multidimensional Item Response Theory*; Springer: New York, NY, USA, 2009; pp. 79–112. [CrossRef]

49. Baker, F.B. *The Basics of Item Response Theory*, 2nd ed.; Office of Educational Research and Improvement (ED): Washington, DC, USA, 2001.

50. Stucky, B.D.; Edelen, M.O. Using hierarchical IRT models to create unidimensional measures from multidimensional data. In *Handbook of Item Response Theory Modeling*; Routledge: London, UK, 2014; pp. 201–224.

51. Balsamo, M.; Carlucci, L.; Innamorati, M.; Lester, D.; Pompili, M. Further insights into the beck hopelessness scale (BHS): Unidimensionality among psychiatric inpatients. *Front. Psychiatry* **2020**, *11*, 727. [CrossRef] [PubMed]

52. Balsamo, M.; Saggino, A.; Carlucci, L. Tailored Screening for Late-Life Depression: A Short Version of the Teate Depression Inventory (TDI-E). *Front. Psychol.* **2019**, *10*, 2693. [CrossRef] [PubMed]

53. Chalmers, R.P. mirt: A Multidimensional Item Response Theory Package for the R Environment. *J. Stat. Softw.* **2012**, *48*, 1–29. [CrossRef]

54. Mokken, R.J. *A Theory and Procedure of Scale Analysis*; De Gruyter: Berlin, Germany, 1971.

55. Sijtsma, K.; Molenaar, I.W. *Introduction to Noparametric Item Response Theory*; Sage: Newcastle upon Tyne, UK, 2002; Volume 5.

56. Chen, W.-H.; Thissen, D. Local Dependence Indexes for Item Pairs Using Item Response Theory. *J. Educ. Behav. Stat.* **1997**, *22*, 265–289. [CrossRef]

57. Muraki, E.; Carlson, J.E. Full-Information Factor Analysis for Polytomous Item Responses. *Appl. Psychol. Meas.* **1995**, *19*, 73–90. [CrossRef]

58. Cai, L.; Monroe, S. *A New Statistic for Evaluating Item Response Theory Models for Ordinal Data*; CRESST Report 839; University of California: Los Angeles, CA, USA, 2014.

59. Schermelleh-Engel, K.; Moosbrugger, H.; Müller, H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* **2003**, *8*, 23–74.

60. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]

61. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]

62. Orlando, M.; Thissen, D. Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Appl. Psychol. Meas.* **2000**, *24*, 50–64. [CrossRef]

63. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]

64. Toland, M.D. Practical Guide to Conducting an Item Response Theory Analysis. *J. Early Adolesc.* **2014**, *34*, 120–151. [CrossRef]

65. Hasmy, A. Compare Unidimensional & Multidimensional Rasch Model for Test with Multidimensional Construct and Items Local Dependence. *J. Educ. Learn.* **2014**, *8*, 187–194. [CrossRef]

66. Gros, D.F.; Simms, L.J.; Antony, M.M. Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads. *Behav. Ther.* **2010**, *41*, 277–284. [CrossRef] [PubMed]

67. Reise, S.P.; Morizot, J.; Hays, R.D. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual. Life Res.* **2007**, *16*, 19–31. [CrossRef]

68. Sireci, S.G.; Thissen, D.; Wainer, H. On the Reliability of Testlet-Based Tests. *J. Educ. Meas.* **1991**, *28*, 237–247. [CrossRef]

69. Stucky, B.D.; Thissen, D.; Orlando Edelen, M. Using Logistic Approximations of Marginal Trace Lines to Develop Short Assessments. *Appl. Psychol. Meas.* **2013**, *37*, 41–57. [CrossRef]

70. Woods, C.M.; Thissen, D. Item Response Theory with Estimation of the Latent Population Distribution Using Spline-Based Densities. *Psychometrika* **2006**, *71*, 281–301. [CrossRef]

71. Tuerlinckx, F.; De Boeck, P. Modeling Local Item Dependencies in Item Response Theory. *Psychol. Belg.* **1998**, *38*, 61. [CrossRef]

72. Anderson, D.R. *Model Based Inference in the Life Sciences: A Primer on Evidence*; Springer: New York, NY, USA, 2008; Volume 31.

73. Calvo, M.G.; Avero, P.; Castillo, M.D.; Miguel-Tobal, J.J. Multidimensional Anxiety and Content-specificity Effects in Preferential Processing of Threat. *Eur. Psychol.* **2003**, *8*, 252–265. [CrossRef]

74. Endler, N.S.; Kocovski, N.L. State and trait anxiety revisited. *J. Anxiety Disord.* **2001**, *15*, 231–245. [CrossRef]

75. Allen, L.; Choate, M. Toward a unified treatment for emotional disorder. *Behav. Ther.* **2004**, *35*, 205–230.

76. Carlucci, L.; Saggino, A.; Balsamo, M. On the efficacy of the unified protocol for transdiagnostic treatment of emotional disorders: A systematic review and meta-analysis. *Clin. Psychol. Rev.* **2021**, *87*, 101999. [CrossRef]

77. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [CrossRef]

78. Chernyshenko, O.S.; Stark, S.; Chan, K.-Y.; Drasgow, F.; Williams, B. Fitting item response theory models to two personality inventories: Issues and insights. *Multivar. Behav. Res.* **2001**, *36*, 523–562. [CrossRef]

79. Krueger, R.F.; Nichol, P.E.; Hicks, B.M.; Markon, K.E.; Patrick, C.J.; Iacono, W.G.; McGue, M. Using Latent Trait Modeling to Conceptualize an Alcohol Problems Continuum. *Psychol. Assess.* **2004**, *16*, 107–119. [CrossRef]

80. Pitt, M.A.; Myung, I.J.; Zhang, S. Toward a method of selecting among computational models of cognition. *Psychol. Rev.* **2002**, *109*, 472–491. [CrossRef] [PubMed]

81. Bornovalova, M.A.; Choate, A.M.; Fatimah, H.; Petersen, K.J.; Wiernik, B.M. Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. *Biol. Psychiatry* **2020**, *88*, 18–27. [CrossRef] [PubMed]

82. Balsamo, M.; Imperatori, C.; Sergi, M.R.; Belvederi Murri, M.; Continisio, M.; Tamburello, A.; Innamorati, M.; Saggino, A. Cognitive Vulnerabilities and Depression in Young Adults: An ROC Curves Analysis. *Depress. Res. Treat.* **2013**, *2013*, 407602. [CrossRef] [PubMed]

83. Innamorati, M.; Tamburello, S.; Contardi, A.; Imperatori, C.; Tamburello, A.; Saggino, A.; Balsamo, M. Psychometric Properties of the Attitudes toward Self-Revised in Italian Young Adults. *Depress. Res. Treat.* **2013**, *2013*, 209216. [CrossRef] [PubMed]

84. Wang, W.-C.; Chen, P.-H.; Cheng, Y.-Y. Improving Measurement Precision of Test Batteries Using Multidimensional Item Response Models. *Psychol. Methods* **2004**, *9*, 116–136. [CrossRef] [PubMed]

85. McCracken, L.M.; Gross, R.T.; Aikens, J.; Carnrike, C.L.M. The assessment of anxiety and fear in persons with chronic pain: A comparison of instruments. *Behav. Res. Ther.* **1996**, *34*, 927–933. [CrossRef] [PubMed]

86. Buckelew, S.P.; Hannay, H.J. Relationships among Anxiety, Defensiveness, Sex, Task Difficulty, and Performance on Various Neuropsychological Tasks. *Percept. Mot. Ski.* **1986**, *63*, 711–718. [CrossRef]

87. Koksal, F.; Power, K.G. Four Systems Anxiety Questionnaire (FSAQ): A Self-Report Measure of Somatic, Cognitive, Behavioral, and Feeling Components. *J. Personal. Assess.* **1990**, *54*, 534–545. [CrossRef] [PubMed]

88. Sijtsma, K.; Junker, B.W. Item Response Theory: Past Performance, Present Developments, and Future Expectations. *Behaviormetrika* **2006**, *33*, 75–102. [CrossRef]

89. Reise, S.P.; Waller, N.G. Item Response Theory and Clinical Measurement. *Annu. Rev. Clin. Psychol.* **2009**, *5*, 27–48. [CrossRef]

90. Balsamo, M. Personality and depression: Evidence of a possible mediating role for anger trait in the relationship between cooperativeness and depression. *Compr. Psychiatry* **2013**, *54*, 46–52. [CrossRef]

91. Balsamo, M. Anger and depression: Evidence of a possible mediating role for rumination. *Psychol. Rep.* **2010**, *106*, 3–12. [CrossRef]

92. Barlow, D.H.; Farchione, T.J.; Bullis, J.R.; Gallagher, M.W.; Murray-Latin, H.; Sauer-Zavala, S.; Bentley, K.H.; Thompson-Hollands, J.; Conklin, L.R.; Boswell, J.F.; et al. The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Compared With Diagnosis-Specific Protocols for Anxiety Disorders. *JAMA Psychiatry* **2017**, *74*, 875. [CrossRef]

93. Colledani, D.; Anselmi, P.; Robusto, E. Using multidimensional item response theory to develop an abbreviated form of the Italian version of Eysenck's IVE questionnaire. *Personal. Individ. Differ.* **2019**, *142*, 45–52. [CrossRef]

94. Finch, H. Item Parameter Estimation for the MIRT Model. *Appl. Psychol. Meas.* **2010**, *34*, 10–26. [CrossRef]