Baroni, D., Caccico, L., Ciandri, S., [...] & Chiorri, C. (2022). Measurement invariance of the Liebowitz Social Anxiety Scale-Self-Report. *Journal of Clinical Psychology*, Advance Online Publication. https://doi.org/10.1002/jclp.23413

RUNNING HEAD: LSAS-SR Measurement invariance

Measurement Invariance of the Liebowitz Social Anxiety Scale-Self-Report

Duccio Baroni[a-c], Laura Caccico[c], Serena Ciandri[c], Cristian Di Gesto[b-c], Laura Di Leonardo[c], Alice Fiesoli[c], Elena Grassi[c], Francesco Lauretta[c], Antonella Lebruto[c], Nicola Marsigli[c], Giulia Rosa Policardo[b-c], Martina Rosadoni[c], Carlo Chiorri[d]*

[a] University of Florence – Department of Education, Languages, Intercultures, Literatures and Psychology, Florence, Italy

[b] University of Florence – Department of Health Sciences, Florence, Italy

[c] Institute of Behavioral and Cognitive Psychology and Psychotherapy, Florence, Italy

[d] Department of Educational Sciences, University of Genova, Italy

*** Corresponding Author**

Carlo Chiorri, Orcid 0000-0002-1640-3897

Department of Educational Sciences

University of Genoa

Corso A. Podestà, 2, 16128

Genova (Italy)

e-mail: carlo.chiorri@unige.it

tel: +39 010 20953709

fax: +39 010 20953728

Measurement Invariance of the Liebowitz Social Anxiety Scale-Self-Report

Abstract

Objective

The Liebowitz Social Anxiety Scale-Self Report (LSAS-SR) is a self-report measure of social anxiety (SA), which has shown adequate psychometric properties across cultures. However, no study has systematically evaluated its measurement invariance between (a) individuals with and without a diagnosis of social anxiety disorder (SAD) and (b) males and females. The current study addresses this issue.

Methods

We collected data on 257 (158 females) Italian individuals diagnosed with SAD and 356 (232 females) community-dwelling adults.

Results

We initially found support for the unidimensionality of the Italian LSAS-SR measurement model in all samples. Using the Graded Response Model, we obtained evidence of partial measurement invariance and differential item functioning bewteen community-dwelling and SAD-diagnosed individuals and evidence of strong measurement invariance between male and female participants.

Conclusions

The results of this study suggest that the Italian LSAS-SR measures the same trait in the same way across the symptom continuum and sexes, making it a psychometrically sound tool for assessment, screening, and research purposes.

Keywords: Liebowitz Social Anxiety Scale; Social Anxiety Disorder; Graded Response Model; Measurement Invariance

Introduction

The main aim of this study was to test the measurement invariance (MI) of the Liebowitz Social Anxiety Scale-Self Report (LSAS-SR, Liebowitz, 1987) between individuals with and without a diagnosis of social anxiety, on the one hand, and males and females, on the other, using the Graded Response Model (GRM, Samejima, 1968). This test is crucial to determine whether observed differences in LSAS-SR scores represent true differences between groups or reflect a non-equivalence in the measurement process. In the following, we initially present the clinical and epidemiological characteristics of social anxiety (SA); then, we report on studies that investigated the basic psychometric properties of the LSAS-SR; finally, we introduce the issue of measurement of invariance and review the results of previous studies that investigated the dimensionality of the LSAS-SR item pool, since the GRM requires that a scale item pool be unidimensional.

Social anxiety

Social Anxiety Disorder (SAD) is a psychological disorder characterized by high discomfort in social situations (American Psychiatric Association [APA], 2013). The core fear of SAD is to be negatively judged by others. People with SAD also worry that actions or behaviors associated with their anxiety will be negatively evaluated by others, leading them to avoid social situations (Skocic et al., 2015). High levels of perceived stress in social interactions and their subsequent avoidance have a severe impact on the individual's well-being, affecting the normal functioning in everyday life across multiple domains (Aderka et al., 2012). Furthermore, SAD is commonly associated with high self-criticism (Cox et al., 2004), other anxiety disorders (Coehlo et al., 2007), obsessive-compulsive disorder (Assunção et al., 2012), depression (Kashdan & Roberts, 2011), eating disorders (Levinson & Rodebaugh, 2012), and alcohol-related disorders (Schneier et al., 2010). Regarding the prevalence of SAD, a cross-national epidemiology survey has shown that in high-income countries (such as the United States [US], Germany, or Italy), the estimated prevalence of SAD was 5.5% for lifetime, 3.1% for 12 months, and 1.7% for 30 days (Stein et al., 2017), and it has been reported that 12.1% of US adults experience social anxiety disorder at some time in their lives (Harvard Medical School, 2007). In summary, SAD is a prominent public health issue that results in substantial impairment in multiple domains. Concerning the aforementioned cross-national survey (Stein et al., 2017), among respondents with a 12-month diagnosis of SAD,

the impairment was more severe in the relationship and social domains than in the home and work domains.

Given the widespread nature of the disorder and the negative impacts on individual functioning, SAD has been the focus of attention for clinicians and researchers (Deller et al., 2020). Several instruments have been developed to assess both symptoms that individuals experience and social situations that could exacerbate them, also considering the number of changes in the definition of SAD (Bögels et al., 2010) made in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013).

The Liebowitz Social Anxiety Scale

Despite the many available methods for assessing social anxiety, self-report instruments are among the most frequently used (Bunnell et al., 2013). Some of the most popular (e.g., *Fear of Negative Evaluation Scale*, Watson & Friend, 1969; *Interaction Anxiousness Scale*, Leary, 1983; *Fear Questionnaire-Social Phobia Subscale*, Marks & Mathews, 1979; *Social Interaction Anxiety Scale and Social Phobia Scale*, Mattick & Clarke, 1998; *Social Phobia Inventory*, Connor et al., 2000) assess the presence of general social distress. However, they often lack specificity (e.g., they miss the distinction between fear and avoidance) and ability to discriminate across diagnostic groups (e.g., Turner et al., 1987).

A useful measure to identify different levels of social anxiety (SA) is the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987). Although initially developed as a clinician-administered scale, some researchers developed and investigated the psychometric properties of a self-reported version of the LSAS (LSAS-SR; Baker et al., 2002; Cox et al., 1998; Fresco et al., 2001; Oakman et al., 2003). The LSAS-SR is one of the most extensively studied scales in terms of its psychometric properties (see, e.g., Osorio et al., 2009). This scale comprises 24 items representing common and daily scenarios: 12 items represent social interaction situations (e.g., ‘Going to a party’), while the others represent performance situations (e.g., ‘Writing while being observed’). The respondent is required to assess the level of fear elicited by each scenario and the subjective probability of avoiding it. The Fear scale ratings range from 0 (no fear) to 3 (severe fear). The Avoidance scale ratings range from 0 to 3 (0 = never; 1 = occasionally; 2 = often; 3 = usually), referring to how often the respondent avoids a specific situation. The availability of both fear and avoidance ratings for the same situations is a unique feature of the LSAS-SR. Avoidance of social situations is not a strict requirement for the diagnosis of SAD, but it has long been considered as a key issue in models of social anxiety (see, e.g., Wells et al., 1995). Recent studies have shown that the transition from fear

to avoidance of social situations is common and rapid, typically occurring within the first year (Sunderland et al., 2016). Hence, LSAS-SR scores can help in distinguish individuals who have fear concerns but have not already developed avoidance behaviors.

Previous studies provided evidence that the LSAS-SR has good psychometric properties. The results of Baker et al. (2002) supported the test-retest reliability of the LSAS-SR scores. The internal consistency of the LSAS-SR was also adequate across different studies, with Cronbach's alpha coefficients ranging from .61 to .98 (Baker et al., 2002; Fresco et al., 2001). Evidence of convergent validity has also been provided, with previous international studies showing that the LSAS-SR scores were correlated with other measures of SA (Beard et al., 2012; Forni dos Santos et al., 2013), such as the Social Phobia Inventory (SPIN; Connor et al., 2000) and the Brief Scale of Social Phobia (BSPS; Davidson et al., 1997). Moreover, Baker et al. (2002) found moderate to excellent correlations between the LSAS-SR scores and the Social Phobia Scale score (SPS; Mattick & Clarke, 1998). Discriminant validity has been investigated using scales that evaluate depression and general anxiety symptoms: in either case, the correlations were weak or moderate (Baker et al., 2002; Levin et al., 2002).

The issue of measurement invariance

A core psychometric property of a measure of symptomatology is its criterion (or known-group) validity, i.e., the ability of its scores to distinguish among individuals with and without a diagnosis of the disorder. Heimberg and Holaway (2007) showed that LSAS-SR discriminates between patients with Social Anxiety Disorder (SAD) and Generalized Anxiety Disorder (GAD), as well as between patients with SAD and nonanxious controls. Mennin et al. (2002) and Rytwinski et al. (2009) also provided cutoff scores for a diagnosis of SAD. However, all of these studies compared the observed scores on the LSAS-SR among groups based on the untested assumption that this measurement is invariant (or equivalent) between participants with SAD and those with no or other disorders. Good criterion validity is crucial for a measure intended to be routinely used in clinical settings for assessment and screening purposes. From a statistical point of view, the comparison of latent and observed scores on an assessment instrument across different groups is valid as long as the assessment instrument provides the same kind of information for all those groups. When item scores differ between individuals with the same level on a trait and this difference depends on a specific group the individuals belong to (i.e., the so-called differential item functioning), the score on the scale has a different meaning across groups. Consequently, the use of the scale and its cutoff scores

for screening purposes becomes questionable, and inaccurate statistical and practical inferences may result (see, e.g., Olino, 2020; Vandenberg & Lance, 2000).

A detailed description of the theoretical and statistical aspects of testing measurement invariance (MI) is beyond the scope of this paper (although the models used here are extensively described in the Materials and methods section). Briefly, as pointed out by Bunnell et al. (2013), in the case of a measure of symptomatology, MI implies that the items and the numerical ratings hold the same meaning for those with and without a specific disorder. Moreover, the constructs represented by the items must be perceived similarly across groups. To the best of our knowledge, only one study (Kubota et al., 2016) has sought to evaluate the measurement invariance of the LSAS-SR between individuals with and without SAD. This study involved participants with SAD and nonclinical university students from Australia and Japan. Since the authors could not find a replicable factor structure in single-group factor analyses, they concluded that the LSAS-SR is not measurement invariant across known groups and national samples. However, they could not formally test measurement invariance using common multiple-group procedures.

A similar argument can be made for the test of sex differences in the LSAS-SR scores. As reported in recent reviews (Asher & Aderka, 2018; Asher et al., 2017), social anxiety tends to have a higher prevalence in females than in males, although the difference is larger in adolescence and decreases with age. Women also report more severe symptoms and a higher number of social fears, although men with SAD are more likely to seek treatment than their female counterparts. This suggests that men may experience more distress than women due to their social anxiety (Asher et al., 2017). However, the course of SAD, its comorbidity, and the functional impairment it causes do not appear to vary substantially with sex (Asher et al., 2017). Regarding the LSAS-SR, a large-scale study in 18 Spanish-speaking countries showed that women tended to obtain higher scores than men on either scale, although the effect size tended to be larger in the Fear (Cohen's *d* ranging from 0.11 to 0.47, median 0.32) than in the Avoidance subscale (Cohen's *d* ranging from 0.00 to 0.35, median 0.15) and tended to remain stable with age. Again, these comparisons were carried out on the observed scores, assuming that the LSAS-SR item scores reflect the latent level of SA regardless of the sex of the participants (that is., there is no differential item functioning [DIF]). As argued by Bunnell et al. (2013), this assumption cannot be taken for granted, as it has been provided evidence of symptom variation across sexes in terms of feared situations and responses to feared stimuli. For example, women with SAD appear to have more severe fears and avoid particular social situations more frequently (e.g., doing everyday activities while being observed, talking to

authority figures, speaking in public). On the other hand, men tend to report more frequently other SA-related difficulties, such as using public restrooms or dealing with dating situations (see, e.g., Xu et al., 2012). Thus, one of the aims of this study was to investigate the measurement invariance between females and males, that is, to test whether differences in observed scores on the LSAS-SR reflect differences in latent scores and whether some items show DIF, i.e., in which differences between sexes remain after controlling for latent mean differences.

The issue of the factor structure of the Liebowitz Social Anxiety Scale

In order to test MI, a measurement model has to be specified. Despite the consistency of research results on the reliability and validity of LSAS-SR scores, there is much less agreement on its factor structure. In his original study, Liebowitz (1987) proposed a two-factor model: performance and social interaction. Later, many studies suggested different factor models. For example, four-factor models were found by Slavkin et al. (1990; interaction with strangers, formal performance/center of attention, drinking and eating while being observed, and behavior in parties/informal situations) and by Safren et al. (1999; social interaction, public speaking, observation by others, and eating and drinking in public). Although the latter model received some support from the results reported by Oakman et al. (2003), the analyses conducted by Baker et al. (2002) on the Fear scale suggested a five-factor structure: social interaction anxiety, nonverbal performance anxiety, ingestion anxiety, public performance anxiety, and assertiveness anxiety.

The many national adaptations of the LSAS-SR did not help to provide a conclusive answer on its measurement model. The scale has been adapted in Brazilian-Portuguese (Caballo et al., 2019; Caballo et al., 2014; Forni dos Santos et al., 2013; Osorio et al., 2009), French (Yao et al., 1999), Japanese (Sugawara et al., 2012), Hebrew (Levin et al., 2002), Persian (Hasani et al., 2017), Spanish (Bobes et al., 1999; Caballo et al., 2014, 2019), and Turkish (Soykan et al., 2003). All these versions recorded excellent psychometric properties but reported different factor structures. For example, Levin et al. (2002) found three factors (the group performance/interaction subscale, the dyadic interaction subscale, and the public activities subscale). Kubota et al. (2016), Osorio et al. (2009), and Sugawara et al. (2012) reported the same four factors found by Safren et al. (1999). Caballo et al. (2019) and Kubota et al. (2016) found five similar (but not identical) factors, while Forni dos Santos et al. (2013) did not report conclusive results. Finally, considering early onset of SAD, Shachar et al. (2014) developed an Israeli version of the LSAS-SR for children and adolescents, which

showed a two-factor structure. As suggested by Osorio et al. (2009), inconsistencies in the results of published studies on SA potentially reflect cultural differences in their samples. Since cultural rules largely shape social interactions (see, e.g., Hong & Woody, 2007), the subjective experience of SA is likely to vary with the cultural context (Kirmayer, 2001).

Summarizing, there is no evidence of a consistently replicable factor structure for the LSAS-SR, and, relevant for this study in which we recruited Italian participants, it seems to depend on the specific cultural context, over and above differences in sampling method, demographic differences in the sampled population (e.g., age, sex, clinical status, etc.), and researcher's decisions throughout the factor-analytic process.

The present study

The present study aimed to test, for both the Fear and Avoidance scales, whether a common measurement model for the Italian version of the LSAS-SR is equivalent between individuals with and without SAD, as well as between female and male groups. This issue was tackled in an Item Response Theory (IRT) framework using the Graded Response Model (GRM, Samejima, 1968). To the best of our knowledge, the LSAS-SR has never been extensively examined from the perspective of IRT (but see Sunderland et al., 2018). IRT provides a detailed item-level analysis, which gives insight into the functioning of individual items and the relationship between latent construct levels and item endorsement. Moreover, IRT analyses enable a comparison between the functioning of individual items among different samples, providing information about their ability to evaluate the construct equivalently between groups. However, the GRM requires that the item pool is unidimensional and, given the inconsistency of previous results about the dimensionality of the LSAS-SR, we used an exploratory approach to test whether an alternative measurement model could fit the data better than a single-factor one, following the suggestions of Schmitt et al. (2018).

Materials and methods

Participants

Two groups of participants were recruited. The first included 257 Italian adults who were referred to a center for the diagnosis and treatment of anxiety disorders (Centro d'Eccellenza per il Disturbo d'Ansia Sociale - CEDAS, Florence, Italy) and who met the criteria of DSM-IV-TR (APA, 2000) for SAD. Participants were interviewed by one of the members of the research team, who were all doctoral psychologists experienced in diagnosing psychiatric disorders, using the *Structured Clinical Interview* for DSM-IV (SCID-IV; First et

al., 2002) in the context of a routine assessment procedure. Interviewers had undergone a training of at least three months that included reviewing written cases, discussing item-by-item administration with an experienced clinician in SAD diagnosis, observing interviews, and administering interviews while being observed and supervised. All interviewers are members of the same Center for Social Anxiety Disorder diagnosis and therapy and perform regular supervision. The diagnostic interviews could not be recorded, thus preventing us from computing inter-rater reliability. Unstructured clinical notes on any doubt that could emerge during the interviews were carefully reviewed with the other team members. An experienced clinician (Nicola Marsigli) in diagnosis of SAD provided the final decision in the diagnostic process. To be eligible for the study, the participants needed to be at least 18 years old and have been diagnosed with social anxiety disorder. They gave their informed consent before completing the measures. We considered exclusion criteria psychotic and acute substance-related disorders, personality disorders, organic mental disorders, and/or regular use of psychotropic medications, while other comorbid mental disorders were included.

The second group comprised 356 Italian community-dwelling adult participants recruited opportunistically by undergraduate and master psychology students among their acquaintances in partial fulfillment of their research training. To be included, these participants must not report any history of psychiatric or psychological disorder or the exclusion criteria listed above.

In either sample, participation in the study was voluntary and we did not provide incentives to the participants. Descriptive statistics of the background variables of the two groups are reported in Table 1. It should be noted that two participants in the community-dwelling sample did not report their sex. As a result, the number of cases in which sex differences were tested does not match the number of cases in which differences between the community-dwelling and the SAD-diagnosed groups were tested.

[Table 1]

Measures

Liebowitz Social Anxiety Scale (LSAS-SR). When we designed the present study, no established Italian version of LSAS-SR was available. We could find a report on the use of the clinician-administered version on a small sample of patients (Ferrari & Bertazzoni, 2006), and translations could occasionally be found on the Internet, but without any evidence of reliability and validity. Therefore, to carry out this study with Italian participants, we had to develop an adaptation of the LSAS-SR to the Italian context. We used the back-translation method proposed by Brislin (1986). Four researchers fluent in English independently

translated the LSAS-SR items. The four different Italian versions were then compared, and the discrepancies were resolved. To ensure accuracy, a native English speaker, not affiliated with the study, independently back-translated the LSAS-SR items into the original language. This version and the original one were compared. Since no significant discrepancies were found, the final Italian version could be considered capable of capture the meaning of the original one (see the Supplementary Materials [SM] SM1 section for more details).

To collect evidence of the adequacy of the psychometric properties of our version, all participants completed a battery that included other measures of SA, depression, anxiety, worry, and obsessive-compulsive symptomatology. A subsample of 55 community-dwelling participants agreed to fill out the scale again at a four-week interval to assess its test-retest reliability. The results are reported in SM2 and support the construct validity, internal consistency, and retest reliability of the LSAS-SR version used in this study.

Procedure

Participants with SAD were asked by their therapists to complete the questionnaires online through a SurveyMonkey link. They were told that they should complete the questionnaires alone in a quiet room, answering the questions as honestly as possible. Community-dwelling participants were recruited by psychology students as a partial fulfillment of the requirements for their research training. Students were instructed to contact participants who they knew did not have a history of psychological and/or psychiatric issues, present them with informed consent, and hand them out the paper-and-pencil version of the questionnaire. As a result, this was a convenience sample and did not reproduce the known demographic characteristics of the Italian population.

Participants were assured that their answers would be kept in strict confidentiality and would be used only for the purposes of this study, and that the results would be reported in aggregate form only and could not be identified individually.

The order of presentation of the questionnaires was randomized in the online version and balanced using the Latin square design in the paper-and-pencil version to avoid biases caused by order and/or sequence. The questionnaire also asked for demographic information, did not ask for personally identifiable information, and took about 30 minutes to complete. Participants were debriefed after they completed their questionnaires.

In the institutions involved in the study, there is no strict need to receive ethical approval to carry out studies that involve only questionnaire administration, provided that participants receive and sign an informed consent form and the study procedures follow with

the American Psychological Association (2017) guidelines on the ethical treatment of human participants and the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards, which was the case of this study.

The informed consent form informed participants about the purpose of the research, expected duration and procedures; their right to decline to participate and to withdraw from the research once participation has begun; that there were no negative consequences of declining or withdrawing; that there were no potential risks, discomfort, or adverse effects; that there was no compensation for participation; that the data from this study would have been published only in aggregate form, and that no individual data would have been disclosed; whom to contact for any question about the research and research participants' rights.

Data analysis strategy

First, we evaluated the frequency distributions of the scores for each item, in each scale (i.e., LSAS-SR Fear and Avoidance), and in each sample to examine the minimum and maximum scores and ensure that all data were within range (i.e., all values on the answer scale had been endorsed at least once). We also assessed the extent of the missing data. There were no missing data in the SAD-diagnosed participants, while there were 10 and 9 items in the Fear and Avoidance scales, respectively, with at least one missing value in the community-dwelling participants. When considering grouping by sex, we found 7 and 7 items in the Fear and Avoidance scales, respectively, in the female sample, and 4 and 3 items in the Fear and Avoidance scales, respectively, in the male sample. The amount of missing data in each item never exceeded three (0.84%, see SM3 for details), and the missing data were imputed using the *imputeMissing* function in the *mirt* package (Chalmers, 2012) in *R*.

We planned to test the measurement model of the LSAS-SR items in each scale using the graded response model (GRM; Samejima, 1968), which represents a family of mathematical models that deal with ordered polytomous categories, as in the case of the LSAS-SR items. The GRM is a two-parameter logistic (2-PL) model developed for use with items that possess two or more ordinal response categories. This model estimates a unique discrimination (slope) parameter for each item across the k ordinal response categories along with $k - 1$ between-category thresholds for items with more than two categories.

However, as for any IRT model, the accurate estimation of GRM item parameters and their subsequent applications critically depend on the degree to which item response data meet the unidimensionality assumption, i.e., whether there is only one latent trait underlying

the data and whether the item response matrix is locally independent after removing a single common latent factor (see, e.g., Hattie, 1985). Reise et al. (2015) have argued that psychological traits usually have various operationalizations consistent with their definition. Thus, to achieve content validity, a large and heterogeneous sample of behaviors must be drawn. This implies that the item pool of complex, multifaceted traits is unlikely to meet strictly the assumptions of unidimensionality, but this does not mean that it cannot be "unidimensionally enough" for an IRT model, i.e., the impact of unmodeled multidimensionality is negligible. The coverage of this issue is beyond the scope of this work, but we provide more information in SM4.

We first examined the dimensionality of the LSAS-SR item pools (separately for each scale) using data from participants in the community-dwelling, SAD-diagnosed, male, and female groups to provide support for the hypothesized measurement model of the LSAS-SR scales in each group as a precursor to measurement equivalence/invariance analyses. Dimensionality was investigated using several methods. First, we used the scree-test, parallel analysis, and minimum average partial (MAP) correlation statistic (Velicer, 1976) on the polychoric correlation matrices of the LSAS-SR items (see SM5 for the rationale of the use of these methods). Following the suggestion by Schmitt et al. (2018), we tested different models in an effort to better understand the data generating process and factor structure, given the inconclusive results of previous studies.

Second, we considered whether the single factor accounted for at least 10% of the variance of each item (i.e., factor loadings on the single factor should be .30 or more) and computed three recommended indices for the assessment of unidimensionality. Omega (McDonald, 1999) is a measure of composite reliability designed for congeneric scales, i.e., scales in which the items may vary in how strongly they are related to the construct measured. From a factor analysis perspective, the factor loadings are not assumed to be equal. This implies that the items do not meet the criteria for tau equivalence but have to be considered as congeneric. This sort of reliability is considered appropriate when the raw scores of the items on a scale are summed up to yield a total score, and thus they are equally weighted (Bentler, 2007). Omega values larger than .80 are considered as a necessary, although not sufficient, condition for unidimensionality (Rodriguez et al. 2016).

The factor score determinacy (FSD) coefficient is the common variance or the correlation of the factor score with the corresponding factor (Krijnen, 2006). It can also be conceived as the multiple correlation of the observed variables with the corresponding factor (Grice, 2001) or as the proportion of variance of the scores on a factor explained by the

corresponding factor score. Thus, the FSD coefficient allows for an evaluation of the validity of factor scores. Gorsuch (1983, p. 260) suggested that an FSD value larger than .90 can be considered adequate for research.

The *H* index of construct replicability (or construct reliability) is a measure of how well a latent variable is represented by a given set of items and therefore replicable across studies (Hancock & Mueller, 2001). It can be considered as an index of the quality of the items. According to the definition provided by Hancock and Mueller (2001, p. 202), "the quantity represented by *H* equals the population squared multiple correlation, P^2, from regressing the construct on its indicators, that is, the proportion of variability in the construct explainable by its own indicator variables". It provides the correlation between a factor and an optimally weighted item composite, and high values of this index indicate that the latent variable is well defined by its indicators and will not change across studies. Hancock and Mueller (2001) suggested .70 as the criterion for considering an *H* index as adequate.

The goodness of fit (GOF) of the GRM models was tested using the C_2 statistic (Cai & Monroe, 2014), which is an omnibus limited-information goodness-of-fit (GOF) statistic that has an approximate χ^2 distribution. As the traditional χ^2 statistic, C_2 is sensitive to sample size, and therefore it is likely to reject the null hypothesis of adequate fit in large samples. However, as in the tradition of structural equation modeling (SEM), fit indices such as the comparative fit index (CFI_{C2}), the Tucker-Lewis index (TLI_{C2}), and the root mean square error of approximation ($RMSEA_{C2}$), along with its 95% confidence interval, can be computed. To the best of our knowledge, there are no recommended guidelines for interpreting these indices when calculated from the C_2 statistic. We thus used the common cutoffs for the fit of a single model, i.e., RMSEA $\leq .06$ and $\leq .08$ for excellent and acceptable fit, respectively, and CFI and TLI $\geq .95$ and $\geq .90$ for excellent and acceptable fit, respectively (see Marsh et al., 2005). As a result, their suitability for the GRM models still has yet to be investigated more thoroughly and they should not be treated as "golden rules" or used for inferential purposes; they should only serve as rough guidelines to be used with caution when evaluating the global fit of IRT models. We also used Orlando and Thissen's (2003) S-X^2 fit index to evaluate item fit, as it has been shown that this statistic performs adequately with the GRM, too (Kang & Chen, 2011).

The same criteria for the GOF assessment were applied when we evaluated the measurement invariance of the LSAS-SR across the samples (SAD-diagnosed vs. community-dwelling, and females vs. males), i.e., whether the association between the LSAS-SR items and the latent factor depended on group membership. For the comparison of GOF across

models, we chose not to consider the chi-square difference test for the reasons provided above, but, again, we relied on criteria widely used in SEM, i.e., a change in the CFI of less than .01 (Chen, 2007; Cheung & Rensvold, 2002) and a change in the RMSEA of less than .015 (Chen, 2007).

We then specified a configural invariance model (M1, all parameters, slopes, and thresholds were freely estimated), a weak or metric invariance model (M2, slopes, or discrimination parameters, were constrained to be invariant across groups), and a strong or scalar invariance model (M3, slopes, and thresholds were constrained to be invariant across groups; more details about this model can be found in SM6. M3 allowed us to test differences in latent means and differential item functioning (DIF), i.e., whether individuals in different samples but with the same level of SA (as measured by the LSAS-SR) had different probabilities of endorsing a certain answer to a certain item. The presence of DIF for all items would prevent considering the estimates of the difference in the latent means as a valid test for differences in mean levels of SA, while the absence of DIF would be the ideal outcome. More realistically, we could expect partial invariance of the thresholds, that is, that thresholds would be invariant for some items and not for others (M3p). To identify the parameters of interest, the invariance of at least two indicators per latent trait is considered sufficient (Byrne et al., 1989). To detect DIF, we used the method described by Meade and Wright (2012, more details in SM6 and considered the expected score standardized difference (ESSD) as an indicator, since it can be interpreted as the commonly used Cohen's *d* (i.e., $|d| < 0.20$ negligible effect; $0.20 \leq |d| < 0.50$: small effect; $0.50 \leq |d| < 0.80$: moderate effect; $|d| \geq 0.80$ large effect).

After obtaining evidence of unidimensionality and measurement invariance across participants with and without diagnosis of SAD for the LSAS scales, we computed cutoff scores using the Receiver Operating Characteristic method (Youden's index; Youden, 1950), along with all the relevant indices (Area Under the Curve, Specificity, Sensitivity, Positive and Negative Predictive Power).

Results

We initially investigated the dimensionality of the LSAS-SR scales. The results of the scree-test, the PA, and the MAP are shown in Figure 1. In all cases, we found evidence of a strong first factor, although with some evidence of multidimensionality.

[Figure 1]

The optimal number of factors suggested by the dimensionality analyses ranged between 1 and 4. We carried out exploratory factor analyses (EFAs) using MINRES extraction from the polychoric correlation matrix, with oblimin rotation, setting to 1, 2, 3, and 4 the number of factors to be extracted. The multifactor solutions did not show evidence of 'approximate simple structure' (Sass & Schmitt, 2010), that is, a solution with all items substantially loading (> .30) on only one factor, with near-zero cross-loadings, and with factors defined by at least three indicators. Instead, we found that in all solutions there was at least one item with more than one 'substantial' (i.e., > |.30|) loading or with no substantial loading at all. Furthermore, some factor correlations exceeded .70, suggesting redundancy (see SM5 for details). In any case, the results did not indicate a convincing, replicable multifactor measurement model for the LSAS-SR items of either scale. Since the LSAS was not developed with a clear multidimensional measurement model in mind, it is not surprising that we could not find conclusive evidence of multidimensionality. In principle, such a result could be achieved by refining the scale post hoc, i.e., removing problematic items and refitting the models. However, even if this procedure led to the desired outcome, it would not guarantee that the content validity of the reduced item pool would not be impaired and that the results would be replicable, given that any post hoc modification is likely to capitalize on the chance characteristics of the data at hand (see, e.g., MacCallum et al., 1992).

On the other hand, we could find support for the single-factor solution, as the bootstrapped confidence intervals of the factor loadings either included .30 or their lower bound was larger than this value (see SM7), and the thresholds of unidimensionality indices described in the Method section were always exceeded (Table 2). Therefore, we assumed that the scales could be considered sufficiently unidimensional and proceeded with fitting a GRM model.

[Table 2]

The GOF indices of the GRM models are reported in Table 2. They show that the GRM models had an acceptable fit to the data. The parameter estimates for these models are reported in SM8. Baker (2001) suggested that items with discrimination parameters lower than .65 poorly discriminate between high and low levels of the latent trait. Item 17 ('Taking a test of your ability, skill, or knowledge') in the Fear scale in the SAD-diagnosed, female, and male groups, and items 3 ('Eating in public'), 4 ('Drinking with others'), 13 ('Urinating in a public bathroom'), and 17 in the Avoidance scale in the SAD-diagnosed sample showed discrimination values lower than this threshold. However, only the 95% confidence interval

of Fear item 17 in the male sample did not include the threshold value. On the other hand, the highest discriminating items were items 15 ('Being the center of attention') and 16 (Speaking up at a meeting') in the community-dwelling group, and items 10 ('Calling someone you don't know very well'), 11 ('Talking face to face with someone you don't know very well'), and 12 ('Meeting strangers') in the SAD-diagnosed group. A similar pattern of results was found for females and males, too.

No item showed a statistically significant S-X^2 statistic after Benjamini-Hochberg's (2000) correction of the p-values for false discovery rate. Although for very few items the uncorrected p-value was lower than .05, we could not find evidence of a systematic tendency (see SM9). Taken together, these results suggested that the GRM model adequately fitted the data in all four groups of participants considered in this study.

We then tested the measurement invariance of the Fear scale based on the criteria described earlier, and found support for the configural (M1) and metric (M2) invariance but not for the scalar invariance (M3; Table 2).

Therefore, we ran a DIF analysis and identified the five no-DIF items with the largest slopes, which were used as anchors (items 8 ['Working while being observed'], 11 ['Talking face to face with someone you don't know very well'], 14 ['Entering a room when others are already seated'], 15 ['Being the center of attention'], and 23 ['Giving a party']). We then tested a partial invariance model (M3p), whose fit was adequate and not substantially different from that of M2. This model also allowed us to test the latent mean difference, which was statistically significant (1.89 [1.63; 2.15]; positive values indicate higher scores in the SAD-diagnosed group), with a very large effect size (d = 2.01 [1.77; 2.25]). ESSDs for the Fear items are reported in Figure 2. We observed a large DIF for items 17 ('Taking a test of your ability, skill, or knowledge') and 7 ('Going to a party'), while items 1 ('Using a telephone in public'), 2 ('Participating in a small group activity'), 3 ('Eating in public'), 13 ('Urinating in a public bathroom'), and 18 ('Expressing disagreement or disapproval to someone you don't know very well') showed a moderate DIF (Figure 2).

[Figure 2]

The same pattern of results was observed for the Avoidance scale. M1 and M2 adequately fit the data and did not substantially differ in fit. M3 did not show an adequate fit, and items 2 ('Participating in a small group activity'), 11 ('Talking face to face with someone you don't know very well'), 12 ('Meeting strangers'), 15 ('Being the center of attention'), and 23 ('Giving a party') were identified as anchors. The partial invariance model showed an adequate fit to the data and did not differ substantially in fit from M2. The latent mean

difference was again statistically significant (1.67 [1.43; 1.94]), with a very large effect size (*d* = 1.57 [1.36; 1.80]). ESSDs for the Avoidance items are reported in Figure 2. We observed a large DIF for items 9 ('Writing while being observed'), 13 ('Urinating in a public bathroom'), and 17 ('Taking a test of your ability, skill, or knowledge'), while items 8 ('Working while being observed') and 24 ('Resisting a high pressure salesperson') showed a moderate DIF (Figure 2).

Given the adequate level of invariance of the LSAS-SR measurement model, the comparison of observed scores can be considered valid. This allowed us to reliably compute cutoff scores. The Area Under the Curve was .907 and .861 for the Fear and Avoidance scales, respectively (see SM10 for more details), indicating that the LSAS-SR scales had a high accuracy in distinguishing community-dwelling participants from SAD-diagnosed participants. The optimal cutoff values were 30 [26; 36] and 28 [20; 32] for the Fear and Avoidance scales, respectively. At these values, the Specificity, Sensitivity, and Positive and Negative predictive values ranged from .728 to .864 (Table SM10.1 and Table SM10.2). More details are reported in SM10.

We found evidence of measurement invariance also between sexes (Table SM8.2 and Table SM8.4). The scalar invariance model for female and male participants did not show a substantially different fit from the metric invariance model. This led us to conclude that no relevant DIF could be found when considering sex differences. The latent means significantly differed (Fear: -0.23 [-0.41; -0.05]; Avoidance: -0.20 [-0.39; -0.02]; negative values indicate higher scores in females), although with small effect sizes (Fear: *d* = -0.21 [-0.43; -0.04]; Avoidance: -0.18 [-0.40; -0.02]).

Discussion

Some individuals tend to avoid social situations due to the high levels of distress experienced. Throughout life, this tendency can have severe and global consequences for them. From a clinical perspective, it is crucial to be able to properly assess the amount of fear triggered by some everyday social situations and the extent to which they are avoided. The self-report version of the LSAS (Liebowitz, 1987) is a quick-to-administer and comprehensive measure to assist in this task, as it allows the assessment of both fear and avoidance.

The main aim of the present study was to investigate the measurement invariance of the Italian LSAS-SR between groups defined by having or not received a diagnosis of SAD, on the one hand, and by sex, on the other. However, before testing this crucial psychometric

property, we had to find a measurement model (i.e., factor structure) that could be considered adequately replicable across these samples, as a previous similar study (Kubota et al., 2016) could not find supporting evidence. The dimensionality analyses pointed out the presence of a strong single factor, but also suggested some evidence of multidimensionality. However, the fit of the single-factor model could be considered as adequate in all groups, while the multifactor solutions did not provide conclusive evidence of a measurement model for the Italian LSAS-SR that was both consistent with the (approximate) simple structure assumption and replicable across groups. The misspecification from forcing multidimensional data into a unidimensional measurement model is likely to lead to severely biased and potentially misleading parameter estimates (Reise et al., 2015), but it has been shown that if there is a strong general factor in the data, then the estimation of the IRT item parameters is acceptably unbiased when it fits to a unidimensional measurement model (see, e.g., Kirisci et al., 2001). It has long been acknowledged (e.g., Humphreys, 1970) that a set of items that are strictly unidimensional is not necessarily desirable (if ever possible), since it might consist of items that are basically the same but written in several slightly different ways, thus being highly intercorrelated. Moreover, such a measure would have a very narrow conceptual bandwidth that will ultimately result in poor predictive power and little theoretical and practical usefulness (Reise et al., 2015). On the contrary, a sufficiently unidimensional set of items might allow the assessment of a single target construct while being sufficiently heterogeneous to validly represent the diverse manifestations of the construct, provide acceptable reliability, and avoid redundancy. The issue of how this essential unidimensionality can be evaluated is beyond the scope of the present study, and it is discussed in more detail in Section 4 of the SM. However, it should be noted that the bifactor modeling approach proposed by Reise et al. (2015) could not be applied here, since the LSAS-SR was not originally developed to obtain a multidimensional measure with a clear, replicable, and stable factor structure, neither such a structure has been found in previous studies (see the Introduction). Since the dimensionality and factor analyses did not provide strong evidence of multidimensionality in any of the groups and the indices we used to assess unidimensionality substantially suggested that the LSAS-SR scales could be considered sufficiently unidimensional, we tested the measurement model of the LSAS-SR items using the GRM.

The GRM showed an adequate fit in all groups and allowed the investigation of the discrimination parameters, which are an index of the items' ability to differentiate subjects with different levels of the latent trait (in this case, level of SA). In all groups, the highest discriminating items (Table SM8.1 and Table SM8.3) involved some sort of interaction with

little-known people and being in a situation where the focus is on the participant. From a clinical point of view, these items seem to tap into the core of SA, which has to do with being the object of evaluation by others. Indeed, these items refer to situations in which individuals expose themselves directly to the scrutiny of others. The lowest discriminating items tapped into the fear of being observed during routine activities (i.e., eating, writing, or drinking). These items are likely to have a lower discrimination ability because they describe everyday situations that usually involve familiar people. Therefore, the responses to these items bore greater similarity across groups. Item 17 emerged as the least discriminant. One possible explanation is that the scenario described by it did not activate the core of SA, as it did not provide direct interaction with other people. However, it might also reflect a cultural issue, peculiar to the Italian context in which the study was carried out. Our translation of the item ("Sottoporsi a un esame scritto", which literally means "Taking a written exam") might tap into a narrower content than the original ("Taking a test of your ability, skill, or knowledge"). However, we had to choose this translation because, in Italy, it is relatively uncommon to take tests that are not written achievement tests and/or for reasons different from educational purposes.

Once we established the measurement model for the Italian LSAS-SR scales separately in each group, we could investigate its measurement invariance between the groups. The use of the LSAS-SR for screening purposes is grounded in the assumption that the association between the LSAS-SR items and the latent factor did not depend on group membership and that the measurements themselves operate equally across groups, thus making the comparison of mean scores meaningful. When comparing community-dwelling and SAD-diagnosed groups, we obtained full support for the configural and metric invariance models. Instead, we found evidence of DIF when we specified a scalar invariance model. We observed a large DIF for items 7 and 17 on the Fear scale and items 9, 13, and 17 on the Avoidance scale, which implies that the scores on the items are not fully explained by differences in the levels on the latent trait. This issue did not affect the validity of the latent mean differences test, given that the criterion of invariance of at least two indicators per latent trait was met (Byrne et al., 1989) and a large effect size was found, supporting the criterion validity of the LSAS-SR, as higher mean levels of SA were observed in the SAD-diagnosed groups. This result is consistent with similar studies on SA measures (see, e.g., Bunnell et al., 2013). However, to the best of our knowledge, this is the first time this result has been obtained for the LSAS-SR. Nevertheless, the inspection of the content of the items that showed the largest DIFs suggests that community-dwelling and SAD-diagnosed participants

might interpret differently the amount of fear and the extent of avoidance that activities such as taking an exam, going to a party, writing while being observed, and urinating in a public bathroom may elicit. These differences can be explained by the fact that each of these situations can elicit different threats besides the social ones. For example, people may avoid using public bathrooms for disgust or fear of contamination or to write in front of others on suspicion that they may read what is being written.

The excellent criterion validity of the LSAS-SR scales was also supported by the ROC curve analyses, suggesting an adequate sensitivity and specificity of the LSAS-SR as a screening tool. We also derived cutoff scores for the Italian population (30 for the Fear scale, 28 for the Avoidance scale), which were in line with those of most published studies. While in US samples cutoff scores of 30 have been proposed (Mennin et al., 2002; Rytwinski et al., 2009), other national studies have proposed cutoff scores of 19 and 26 (Bobes et al., 1999).

Another aim of this study was to examine the measurement invariance of the Italian LSAS-SR between females and males despite their diagnosis status (i.e., whether they have been diagnosed with SAD or not). We found support for configural, metric, and scalar invariance and, perhaps more interestingly, no evidence of DIF. This result implies that differences in observed mean scores could actually be ascribed to actual differences in the level of SA and rules out the possible effect of non-equivalence between sexes in the functioning of the LSAS-SR items. Although with small effect sizes, females reported higher mean levels of SA than males. This result is consistent with other studies on SA measures (Bunnell et al., 2013), and with recent reviews (Asher et al., 2017) and research (Asher & Aderka, 2018), which have found that women are more likely to have SAD, have more severe symptoms, and a higher number of social fears, especially at younger ages. So far, no conclusive explanation for these differences has been established. Asher et al. (2017) proposed an interpretation in terms of self-construal theory (Cross et al., 2011), which assumes that women tend to develop and maintain an interdependent self-construal, in which others are represented as part of the self, while in men's self-construal, others are separate from the self. As a result, women's sense of self depends on their relationships with significant others more than men's. This potentially explains why women can experience more dissatisfaction with life and negative affect than men when they feel misunderstood in interpersonal interactions and perceive a lack of relationship harmony (Reid, 2004). Therefore, these differences in self-construal can explain the greater level of anxiety experienced by women in social situations, which may represent potential sources of scrutiny, negative evaluation, and potential rejection by others. Alternative explanations, relevant in the

case of scores on a self-report measure, suggest that men tend to underreport their SA levels and/or interpret some questions differently, which is consistent with theories about self-discrepancy (Higgins, 1987) and identity-discrepancy (Marcussen & Large, 2003). As traditional gender stereotypes depict men as self-confident, assertive, and dominant (Wood & Eagly, 2012), they might perceive a higher discrepancy between their actual and ideal/ought selves due to their social anxiety. They might also be more reluctant to report such feelings than women in the same situation and, as such, might perceive less discrepancy in their ought or ideal social roles—although, according to Asher et al. (2017), this might explain why men are more likely to seek treatment for SAD. However, the lack of DIF suggests that this might not be the case, at least for the content of the items included in the LSAS-SR.

Limitations

Some limitations of the present work should be acknowledged. First, the data were collected on Italian opportunistic samples. Although this is almost unavoidable for clinical participants and, more generally, is common in this field of research, this sampling strategy undermines the understanding of the full extent and degree of generalizability of the present results. Therefore, they may not necessarily generalize to other Italian contexts. Given the variability of LSAS-SR psychometric properties across cultural contexts (see Introduction), these results may not generalize to the English or other translated versions of the instrument. However, this issue would be better addressed by a cross-national study that would control for sampling method, demographic differences, and statistical methods. On a related note, the data in Table 1 indicate some differences between the community-dwelling and the SAD-diagnosed groups in the background variables, the most relevant of which is the moderate difference in age (d = 0.53). As the SAD-diagnosed participants were younger, this can explain why they were slightly less educated (r = .27), more likely to be non-married (r = .18), and less likely to work (r = .19). We thus used Propensity Score Analysis (Rosenbaum & Rubin, 1983) to compute the conditional probability (i.e., propensity) of being in the SAD-diagnosed group given the background variables and to weight the data based on these propensity scores. We then performed the same analyses of the manuscript using the weighted data, and we could not find substantial differences in the results. Therefore, we concluded that these differences in background variables did not have a substantial effect (if any) on the results of the invariance tests.

Second, it might be argued that the sample sizes could not be adequate. For instance, Reise and Yu (1990) reported that as many as 500 participants are needed to achieve an

adequate calibration under the graded response model. A common procedure to establish the adequacy of the sample size for a latent variable model is to run a Monte Carlo simulation. Muthén and Muthén (2002) suggested that when parameter and standard error biases do not exceed 10% for any parameter in the model, and coverage (i.e., the proportion of replications for which the 95% confidence interval contains the true parameter value) remains between .91 and .98, the sample size allows keeping statistical power close to .80. We carried out these analyses (using the *simdata* function in the *R* package *mirt*), but their results were impaired by the low base rate of endorsement of the higher score in some items (see SM3). This led to many replications in which the highest score for some items was 2 instead of 3, especially in the community-dwelling participants. This prevented us from performing this analysis for this sample and for the invariance models. However, the simulations were successful for the SAD-diagnosed sample (see SM11) and revealed that the bias obtained with the available sample size (n = 257) met the Muthéns' criteria for parameter estimates and coverage, while some items did not meet the criterion for standard errors. These results indicated that, while the study could have been slightly underpowered, any resulting bias was unlikely to severely undermine the validity of the results. However, they also suggest that similar future studies should try and include a higher number of participants.

Third, another limitation can be found in the collection of data on the self-report version of the LSAS. As with any other scale of this sort, the data could have suffered from biases such as social desirability and short-term recall bias. Since a clinician-administered version of the LSAS is available, future studies may evaluate the consistency of the scores across the administration methods.

Fourth, it should be noted that participants in the community-dwelling sample did not undergo a formal diagnostic assessment but were included in a "non-diagnosed" group simply because of a lack of self-reporting of psychiatric or psychological disorders. As a result, we cannot exclude that some community-dwelling participants could actually meet the criteria for a diagnosis of SAD (or for any other psychological disorder). Unfortunately, it was not possible to screen these participants as we did with those of the SAD-diagnosed sample, and for the Italian versions of other measures of SA there is no robust evidence of reliable cutoff scores that could assist in the screening process.

This issue is related to another limitation of this study, namely, the different administration methods for the two samples (online for SAD-diagnosed participants, paper-and-pencil for community-dwelling participants). As the paper-and-pencil administration involved direct interaction with participants, it potentially allowed data collectors to screen

for ineligible participants (e.g., individuals that clearly did not fill out the questionnaire seriously). The same could not have been done with online administration unless all participants were contacted directly and were very well known by recruiters. Although not impossible in practice, this would have had the side effect of recruiting a much smaller and/or even less representative sample of community-dwelling participants. On the other hand, not all participants in the SAD-diagnosed group lived in the city of the institution to which they were referred. Therefore, administering the questionnaires online allowed us to avoid using session time for data collection and/or potential inconveniences such as, e.g., returning the batteries via regular mail. In general, there is mixed evidence about the equivalence of online vs. paper-and-pencil versions of the same questionnaires, especially in clinical contexts. While some studies report substantial differences (e.g., Buchanan, 2003), others report substantial equivalence (e.g., Campbell et al., 2015). Two previous studies (Hedman et al., 2010; Hirai et al., 2011) investigated the effect of paper-and-pencil vs. online administration of social anxiety measures in nondiagnosed participants, and found similar means, internal consistency reliability estimates, construct validity, and sex-specific mean scores (with no sex by group interaction) across the administration groups. However, Hirai et al. (2011) reported a lack of measurement invariance across administration methods at the weak invariance level (i.e., factor loadings). Our results are at odds with this finding, as we found evidence of invariance of discrimination parameters (i.e., the IRT equivalent of factor loadings), although we cannot exclude that some confounding effects might have played a role. However, without a study specifically focused on this issue, it is not possible to draw definitive conclusions.

Conclusions

Despite its limitations, the present work suggests that the Italian LSAS-SR is a sound psychometric tool for the assessment of SA for screening and assessment purposes. We provided evidence for the measurement invariance of a single-factor structure for either scale in community-dwelling and SAD-diagnosed participants, as well as in female and male participants. As such, the Italian LSAS-SR can be confidently used to compare levels of SA across these samples. From these premises, future studies should investigate its measurement invariance across time and other subpopulations defined by age (e.g., adolescents vs. adults) or other relevant background characteristics.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The Authors wish to thank **********

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

References

Aderka, I. M., Hofmann, S. G., Nickerson, A., Hermesh, H., Gilboa–Schechtman, E., & Marom, S. (2012). Functional impairment in social anxiety disorder. *Journal of Anxiety Disorders*, *26*(3), 393–400. https://doi.org/10.1016/j.janxdis.2012.01.003.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. http://apa.org/ethics/code/index.aspx.

Asher, M., & Aderka, I. M. (2018). Gender differences in social anxiety disorder. *Journal of Clinical Psychology, 74*(10), 1730–1741. https://doi.org/10.1002/jclp.22624.

Asher, M., Asnaani, A., & Aderka, I. M. (2017). Gender differences in social anxiety disorder: A review. *Clinical Psychology Review, 56*, 1–12. https://doi.org/10.1016/j.cpr.2017.05.004.

Assunção, M. C., da Conceição Costa, D. L., de Mathis, M. A., Gedanke Shavitt, R., Ferrão, Y. A., do Rosário, M. C., Miguel, E. C., & Rodrigues Torres A. (2012). Social phobia in obsessive–compulsive disorder: Prevalence and correlates. *Journal of Affective Disorders, 143*(1–3), 138–147. https://doi.org/10.1016/j.jad.2012.05.044.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Baker, S. L., Heinrichs, N., Kim, H. J., & Hofmann, S. G. (2002). The Liebowitz Social Anxiety Scale as a self–report instrument: A preliminary psychometric analysis.

Behaviour Research and Therapy, 40, 701–715. https://doi.org/10.1016/S0005–7967(01)00060–2.

Beard, C., Rodriguez, B. F., Weisberg, R. B., Ashley Perry, A., & Keller, M. B. (2012). Psychometric properties of the Liebowitz Social Anxiety Scale in a longitudinal study of Latinos with anxiety disorders. *Hispanic Journal of Behavioral Sciences*, *34*(2), 269–278. https://doi.org/10.1177/0739986312436660.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83. https://doi.org/10.3102/10769986025001060.

Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit–weighted composites. In S. Lee (Ed.), *Handbook of computing and statistics with applications: Vol. 1* (pp. 1–19). Elsevier.

Bobes, J., Badia, X., Luque, A., Garcia, M., González, M. P., & Dal–Ré, R. (1999). Validation of the Spanish version of the Liebowitz social anxiety scale, social anxiety and distress scale and Sheehan disability inventory for the evaluation of social phobia. *Medicina Clínica, 112*(14), 530–538.

Bögels, S. M., Alden, L., Beidel, D. C., Clark, L. A., Pine, D. S., Stein, M. B., & Voncken, M. (2010). Social anxiety disorder: questions and answers for the DSM–V. *Depression and Anxiety*, *27*(2), 168–189. https://doi.org/10.1002/da.20670.

Brislin, R. W. (1986). The wording and translation of research instruments. In: W. L. Lonner, & J.W. Berry (Eds.), *Field methods in cross–cultural research* (pp. 137–164). Sage Publications.

Buchanan, T. (2003). Internet–based questionnaire assessment: Appropriate use in clinical contexts. *Cognitive Behaviour Therapy*, *32*(3), 100–109. https://doi.org/10.1080/16506070310000957.

Bunnell, B. E., Joseph, D. L., & Beidel, D. C. (2013). Measurement invariance of the Social Phobia and Anxiety Inventory. *Journal of Anxiety Disorders*, *27*(1), 84–91. https://doi.org/10.1016/j.janxdis.2012.09.001.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033–2909.105.3.456

Caballo, V. E., Salazar, I. C., Arias, V., Hofmann, S. G., & Curtiss, J. (2019). Psychometric properties of the Liebowitz Social Anxiety Scale in a large cross–cultural Spanish and

Portuguese speaking sample. *Brazilian Journal of Psychiatry, 41(*2), 122–130. https://doi.org/10.1590/1516–4446–2018–0006.

Caballo, V. E., Salazar, I. C., Irurtia, M. J., Arias, B., Hofmann, S. G., & Ciso–A Research Team. (2014). Differences in social anxiety between men and women across 18 countries. *Personality and Individual Differences, 64*, 35–40. https://doi.org/10.1016/j.paid.2014.02.013.

Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Technical Report. http://cresst.org/publications/cresst–publication–3208/.

Campbell, N., Ali, F., Finlay, A. Y., & Salek, S. S. (2015). Equivalence of electronic and paper–based patient–reported outcome measures. In *Quality of Life Research* (Vol. 24, Issue 8, pp. 1949–1961). Kluwer Academic Publishers. https://doi.org/10.1007/s11136–015–0937–3.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: *A Multidisciplinary Journal, 4*(3), 464–504. https://doi.org/10.1080/10705510701301834.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness–of–fit indexes for testing measurement invariance. *Structural equation modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Coelho, H. F., Cooper, P. J., & Murray, L. (2007). A family study of co–morbidity between generalized social phobia and generalized anxiety disorder in a non–patient sample. *Journal of Affective Disorders, 100*(1–3), 103–113. https://doi.org/10.1016/j.jad.2006.10.001.

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Foa, E., & Weisler, R. H. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). *British Journal of Psychiatry, 76*, 379–386. https://doi.org/10.1192/bjp.176.4.379.

Cox, B. J., Fleet, C., & Stein, M. B. (2004). Self–criticism and social phobia in the US national comorbidity survey. *Journal of Affective Disorders, 82*(15), 227–234. https://doi.org/10.1016/j.jad.2003.12.012.

Cox, B. J., Ross, L., Swinson, R. P., & Direnfeld, D. M. (1998). A comparison of social phobia outcome measures in cognitive–behavioral group therapy. *Behavior Modification, 22*(3), 285–297. https://doi.org/10.1177/01454455980223004.

Cross, S. E., Hardin, E. E., & Gercek–Swing, B. (2011). The what, how, why, and where of self–construal. *Personality and Social Psychology Review*, *15*(2), 142–179. https://doi.org/10.1177/1088868310373752.

Davidson, J. R., Miner, C. M., De Veaugh–Geiss, J., Tupler, L. A., Colket, J. T., & Potts, N. L. S. (1997). The brief social phobia scale: A psychometric evaluation. *Psychological Medicine, 27*(1), 161–166. https://doi.org/10.1017/s0033291796004217.

Deller, J., Perrotte, J., Wainwright, K., Brunsman, J., & Osman, A. (2020). Dimensionality, reliability, invariance, and validity of the Multidimensional Social Anxiety Response Inventory–21 (MSARI–21). *Journal of Personality Assessment*, *102*(4), 527–537. https://doi.org/10.1080/00223891.2019.1569529.

Ferrari, A. & Bertazzoni, S. (2006). Fobia sociale in utenti di un centro di salute mentale impegnati in attività di laboratorio protetto. *Lo Spallanzani*, *20*, 93–98.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM–IV–TR axis I disorders, research version, patient edition (SCID–I/P).* Biometrics Research, New York State Psychiatric Institute.

Forni dos Santos, L., Loureiro, S. R., Crippa, J. A. d. S., & Osorio, F. d. L. (2013). Psychometric validation study of the Liebowitz Social Anxiety Scale – Self–Reported version for Brazilian Portuguese. *PlosOne, 8*(7), e70235. https://doi.org/10.1371/journal.pone.0070235.

Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001) The Liebowitz Social Anxiety Scale: A comparison of psychometric properties of self–reported and clinical–administered formats. *Psychological Medicine, 31*, 1025–1035. https://doi.org/10.1017/S003329170105405.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. https://doi.org/10.1037/1082–989X.6.4.430.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future: A festschrift in honor of Karl Jöreskog* (pp. 195–216). Scientific Software International.

Harvard Medical School (2007). *National Comorbidity Survey (NCS)*. Retrieved from https://www.hcp.med.harvard.edu/ncs/index.php.

Hasani, J., Akbari, E., & Fayazi., M. (2017). Reliability, validity, and confirmatory factor structure of Persian version of Liebowitz Social Anxiety Scale (LSAS). *Social Psychology Research, 6*(24), 35–51.

Hattie, J. (1985). Assessing the unidimensionality of test and items. *Applied Psychological Measurement, 9*, 139–164. https://doi.org/10.1016/S0305–0483(96)00051–5.

Hedman, E., Ljótsson, B., Rück, C., Furmark, T., Carlbring, P., Lindefors, N., & Andersson, G. (2010). Internet administration of self–report measures commonly used in research on social anxiety disorder: A psychometric evaluation. *Computers in Human Behavior, 26*(4), 736–740. https://doi.org/10.1016/j.chb.2010.01.010.

Heimberg, R. G., & Holaway, R. M. (2007). Examination of the known–groups validity of the Liebowitz Social Anxiety Scale. *Depression and Anxiety, 24*, 447–454. https://doi.org/10.1002/da.20277.

Higgins, E. T. (1987). Self–Discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319–340. doi:10.1037/0033–295X.94.3.319.

Hirai, M., Vernon, L. L., Clum, G. A., & Skidmore, S. T. (2011). Psychometric properties and administration measurement invariance of social phobia symptom measures: Paper–pencil vs. internet administrations. *Journal of Psychopathology and Behavioral Assessment, 33*(4), 470–479. https://doi.org/10.1007/s10862–011–9257–2.

Hong, J. J., & Woody, S. R. (2007). Cultural mediators of self–reported social anxiety. *Behaviour Research and Therapy, 45*(8), 1779–1789. https://doi.org/10.1016/j.brat.2007.01.011.

Humphreys, L. G. (1970). A skeptical look at the factor pure test. In C. E. Lunneborg (Ed.), *Current problems and techniques in multivariate psychology: Proceedings of a conference honoring Professor Paul Horst* (pp. 23–32). University of Washington. https://doi.org/10.1177/0013164494054002002.

Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X^2 item fit index for the graded response model. *Asia Pacific Education Review, 12*(1), 89–96. https://doi.org/10.1007/s12564-010-9082-4.

Kashdan T. B., & Roberts, J. E. (2011). Comorbid social anxiety disorder in clients with depressive disorders: Predicting changes in depressive symptoms, therapeutic relationships, and focus of attention in group treatment. *Behaviour Research and Therapy, 49*, 875–884. https://doi.org/10.1016/j.brat.2011.10.002.

Kirisci, L., Hsu, T., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*(2), 146–162. doi:10.1177/01466210122031975.

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: Implications for diagnosis and treatment. *Journal of Clinical Psychiatry, 62*(Suppl. 13), 22–30.

Krijnen, W. P. (2006). Some results on mean square error for factor score prediction. *Psychometrika, 71*(2), 395–409. https://doi.org/10.1007/s11336–004–1220–7.

Kubota, R., Chen, J., Peters, L., Skuza, P., Furukawa, T. A., & Nagata, T. (2016, July 24–29). *Liebowitz Social Anxiety Scale (LSAS): Testing Measurement Equivalence across Cultures* [Conference presentation]. ICP 2016 31st International Congress of Psychology, Yokohama, Japan.

Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of Personality Assessment, 47*(1), 66–75. https://doi.org/10.1207/s15327752jpa4701_8.

Levin, J. B., Marom, S., Gur, S., Wechter, D., & Hermesh, H. (2002). Psychometric properties and three proposed subscales of a self–report version of the Liebowitz Social Anxiety Scale translated into Hebrew. *Depression and Anxiety, 16*, 143–151. doi:10.1002/da.10064.

Levinson, C. A., & Rodebaugh, T. L. (2012). Social anxiety and eating disorder comorbidity: The role of negative social evaluation fears. *Eating Behaviors, 13*(1), 27–35. https://doi.org/10.1016/j.eatbeh.2011.11.006.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22*, 147–173. https://doi.org/10.1159/000414022.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490–504. https://doi.org/10.1037/0033–2909.111.3.490.

Marcussen, K., & Large, M. D. (2003). Using identity discrepancy theory to predict psychological distress. In P. J. Burke, T. J. Owens, R. T. Serpe, & P. A. Thoits (Eds.), *Advances in identity theory and research* (pp. 151–164). Kluwer Academic/Plenum. https://doi.org/10.1007/978–1–4419–9188–1_11.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self–rating for phobic patients. *Behaviour Research and Therapy, 17*(3), 263–267. https://doi.org/10.1016/0005–7967(79)90041–X.

Marsh, H. W., Hau, K.–T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu–Olivares & J. McArdle (Eds.), *Contemporary psychometrics. A Festschrift for Roderick P. McDonald* (pp. 275–340). Lawrence Erlbaum.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005–7967(97)10031–6.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates Publishers.

Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, *97*(5), 1016–1031. https://doi.org/10.1037/a0027934.

Mennin, D. S., Fresco, D. M., Heimberg, R. G., Schneier, F. R., Davies, S. O., & Liebowitz, M. R. (2002). Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale. *Journal of Anxiety Disorders*, *16*(6), 661–673. https://doi.org/10.1016/S0887–6185(02)00134–2.

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8.

Oakman, J., Van Ameringer, M., Mancini, C., & Farvolden, P. (2003). A Confirmatory Factors Analysis of a Self–Report Version of a Liebowitz Social Anxiety Scale. *Journal of Clinical Psychology*, *59*(1), 149–161. https://doi.org/10.1002/jclp.10124.

Olino, T. M. (2020). Clinical applications of measurement invariance. *Journal of Personality Assessment*, *102*(5), 727–729. https://doi.org/10.1080/00223891.2020.1793766

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S–X^2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*(4), 289–298. https://doi.org/10.1177/0146621603027004004.

Osorio, F. L, Crippa, J. A. S, & Loureiro, S. R. (2009). Instruments for the assessment of Social Anxiety Disorder. *Revista de Psiquiatria Clınica*, *32*, 73–83. doi: 10.1590/S0101–60832005000200003.

Reid, A. (2004). Gender and sources of subjective well–being. *Sex Roles*, *51*(11–12), 617–629. doi:10.1007/s11199–004–0714–1.

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27*(2), 133–144. https://doi.org/10.1111/j.1745–3984.1990.tb00738.x.

Reise, S. P., Cook, K. F., & Moore, T. M. (2015). Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling. Applications to typical performance assessment* (pp. 13–40). Routledge.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/https://doi.org/10.1093/biomet/70.1.41

Rytwinski, N. K., Fresco, D. M., Heimberg, R. G., Coles, M. E., Liebowitz, M. R., Cissell, S., Stein, M. B., & Hofmann, S. G. (2009). Screening for Social Anxiety Disorder with the self–report version of the Liebowitz Social Anxiety Scale. *Depression and Anxiety, 26*, 34–38. https://doi.org/10.1002/da.20503.

Safren, S. A., Heimberg, R. G., Horner, K. J., Juster, H. R., Schneier, F. R., & Liebowitz, M. R. (1999). Factor structure of social fears: The Liebowitz Social Anxiety Scale. *Journal of Anxiety Disorders, 13*, 253–270. https://doi.org/10.1016/S0887–6185(99)00003–1.

Samejima, F. (1968). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph, 1968*(1), i–169. https://doi.org/10.1002/j.2333–8504.1968.tb00153.x.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*(1), 73–103. https://doi.org/10.1080/00273170903504810.

Schneier, F. R., Foose, T. E., Hasin, D. S., Heimberg, R. G., Shang–Min Liu, Grant, B. F., & Blanco, C. (2010). Social anxiety disorder and alcohol use disorder comorbidity in the national epidemiologic survey on alcohol and related conditions. *Psychological Medicine, 40*(6), 977–988. https://doi.org/10.1017/S0033291709991231.

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the "best" factor structure and moving measurement validation forward: An illustration. *Journal of Personality Assessment, 100*(4), 345–362. https://doi.org/10.1080/00223891.2018.1449116

Shachar, I., Aderka, I. M., & Gilboa–Schechtman, E. (2014). The factor structure of the Liebowitz social anxiety scale for children and adolescents: Development of a brief version. *Child Psychiatry and Human Development*, *45*(3), 285–293. https://doi.org/10.1007/s10578–013–0398–2.

Skocic, S., Jackson, H., & Hulbert, C. (2015). Beyond DSM–5: An alternative approach to assessing Social Anxiety Disorder. *Journal of Anxiety Disorders*, *30*, 8–15. https://doi.org/10.1016/j.janxdis.2014.12.006.

Slavkin, S. L., Holt, C. S., Heimberg, R. G., Jaccard, J. J., & Liebowitz, M. R. (1990, November). *The Liebowitz Social Anxiety Scale: An exploratory analysis of construct validity* [Conference presentation]. 24th Annual Meeting of the Association for Advancement of Behavior Therapy, San Francisco, CA, United States.

Soykan, C., Ozgüven, H. D., & Gençöz, T. (2003). Liebowitz Social Anxiety Scale: the Turkish Version. *Psychological Reports*, *93*, 1059–1069. https://doi.org/10.2466/pr0.2003.93.3f.1059.

Stein, D. J., Lim, C. C. W., Roest, A. M., de Jonge, P., Aguilar–Gaxiola, S., Al–Hamzawi, A., Alonso, J., Benjet, C., Bromet, E. J., Bruffaerts, R., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Harris, M. G., He, Y., Hinkov, H., Horiguchi, I., Hu, C., … Williams, D. R. (2017). The cross–national epidemiology of social anxiety disorder: Data from the World Mental Health Survey Initiative. *BMC Medicine*, *15*(1), 143. https://doi.org/10.1186/s12916–017–0889–2.

Sugawara, N., Yasui-Furukori, N., Kaneda, A., Sato, Y., Tsuchimine, S., Fujii, A., Danjo, K., Takahashi, I., Matsuzaka, M., & Kaneko, S. (2012), Factor structure of the Liebowitz Social Anxiety Scale in community-dwelling subjects in Japan. *Psychiatry and Clinical Neurosciences*, *66*, 525–528. doi:10.1111/j.1440–1819.2012.02381.x.

Sunderland, M., Batterham, P., Calear, A., Carragher, N., Baillie, A., & Slade, T. (2018). High agreement was obtained across scores from multiple equated scales for social anxiety disorder using item response theory. *Journal of Clinical Epidemiology*, *99*, 132–143. https://doi.org/10.1016/J.JCLINEPI.2018.04.003

Sunderland, M., Crome, E., Stapinski, L., Baillie, A. J., & Rapee, R. M. (2016). From fear to avoidance: Factors associated with the onset of avoidance in people who fear social situations. *Journal of Experimental Psychopathology, 7*(4), 534–548. https://doi.org/10.5127/JEP.055216.

Turner, S. M., McCanna, M., & Beidel, D. C. (1987). Validity of the social avoidance and distress and fear of negative evaluation scales. *Behaviour Research and Therapy*, *25*(2), 113–115. https://doi.org/10.1016/0005–7967(87)90081–7.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002.

Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557.

Watson, D., & Friend, R. (1969). Measurement of social–evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33*(4), 448–457.https://doi.org/10.1037/h0027806.

Wells, A., Clark, D. M., Salkovskis, P., Ludgate, J., Hackmann, A., & Gelder, M. (1995). Social phobia: The role of in-situation safety behaviors in maintaining anxiety and negative beliefs. *Behavior Therapy*, *26*(1), 153–161. https://doi.org/10.1016/S0005-7894(05)80088-7

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In J. M. Olson & P. Zanna (Eds.), *Advances in experimental social psychology* (pp. 55–123). Elsevier. https://doi.org/10.1016/B978–0–12–394281–4.00002–7.

Xu, Y., Schneier, F., Heimberg, R. G., Princisvalle, K., Liebowitz, M. R., Wang, S., & Blanco, C. (2012). Gender differences in social anxiety disorder: Results from the national epidemiologic sample on alcohol and related conditions. *Journal of Anxiety Disorders*, *26*(1), 12–19. https://doi.org/10.1016/j.janxdis.2011.08.006.

Yao, S. N., Note, I., Fanget, F., Albuisson, E., Bouvard, M., Jalenques, I., & Cottraux, J. (1999). Social anxiety in patients with social phobia: validation of the Liebowitz social anxiety scale: The French version. *L'Encéphale*, *25*, 429–435.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097–0142(1950)3:1<32::aid–cncr2820030106>3.0.co;2–3.

Figure captions

Figure 1 Scree-plots, results from the parallel analysis and the Minimum Average Partial (MAP) correlation statistic for the Fear and Avoidance scales of the Liebowitz Social-Anxiety Scale-Self Report in community-dwelling participants (n = 356), participants diagnosed with a social anxiety disorder (SAD), female (n = 390) and male (n = 221) participants.

Figure 2 Expected Score Standardized Difference (ESSD) and Differential Item Functioning of the Liebowitz Social Anxiety Scale - Self-Report. Black dots represent the participants diagnosed the social anxiety disorder (SAD), red dots the community-dwelling sample.

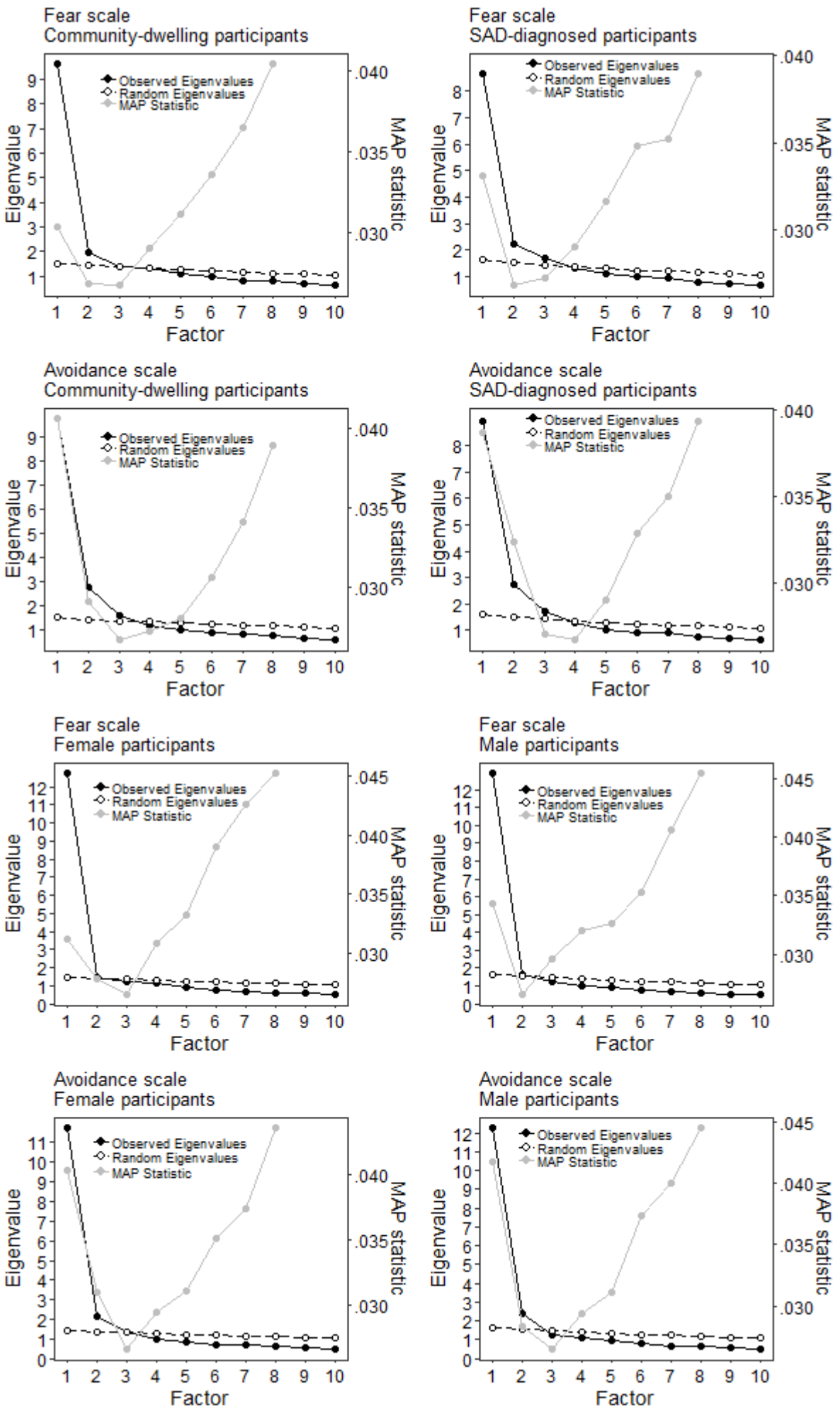
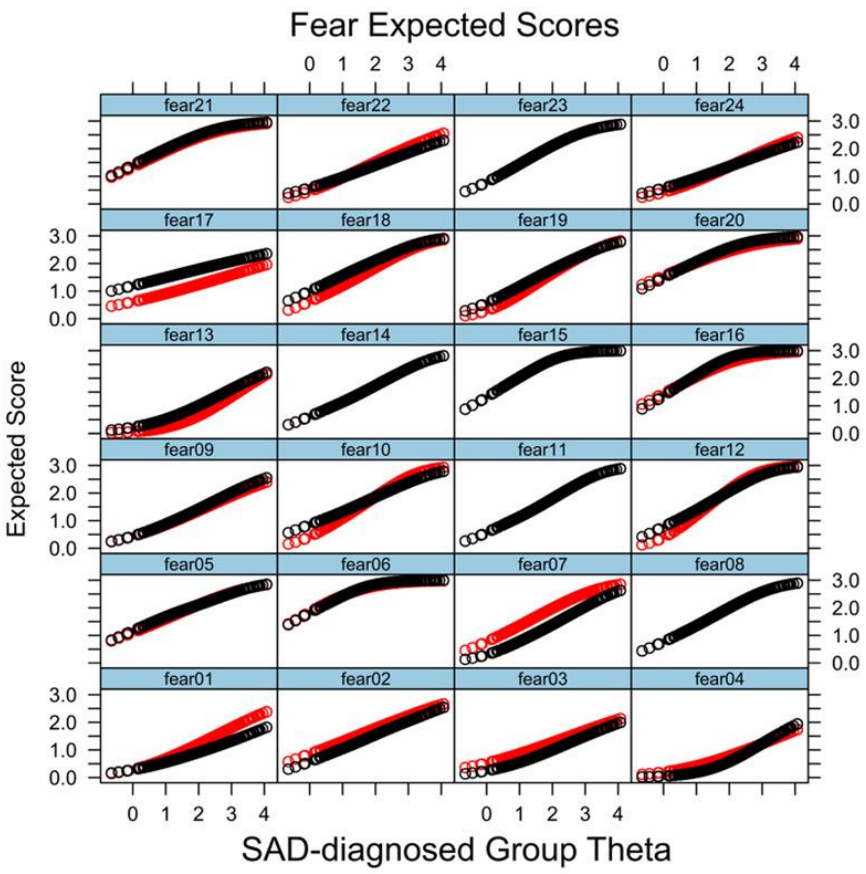

Fear scale
Community-dwelling participants
Fear scale
SAD-diagnosed participants
Avoidance scale
Community-dwelling participants
Avoidance scale
SAD-diagnosed participants
Fear scale
Female participants
Fear scale
Male participants
Avoidance scale
Female participants
Avoidance scale
Male participants
Observed Eigenvalues
Random Eigenvalues
MAP Statistic
Eigenvalue
MAP statistic
Factor

Figure 1

Item	ESSD
fear01	0.67 [0.53; 0.81]
fear02	0.53 [0.39; 0.67]
fear03	0.68 [0.54; 0.82]
fear04	0.38 [0.24; 0.52]
fear05	-0.05 [-0.18; 0.08]
fear06	-0.09 [-0.22; 0.04]
fear07	1.05 [0.91; 1.19]
fear08	Anchor
fear09	-0.13 [-0.26; 0.00]
fear10	-0.07 [-0.20; 0.06]
fear11	Anchor
fear12	-0.04 [-0.17; 0.09]
fear13	-0.66 [-0.80; -0.52]
fear14	Anchor
fear15	Anchor
fear16	-0.28 [-0.42; -0.14]
fear17	-1.95 [-2.11; -1.79]
fear18	-0.53 [-0.67; -0.39]
fear19	-0.28 [-0.42; -0.14]
fear20	-0.25 [-0.38; -0.11]
fear21	-0.20 [-0.33; -0.07]
fear22	0.20 [0.07; 0.33]
fear23	Anchor
fear24	-0.03 [-0.16; 0.10]



Item	ESSD
avoid01	-0.30 [-0.44; -0.16]
avoid02	Anchor
avoid03	-0.28 [-0.42; -0.14]
avoid04	-0.10 [-0.23; 0.03]
avoid05	-0.17 [-0.30; -0.04]
avoid06	0.10 [-0.03; 0.23]
avoid07	0.39 [0.25; 0.53]
avoid08	-0.53 [-0.67; -0.39]
avoid09	-0.83 [-0.97; -0.69]
avoid10	-0.24 [-0.37; -0.11]
avoid11	Anchor
avoid12	Anchor
avoid13	-1.02 [-1.16; -0.88]
avoid14	-0.23 [-0.36; -0.10]
avoid15	Anchor
avoid16	-0.19 [-0.32; -0.06]
avoid17	-1.34 [-1.49; -1.19]
avoid18	-0.14 [-0.27; -0.01]
avoid19	0.05 [-0.08; 0.18]
avoid20	-0.11 [-0.24; 0.02]
avoid21	-0.07 [-0.20; 0.06]
avoid22	0.35 [0.21; 0.49]
avoid23	Anchor
avoid24	-0.60 [-0.74; -0.46]

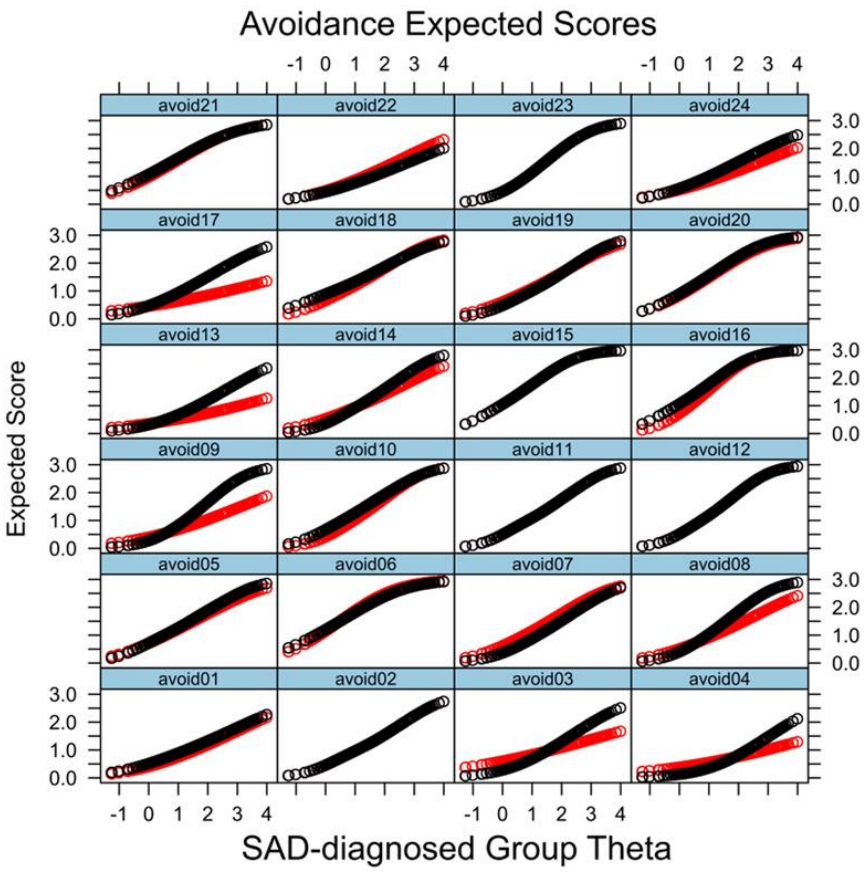


Figure 2

Supplementary Materials for the manuscript

Measurement Invariance and Psychometric Properties of the Liebowitz Social Anxiety Scale-Self Report Across Clinical and Non-clinical Groups

SM1 Italian Liebowitz Social Anxiety Scale (I-LSAS)

Fear subscale

Di seguito troverà una lista di situazioni quotidiane. Le chiediamo di valutare quanto ogni situazione la rende ansioso o spaventato. Se deve dare un punteggio ad una situazione di cui normalmente non ha esperienza, provi ad immaginare cosa succederebbe se dovesse affrontarla e quindi indichi con un punteggio quanta ansia proverebbe.
Risponda pensando a quanto ogni situazione Le ha dato disagio nell'**ultima settimana**.

	Nessuna ansia	Ansia lieve	Ansia moderata	Ansia grave
01. Telefonare in pubblico [Using a telephone in public]	☐0	☐1	☐2	☐3
02. Partecipare ad un'attività in un piccolo gruppo [Participating in a small group activity]	☐0	☐1	☐2	☐3
03. Mangiare di fronte ad altre persone [Eating in public]	☐0	☐1	☐2	☐3
04. Bere con altri in pubblico [Drinking with others]	☐0	☐1	☐2	☐3
05. Parlare con qualcuno che ha una posizione di autorità [Talking to someone in authority]	☐0	☐1	☐2	☐3
06. Recitare, esibirsi o parlare davanti ad un pubblico [Acting, performing, or speaking in front of an audience]	☐0	☐1	☐2	☐3
07. Andare a una festa [Going to a party]	☐0	☐1	☐2	☐3
08. Lavorare mentre si è osservati [Working while being observed]	☐0	☐1	☐2	☐3
09. Scrivere mentre si è osservati [Writing while being observed]	☐0	☐1	☐2	☐3
10. Telefonare a qualcuno che si conosce poco [Calling someone you don't know very well]	☐0	☐1	☐2	☐3
11. Parlare di persona con qualcuno che si conosce poco [Talking face to face with someone you don't know very well]	☐0	☐1	☐2	☐3
12. Incontrarsi con persone sconosciute [Meeting strangers]	☐0	☐1	☐2	☐3
13. Urinare in un bagno pubblico [Urinating in a public bathroom]	☐0	☐1	☐2	☐3
14. Entrare in una stanza dove ci sono già altre persone sedute [Entering a room when others are already seated]	☐0	☐1	☐2	☐3
15. Essere al centro dell'attenzione [Being the center of attention]	☐0	☐1	☐2	☐3
16. Prendere la parola in una riunione [Speaking up at a meeting]	☐0	☐1	☐2	☐3
17. Sottoporsi a un esame scritto [Taking a test of your ability, skill, or knowledge]	☐0	☐1	☐2	☐3
18. Esprimere disaccordo o disapprovazione a qualcuno che si conosce poco [Expressing disagreement or disapproval to someone you don't know very well]	☐0	☐1	☐2	☐3
19. Guardare negli occhi qualcuno che si conosce poco [Looking someone who you don't know very well straight in the eyes]	☐0	☐1	☐2	☐3
20. Tenere una relazione davanti a un gruppo di persone [Giving a prepared oral talk to a group]	☐0	☐1	☐2	☐3
21. Provarci con qualcuno per iniziare una relazione sentimentale/sessuale [Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship]	☐0	☐1	☐2	☐3
22. Restituire della merce in negozio [Returning goods to a store for a refund]	☐0	☐1	☐2	☐3
23. Dare una festa [Giving a party]	☐0	☐1	☐2	☐3
24. Resistere ad un venditore molto insistente [Resisting a high pressure sales person]	☐0	☐1	☐2	☐3

Avoidance subscale

Ora Le chiediamo di valutare quanto ha evitato o eviterebbe ogni situazione. Se deve dare un punteggio ad una situazione di cui normalmente non ha esperienza, provi ad immaginare cosa succederebbe se dovesse affrontarla e quindi indichi con un punteggio quanto la eviterebbe.
Risponda pensando a cosa ha fatto nell'**ultima settimana**.

	Mai	Qualche volta	Spesso	Sempre
01. Telefonare in pubblico [Using a telephone in public]	☐0	☐1	☐2	☐3
02. Partecipare ad un'attività in un piccolo gruppo [Participating in a small group activity]	☐0	☐1	☐2	☐3
03. Mangiare di fronte ad altre persone [Eating in public]	☐0	☐1	☐2	☐3
04. Bere con altri in pubblico [Drinking with others]	☐0	☐1	☐2	☐3
05. Parlare con qualcuno che ha una posizione di autorità [Talking to someone in authority]	☐0	☐1	☐2	☐3
06. Recitare, esibirsi o parlare davanti ad un pubblico [Acting, performing, or speaking in front of an audience]	☐0	☐1	☐2	☐3
07. Andare a una festa [Going to a party]	☐0	☐1	☐2	☐3
08. Lavorare mentre si è osservati [Working while being observed]	☐0	☐1	☐2	☐3
09. Scrivere mentre si è osservati [Writing while being observed]	☐0	☐1	☐2	☐3
10. Telefonare a qualcuno che si conosce poco [Calling someone you don't know very well]	☐0	☐1	☐2	☐3
11. Parlare di persona con qualcuno che si conosce poco [Talking face to face with someone you don't know very well]	☐0	☐1	☐2	☐3
12. Incontrarsi con persone sconosciute [Meeting strangers]	☐0	☐1	☐2	☐3
13. Urinare in un bagno pubblico [Urinating in a public bathroom]	☐0	☐1	☐2	☐3
14. Entrare in una stanza dove ci sono già altre persone sedute [Entering a room when others are already seated]	☐0	☐1	☐2	☐3
15. Essere al centro dell'attenzione [Being the center of attention]	☐0	☐1	☐2	☐3
16. Prendere la parola in una riunione [Speaking up at a meeting]	☐0	☐1	☐2	☐3
17. Sottoporsi a un esame scritto [Taking a test of your ability, skill, or knowledge]	☐0	☐1	☐2	☐3
18. Esprimere disaccordo o disapprovazione a qualcuno che si conosce poco [Expressing disagreement or disapproval to someone you don't know very well]	☐0	☐1	☐2	☐3
19. Guardare negli occhi qualcuno che si conosce poco [Looking someone who you don't know very well straight in the eyes]	☐0	☐1	☐2	☐3
20. Tenere una relazione davanti a un gruppo di persone [Giving a prepared oral talk to a group]	☐0	☐1	☐2	☐3
21. Provarci con qualcuno per iniziare una relazione sentimentale/sessuale [Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship]	☐0	☐1	☐2	☐3
22. Restituire della merce in negozio [Returning goods to a store for a refund]	☐0	☐1	☐2	☐3
23. Dare una festa [Giving a party]	☐0	☐1	☐2	☐3
24. Resistere ad un venditore molto insistente [Resisting a high pressure sales person]	☐0	☐1	☐2	☐3

SM2 Tests of validity and reliability of the Italian version of the Liebowitz Social Anxiety Scale - Self-report.

Rationale of the construct validity tests

To test the construct validity of the Italian Liebowitz Social Anxiety Scale - Self-report scales, we investigated the pattern of association of its scores with other measures of social anxiety (SA) (convergent validity) and with measures of depression, general anxiety, worry, and obsessive-compulsive symptomatology (discriminant validity) in participants diagnosed with Social Anxiety Disorder (SAD-diagnoses) and community-dwelling participants.

Social anxiety disorder (SAD) is known to be comorbid with depression and general anxiety (Lydiard, 2001; Ruscio et al., 2008) and with obsessive-compulsive disorder in either clinical samples or community-dwelling participants (Assunção et al., 2012; Baldwin et al., 2008; Carpita et al., 2020). There is less evidence about worry, i.e., the tendency to experience persistent, uncontrollable intrusive thoughts and images that can be upsetting, and focus on solving problems or exploring feared outcomes that may occur in the future (Borkovec et al., 1983). Worry is a defining characteristic of the general anxiety disorder (GAD) and there are aspects of SAD for which worry may be considered, at least in part, to be a maintaining factor (e.g., individuals with SAD engage in anticipatory anxiety before social interactions and post-event processing following social interactions, both of which could be considered as 'worry-like'; see Clark & Wells, 1995). GAD-like worry has been found in socially anxious adults (Starcevic et al.,2007), and Hearn et al. (2017) found that worry correlated with SAD symptoms and severity in youth.

First, we computed the Pearson correlation of LSAS-SR scores with the other measures in both groups. Then, we tested whether there were differences between the SAD-diagnosed and community-dwelling groups in the pattern of association of these scores with Steiger's (1980) test of the equality of correlation matrices, followed by pairwise comparisons (with Benjamini-Hochberg's (2000) adaptive false discovery rate *p*-value

adjustment procedure). We then tested whether the convergent correlations (i.e., correlations with SPS, SIAS, and SPIN) were statistically stronger than the discriminant correlations (i.e., correlations with BDI, BAI, PSWQ, and MOCQ) using the $Z_{contrast}$ test (Meng et al., 1992; Westen & Rosenthal, 2003). This was followed by contrasts between each convergent correlation with each discriminant correlation (again with Benjamini-Hochberg's (2000) adaptive false discovery rate *p*-value adjustment procedure).

Measures

All participants completed a battery comprising a demographic information schedule, the LSAS-SR (as described in the manuscript), and the following questionnaires.

Social Phobia Scale. The Italian version (Sica et al., 2007) of the Social Phobia Scale (SPS; Mattick & Clarke, 1998) was used for assessing the fear of being watched or judged in routine activities. The scale consists of 20 items (e.g., "*I become anxious if I have to write in front of other people*") rated on a 5-point scale (0 = "not at all"; 4 = "extremely") according to the anxiety evoked by each situation with higher scores indicating greater levels of anxiety. Descriptive statistics and Cronbach's alphas for the two groups of participants are reported in Table SM2.1.

Social Interaction Anxiety Scale. The Italian version (Sica et al., 2007) of the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) was used for evaluating the fear of social interactions. The scale consists of 19 items (e.g., "*I become tense if I have to talk about myself or my feelings*") rated on a 6-point scale (0 = "not at all"; 5 = "extremely") that evaluates the fear experienced in social interaction situations. Higher scores indicate greater levels of anxiety. Descriptive statistics and Cronbach's alphas for the two groups of participants are reported in Table SM2.1.

Social Phobia Inventory. The Italian version (Gori et al., 2013) of the Social Phobia Inventory (SPIN; Connor et al., 2000) was used to assess fear, avoidance, and physiological changes related to social anxiety. It's composed of 17 items rated on a 5-point scale (0 = "not at all", 4 = "always"). The original version of the scale was conceived to provide scores in three subscales (Fear, Avoidance, Physiological), while the study on the Italian version reported a three-factor structure with a somewhat different meaning (Fear, Avoidance, and Authority problems). Again, we could not find convincing evidence of the replicability of these measurement models in our data. Since we found that all the items had substantial loadings (≥|.30|) on a single factor in either sample, we decided to compute only a total score. Descriptive statistics and Cronbach's alphas of the total score for the two groups of participants are reported in Table SM2.1.

Beck Depression Inventory-II. The Italian version (Ghisi et al., 2006) of the Beck Depression Inventory-II (BDI-II; Beck et al., 1996) was used for assessing the presence and the severity of depression symptoms. BDI-II consists of 21 items rated on a 4 points Likert-type scale (e.g., 0 = "*I do not feel sad*", 3 = "*I am so sad or unhappy that I can't stand it*"). Higher scores reflect higher severity of depression symptoms. Although two-factor measurement models were reported for the BDI-II, convincing support for the one-factor model has been also provided (see, e.g., Huang & Chen 2014). Moreover, consistent with the results by Huang and Chen (2014), in this study the cognitive and somatic-affective subscale scores were highly correlated (.78 and .74 in the general population and patient sample, respectively), hence we considered only the total score. Descriptive statistics and Cronbach's alphas of the total score for the two groups of participants are reported in Table SM2.1.

Beck Anxiety Inventory. The Italian version (Sica et al., 2006) of the Beck Anxiety Inventory (BAI; Beck et al., 1988) was used to evaluate the intensity of general anxiety symptoms. The BAI consists of 21 items rated on a 4-point scale (0 = *not at all*; 3 =

severely). Descriptive statistics and Cronbach's alphas of the total score for the two groups of participants are reported in Table SM2.1.

Penn State Worry Questionnaire. The Italian version (Morani et al., 1999) of the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990) was used to measures the inclination to worry, independently from situations or moments, with excessive frequency and intensity. The PSWQ consists of 16 items (e.g., "*If I do not have enough time to do everything, I do not worry about it*") rated on a 5-point scale (1= *not at all typical of me*; 5 = *very typical of me*); higher scores reflect greater tendency to worry. Descriptive statistics and Cronbach's alphas of the total score for the two groups of participants are reported in Table SM2.1.

*Maudsley Obsessional Compulsive Questionnaire***.** The Italian version (Sanavio & Vidotto, 1985) of the Maudsley Obsessional Compulsive Questionnaire (MOCQ; Hodgson & Rachman, 1977) was used to assess the presence of obsessive-compulsive symptoms. MOCQ is composed of 21 true/false items. The original study found a four-factor structure and the study on the Italian version reported evidence of a three-factor structure, but the reliabilities of the scales suggested an inadequate measurement model, as some of them were as low as .33 using the original scoring or .45 using the revised Sanavio and Vidotto's scoring, in either group. We thus investigated the dimensionality of the MOCQ with the methods described in the manuscript. While parallel analysis suggested up to four factors, MAP suggested two (Figure SM2.1).

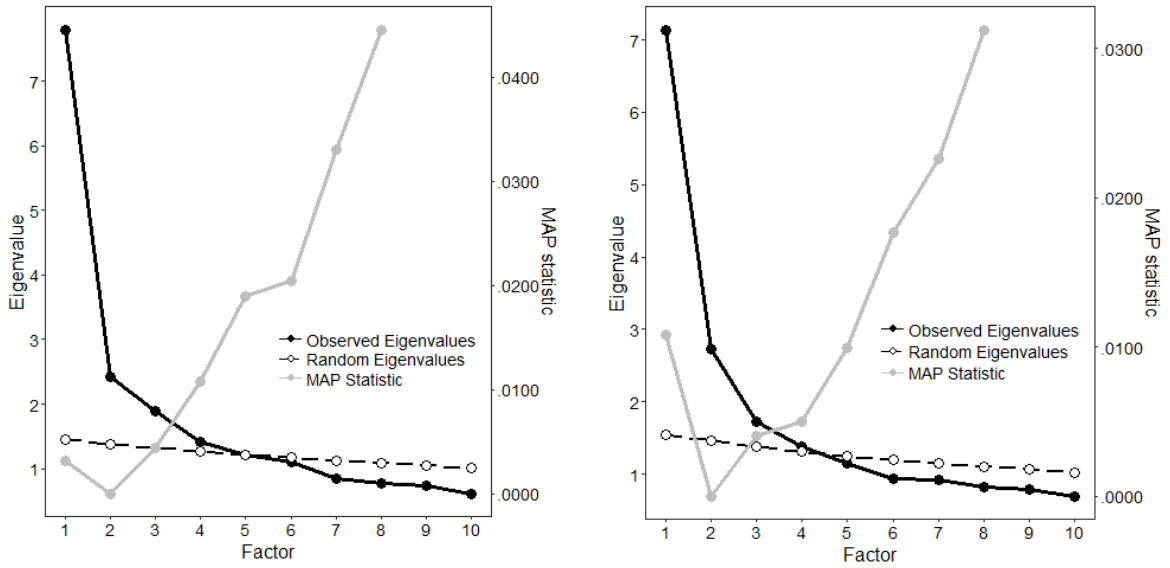


Figure SM2.1 Dimensionality analyses for the Maudsley Obsessional Compulsive Questionnaire in the community-dwelling (left) and in the SAD-diagnosed (right) samples. MAP = Minimum Average Partial correlation.

We then performed exploratory factor analyses on the tetrachoric correlation matrix using the MINRES extraction method and *oblimin* rotation (the default in the *fa* function in the *psych* package [Revelle, 2015] in *R*). While the three- and four-factor solutions did not provide a simple structure, the two-factor solution showed evidence of simple structure, with moderately correlated factors (Table SM2.1a and b).

Table SM2.1a Results of the exploratory factor analyses on the Maudsley Obsessional Compulsive Questionnaire in the community-dwelling sample (n = 356). Bolded items are larger than |.30|

Item	F1	F2	F1	F2	F3	F1	F2	F3	F4	Content
mocq01	-.04	**.63**	-.04	**.61**	.06	-.11	**.56**	.07	.15	I avoid using public telephones because of possible contamination.
mocq02	**.42**	.16	.06	.11	**.74**	.06	.09	**.90**	-.14	I frequently get nasty thoughts and have difficulty in getting rid of them.
mocq03	-.03	**.70**	-.16	**.69**	.27	**-.31**	**.60**	.25	**.31**	I don't worry unduly about contamination if I touch an animal
mocq04	**.63**	.23	**.62**	.25	.06	**.58**	.28	.13	.04	I frequently have to check things (e.g. gas or water taps, doors, etc.) several times.
mocq05	**.42**	.09	.08	.03	**.74**	.05	.00	**.82**	-.04	I find that almost every day I am upset by unpleasant thoughts that come into my mind against my will.
mocq06	**.66**	.04	**.44**	.02	**.45**	.20	-.06	**.42**	**.45**	I usually have serious doubts about the simple everyday things I do
mocq07	**.49**	.17	**.48**	.19	.03	.09	.06	-.16	**.97**	I tend to get behind in my work because I repeat things over and over again.
mocq08	.21	**.61**	.24	**.61**	-.02	.18	**.58**	.03	.14	I use only an average amount of soap.
mocq09	**.59**	.27	**.50**	.28	.20	**.36**	.25	.22	.26	I do not check letters over and over again before posting them.
mocq10	.03	**.36**	.18	**.40**	-.27	.30	**.47**	-.18	-.23	I am not excessively concerned about cleanliness.
mocq11	**.63**	.06	**.65**	.08	-.03	**.51**	.09	-.01	.26	One of my major problems is that I pay too much attention to detail.
mocq12	**.81**	.10	**.78**	.13	.07	**.72**	.17	.15	.07	My major problem is repeated checking.
mocq13	-.01	**.55**	.10	**.57**	-.19	.15	**.60**	-.11	-.10	I am not unduly concerned about germs and diseases.
mocq14	**.85**	-.21	**.85**	-.17	-.01	**.84**	-.10	.06	-.01	I do not tend to check things more than once.
mocq15	**.61**	-.19	**.68**	-.16	-.12	**.62**	-.11	-.07	.06	I do not stick to a very strict routine when doing ordinary things.
mocq16	-.10	**.61**	-.15	**.58**	.13	-.25	**.51**	.12	.22	My hands do not feel dirty after touching money.
mocq17	**.38**	.29	**.45**	**.32**	-.11	**.42**	**.34**	-.05	.05	I take rather a long time to complete my washing in the morning.
mocq18	.10	**.75**	.15	**.74**	-.04	.19	**.77**	.06	-.10	I do not use a great deal of antiseptics.
mocq19	**.90**	.01	**.81**	.04	.17	**.67**	.05	.22	.22	I spend a lot of time every day checking things over and over again.
mocq20	**.37**	.15	**.54**	.19	-.29	**.48**	.22	-.24	.10	Hanging and folding my clothes at night does not. take up a lot of time.
mocq21	**.77**	-.07	**.54**	-.08	**.45**	.29	-.16	**.41**	**.48**	Even when I do something very carefully I often feel that it is not quite right.
r with F1		.43		.36	.34		.26	.33	.36	
r with F2					.25			.25	.27	
r with F3									.22	

Table SM2.1b Results of the exploratory factor analyses on the Maudsley Obsessional Compulsive Questionnaire in the SAD-diagnosed sample (n = 257). Bolded items are larger than |.30|

Item	F1	F2	F1	F2	F3	F1	F2	F3	F4	Content
mocq01	-.14	**.83**	-.05	**.81**	-.11	.01	**.79**	-.15	.06	I avoid using public telephones because of possible contamination.
mocq02	**.57**	-.08	-.01	.02	**.96**	-.02	.02	**.97**	-.12	I frequently get nasty thoughts and have difficulty in getting rid of them.
mocq03	.12	**.47**	-.03	**.51**	.21	.16	**.68**	.00	**-.37**	I don't worry unduly about contamination if I touch an animal
mocq04	**.50**	.18	**.55**	.15	-.01	**.53**	.12	.01	.11	I frequently have to check things (e.g. gas or water taps, doors, etc.) several times.
mocq05	**.58**	.00	.18	.08	**.65**	.03	-.02	**.86**	.17	I find that almost every day I am upset by unpleasant thoughts that come into my mind against my will.
mocq06	**.62**	.00	**.44**	.03	.29	**.44**	.02	.28	.02	I usually have serious doubts about the simple everyday things I do
mocq07	**.46**	.14	**.38**	.14	.15	**.36**	.11	.16	.08	I tend to get behind in my work because I repeat things over and over again.
mocq08	.29	**.60**	.15	**.63**	.22	.17	**.59**	.22	.09	I use only an average amount of soap.
mocq09	**.47**	.08	**.50**	.06	.00	**.46**	.02	.04	.14	I do not check letters over and over again before posting them.
mocq10	-.10	**.64**	.03	**.60**	-.17	-.10	**.47**	.00	**.42**	I am not excessively concerned about cleanliness.
mocq11	**.42**	.19	**.42**	.18	.05	**.49**	.21	-.03	-.06	One of my major problems is that I pay too much attention to detail.
mocq12	**.74**	.06	**.85**	.00	-.06	**.87**	.00	-.10	.05	My major problem is repeated checking.
mocq13	.00	**.62**	-.04	**.63**	.07	.00	**.62**	.05	.04	I am not unduly concerned about germs and diseases.
mocq14	**.82**	.00	**.85**	-.03	.06	**.77**	-.10	.13	.21	I do not tend to check things more than once.
mocq15	.17	**.45**	**.33**	**.40**	-.17	.14	.22	.04	**.53**	I do not stick to a very strict routine when doing ordinary things.
mocq16	-.13	**.69**	-.23	**.73**	.12	-.14	**.77**	.05	-.08	My hands do not feel dirty after touching money.
mocq17	.23	**.48**	.27	**.47**	-.01	.22	**.40**	.05	.20	I take rather a long time to complete my washing in the morning.
mocq18	.13	**.81**	.15	**.80**	.01	.09	**.69**	.11	**.31**	I do not use a great deal of antiseptics.
mocq19	**.78**	.02	**.78**	-.01	.07	**.88**	.03	-.04	-.10	I spend a lot of time every day checking things over and over again.
mocq20	.15	**.35**	**.40**	.28	-.29	.18	.07	-.05	**.65**	Hanging and folding my clothes at night does not. take up a lot of time.
mocq21	**.75**	-.11	**.47**	-.05	**.43**	**.59**	.03	.29	-.26	Even when I do something very carefully I often feel that it is not quite right.
r with F1		.40		.37	.34		.33	.39	.19	
r with F2					.16			.19	.21	
r with F3									.02	

Note: r with F*: Pearson correlation with F*.

Beyond the slight differences in the loading patterns between the two groups, the content of the items suggested that the two factors tapped into checking and cleaning concerns (excluding items 15, 17, and 20 that did not substantively load on the same factor in either sample). However, the cleaning factor still had inadequate reliability (.56) in the community-dwelling sample. We thus checked the properties of a single-factor solution, and found that it accounted for 34% and 31% of variance in the community-dwelling and SAD-diagnosed samples, respectively, and that the unidimensionality indices met the criteria reported in the manuscript (Community-dwelling sample: Cronbach's alpha: .80 [.77; .83]; omega = .92 [.90; .95]; Factor score determinacy [FSD] = .98 [.95; .98]; *H* index = .95 [.91; .96]; SAD-diagnosed sample: Cronbach's alpha: .80 [.77; .84]; omega = .91 [.89; .94]; Factor score determinacy [FSD] = .96 [.95; .97]; *H* index = .93 [.90; .94]). The factor loadings on the single factor were all equal to or larger than .30 (Table SM2.2). Taken together, these results lead us to consider a single total score for the MOCQ in subsequent analyses.

Table SM2.2 Results of the single-factor exploratory factor analysis on the Maudsley Obsessional Compulsive Questionnaire (factor loadings and their 95% bootstrapped confidence interval).

Item	Community-dwelling sample (*n* = 356)	SAD-diagnosed sample (*n* = 257)
mocq01	.40 [.07; .68]	.51 [.23; .75]
mocq02	.51 [.30; .68]	.43 [.24; .63]
mocq03	.45 [.19; .67]	.47 [.27; .64]
mocq04	.77 [.66; .86]	.58 [.43; .72]
mocq05	.47 [.23; .67]	.51 [.35; .67]
mocq06	.65 [.49; .79]	.54 [.38; .70]
mocq07	.59 [.29; .80]	.51 [.33; .67]
mocq08	.61 [.35; .83]	.73 [.57; .85]
mocq09	.76 [.64; .85]	.48 [.32; .62]
mocq10	.29 [.10; .46]	.41 [.23; .56]
mocq11	.63 [.50; .75]	.53 [.38; .67]
mocq12	.84 [.75; .92]	.68 [.55; .80]
mocq13	.37 [.17; .54]	.49 [.32; .64]
mocq14	.65 [.54; .75]	.71 [.58; .82]
mocq15	.43 [.25; .60]	.50 [.34; .64]
mocq16	.32 [.13; .49]	.42 [.23; .59]
mocq17	.56 [.30; .77]	.59 [.42; .75]
mocq18	.60 [.26; .84]	.74 [.50; .91]
mocq19	.85 [.73; .94]	.68 [.54; .80]
mocq20	.46 [.24; .64]	.41 [.14; .64]
mocq21	.67 [.56; .77]	.56 [.39; .73]

We then tested the construct validity of the LSAS-SR scales in both groups. The correlation matrix of the LSAS-SR scores with the scores on the other measures is reported in Table SM2.3. The correlation matrices of the two groups were very similar, and Steiger's (1980) test of the equality of the correlation matrices was not significant ($\chi^2(36) = 44.03$, $p = .168$, $r = .26$ [.19; .34]). Further, upon performing pairwise comparisons between the coefficients, we found no evidence of substantial differences (Table SM2.4).

Although two coefficients (SIAS with SPIN, and SPS with PSWQ) differed statistically, their difference was no longer significant after the correction made by Benjamini-Hochberg (2000) adaptive false discovery rate controlling procedure, and the effect size did not exceed .11. These results suggest an equivalence of the pattern of association between the measures in the two groups.

We then tested whether the convergent correlations (i.e., those with SPS, SIAS, and SPIN) were statistically stronger than the discriminant correlations (i.e., those with BDI, BAI, PSWQ, and MOCQ). The $Z_{contrast}$ test (Meng et al., 1992; Westen & Rosenthal, 2003) was significant for both scales in both samples (Community-dwelling: Fear: $Z = 13.55$, $p < .001$, $r = .71$ [.66; .75]; Avoidance: $Z = 6.63$, $p < .001$, $r = .35$ [.25; .43]; SAD-diagnosed: Fear: $Z = 13.63$, $p < .001$, $r = .85$ [.82; .87]; Avoidance: $Z = 9.96$, $p < .001$, $r = .62$ [.54; .68]). When we performed the post-hoc pairwise comparisons of each convergent correlation with each discriminant correlation in both scales and groups, all contrasts were significant at $p < .05$ with and without adjusting the *p*-values for the false discovery rate (see Table SM2.5a and SM2.5b for details). In the SAD-diagnosed group, the effect sizes *r* of the contrasts ranged from .28 to .63 and .19 to .50 in the Fear and Avoidance scales, respectively. In the community-dwelling group, the effect sizes *r* of the contrasts ranged from .23 to .59 and .12 to .32 in the Fear and Avoidance scales, respectively. Taken together, these results suggest an adequate construct validity of the LSAS-SR scales.

Table SM2.3 Correlation matrix, reliability, and descriptive statistics of the scores on the measures of social anxiety, depression, anxiety, worry, and obsessive-compulsive symptoms in the patient (n = 257, lower triangle, rows) and general population (n = 301, upper triangle, columns) samples.

Scale	1	2	3	4	5	6	7	8	9	M	SD
1. LSAS-Fear	.90 (.88; .91)	.68	.78	.70	.71	.54	.48	.44	-.36	41.37	11.96
	.91 (.89; .92)	(.62; .73)	(.73; .82)	(.63; .75)	(.65; .77)	(.45; .62)	(.39; .57)	(.35; .53)	(-.45; -.25)		
2. LSAS-Avoidance	.68	.90 (.88; .92)	.55	.53	.59	.42	.42	.44	-.33	34.77	13.28
	(.61; .74)	.90 (.89; .92)	(.47; .62)	(.44; .61)	(.51; .66)	(.33; .51)	(.33; .51)	(.34; .52)	(-.43; -.22)		
3. SPS	.79	.63	.92 (90; .93)	.76	.80	.58	.53	.57	-.44	43.23	15.15
	(.74; .83)	(.55; .70)	.92 (.91; .93)	(.70; .80)	(.75; .84)	(.50; .65)	(.44; .61)	(.48; .64)	(-.53; -.35)		
4. SIAS	.73	.60	.69	.90 (.88; .92)	.82	.60	.48	.51	-.33	65.42	17.81
	(.67; .78)	(.51; .67)	(.62; .75)	.91 (.90; .93)	(.77; .85)	(.52; .66)	(.39; .56)	(.42; .59)	(-.42; -.22)		
5. SPIN	.77	.66	.79	.76	.87 (.85; .90)	.52	.50	.55	-.32	44.29	11.59
	(.72; .82)	(.58; .72)	(.74; .83)	(.70; .81)	.91 (.90; .93)	(.43; .60)	(.40; .58)	(.47; .63)	(-.42; -.22)		
6. BDI	.53	.44	.55	.53	.56	.92 (.91; .93)	.57	.50	-.38	25.85	13.11
	(.44; .61)	(.33; .53)	(.46; .63)	(.44; .61)	(.47; .64)	.92 (.91; .93)	(.49; .64)	(.41; .58)	(-.47; -.28)		
7. BAI	.45	.35	.57	.40	.54	.55	.89 (.87; .91)	.45	-.36	26.21	11.71
	(.35; .55)	(.24; .45)	(.49; .65)	(.29; .50)	(.45; .62)	(.45; .63)	.91 (.90; .93)	(.35; .53)	(-.46; -.26)		
8. PSWQ	.41	.37	.40	.47	.52	.44	.41	.88 (.86; .90)	-.46	63.60	11.48
	(.31; .51)	(.25; .47)	(.29; .50)	(.37; .56)	(.43; .61)	(.34; .54)	(.31; .51)	.91 (.90; .92)	(-.55; -.37)		
9. MOCQ	-.31	-.21	-.31	-.32	-.35	-.41	-.39	-.34	.80 (.77; .84)	13.74	4.09
	(-.42; -.19)	(-.33; -.09)	(-.42; -.20)	(-.43; -.21)	(-.45; -.24)	(-.51; -.31)	(-.49; -.29)	(-.45; -.23)	.80 (.77; .83)		
M	19.94	16.01	14.92	22.70	14.29	7.85	9.46	43.71	16.82		
SD	10.54	10.74	11.96	15.23	11.26	8.20	8.90	12.74	3.59		

Note: all correlations significant at $p < .001$; bracketed values on the main diagonal are Cronbach's alphas for the clinical (upper) and general population (lower) sample with their 95% confidence intervals; bracketed values in the off-diagonal cells are 95% confidence interval for the correlation coefficients; LSAS: Liebowitz Social Anxiety Scale; SPS: Social Phobia Scale; SIAS: Social Interaction Anxiety scale; SPIN: Social Phobia Inventory; BDI: Beck Depression Inventory; BAI: Beck Anxiety Inventory; PSWQ: Penn State Worry Questionnaire; MOCQ: Maudsley Obsessional-Compulsive Questionnaire. M: mean; SD: standard deviation

Table SM2.4 Pairwise comparison of single coefficients of the matrices reported in Table SM2.1

Variable 1	Variable 2	SAD-diagnosed (*n* = 257)	Community-dwelling (*n* = 356)	Z-value	p	adj-p	r
LSAS-F	LSAS-A	.68	.68	0.00	1.000	1.000	.01 [-.07; .08]
LSAS-F	SPS	.79	.79	0.32	.751	.835	.00 [-.08; .08]
LSAS-A	SPS	.63	.57	0.75	.455	.835	.05 [-.03; .13]
LSAS-F	SIAS	.73	.71	1.62	.106	.592	.02 [-.05; .10]
LSAS-A	SIAS	.60	.55	-0.17	.865	.916	.03 [-.04; .11]
SPS	SIAS	.69	.77	-0.47	.642	.835	.08 [-.00; .15]
LSAS-F	SPIN	.77	.73	-0.45	.656	.835	.05 [-.03; .13]
LSAS-A	SPIN	.66	.61	0.68	.494	.835	.04 [-.04; .12]
SPS	SPIN	.79	.81	1.50	.135	.607	.02 [-.06; .10]
SIAS	SPIN	.76	.82	1.25	.211	.682	.08 [.00; .16]
LSAS-F	BDI	.53	.55	1.40	.162	.647	.01 [-.07; .09]
LSAS-A	BDI	.44	.44	0.30	.766	.835	.00 [-.08; .08]
SPS	BDI	.55	.59	-1.00	.317	.816	.03 [-.05; .11]
SIAS	BDI	.53	.61	-1.02	.308	.816	.05 [-.03; .13]
SPIN	BDI	.56	.53	1.58	.115	.592	.02 [-.06; .10]
LSAS-F	BAI	.45	.50	-1.80	.072	.589	.03 [-.05; .11]
LSAS-A	BAI	.35	.44	-0.33	.741	.835	.05 [-.03; .13]
SPS	BAI	.58	.54	-0.54	.592	.835	.02 [-.06; .10]
SIAS	BAI	.40	.49	0.70	.486	.835	.06 [-.02; .14]
SPIN	BAI	.54	.51	-2.72	.007	.234	.02 [-.06; .10]
BDI	BAI	.55	.58	1.84	.065	.589	.02 [-.06; .10]
LSAS-F	PSWQ	.41	.46	-1.95	.051	.589	.03 [-.05; .11]
LSAS-A	PSWQ	.37	.46	-1.25	.211	.682	.05 [-.03; .13]
SPS	PSWQ	.40	.57	-1.21	.227	.682	.11 [.03; .19]
SIAS	PSWQ	.47	.51	-0.64	.522	.835	.03 [-.05; .10]
SPIN	PSWQ	.52	.56	0.14	.892	.917	.02 [-.05; .10]
BDI	PSWQ	.44	.51	0.69	.492	.835	.04 [-.04; .12]
BAI	PSWQ	.41	.45	0.67	.505	.835	.02 [-.05; .10]
LSAS-F	MOCQ	-.31	-.37	-0.51	.609	.835	.03 [-.05; .11]
LSAS-A	MOCQ	-.21	-.34	-0.41	.681	.835	.07 [-.01; .15]
SPS	MOCQ	-.31	-.45	-0.35	.723	.835	.08 [.00; .16]
SIAS	MOCQ	-.32	-.34	-0.94	.349	.835	.01 [-.07; .09]
SPIN	MOCQ	-.35	-.33	-0.43	.666	.835	.01 [-.07; .09]
BDI	MOCQ	-.41	-.39	-0.60	.551	.835	.02 [-.06; .10]
BAI	MOCQ	-.39	-.37	-0.42	.671	.835	.01 [-.07; .09]
PSWQ	MOCQ	-.35	-.48	1.74	.082	.589	.08 [-.00; .16]

Note: Patients: *n* = 257; General population: *n* = 356; p = *p*-value; adj-p: adjusted *p*-value following the Benjamini-Hochberg (2000)'s adaptive false discovery rate controlling procedure. LSAS: Liebowitz Social Anxiety Scale; LSAS-F: LSAS-Fear; LSAS-A: LSAS-Avoidance; SPS: Social Phobia Scale; SIAS: Social Interaction Anxiety scale; SPIN: Social Phobia Inventory; BDI: Beck Depression Inventory; BAI: Beck Anxiety Inventory; PSWQ: Penn State Worry Questionnaire; MOCQ: Maudsley Obsessional-Compulsive Questionnaire.

Table SM2.5a Pairwise contrasts of convergent (rows) vs discriminant (columns) correlations for the Liebovitz Social Anxiety Scales – SAD-diagnosed sample (n = 257)

Scale	Statistic	Fear				Avoidance			
		BDI	BAI	PSWQ	MOCQ	BDI	BAI	PSWQ	MOCQ
		Patients							
SPS	Z	6.47	7.89	8.55	10.09	3.83	5.28	5.06	7.41
	p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	adj-p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	r	.40 [.30; .49]	.49 [.40; .57]	.53 [.44; .61]	.63 [.56; .69]	.24 [.12; .35]	.33 [.22; .43]	.32 [.20; .42]	.46 [.36; .55]
SIAS	Z	4.55	5.97	6.63	8.18	3.10	4.55	4.33	6.68
	p	<.001	<.001	<.001	<.001	.002	<.001	<.001	<.001
	adj-p	<.001	<.001	<.001	<.001	.002	<.001	<.001	<.001
	r	.28 [.17; .39]	.37 [.26; .47]	.41 [.31; .50]	.51 [.42; .59]	.19 [.07; .31]	.28 [.17; .39]	.27 [.15; .38]	.42 [.31; .51]
SPIN	Z	5.78	7.20	7.86	9.41	4.51	5.96	5.73	8.08
	p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	adj-p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	r	.36 [.25; .46]	.45 [.35; .53]	.49 [.40; .57]	.59 [.51; .65]	.28 [.17; .39]	.37 [.26; .47]	.36 [.25; .45]	.50 [.41; .58]

Note: Z: z-value; p = p-value; adj-p: adjusted p-value following the Benjamini-Hochberg (2000)'s adaptive false discovery rate controlling procedure. SPS: Social Phobia Scale; SIAS: Social Interaction Anxiety scale; SPIN: Social Phobia Inventory; BDI: Beck Depression Inventory; BAI: Beck Anxiety Inventory; PSWQ: Penn State Worry Questionnaire; MOCQ: Maudsley Obsessional-Compulsive Questionnaire.

Table SM2.5b Pairwise contrasts of convergent (rows) vs discriminant (columns) correlations for the Liebovitz Social Anxiety Scales – community-dwelling sample (*n* = 356)

Scale	Statistic	Fear				Avoidance			
		BDI	BAI	PSWQ	MOCQ	BDI	BAI	PSWQ	MOCQ
		General population							
SPS	Z	7.34	8.56	9.44	11.23	2.89	2.91	2.67	4.97
	p	<.001	<.001	<.001	<.001	.004	.004	.008	<.001
	adj-p	<.001	<.001	<.001	<.001	.005	.005	.009	<.001
	r	.39 [.30; .47]	.45 [.37; .53]	.50 [.42; .57]	.59 [.53; .65]	.15 [.05; .25]	.15 [.05; .25]	.14 [.04; .24]	.26 [.16; .35]
SIAS	Z	4.28	5.50	6.38	8.16	2.44	2.46	2.22	4.52
	p	<.001	<.001	<.001	<.001	.015	.014	.027	<.001
	adj-p	<.001	<.001	<.001	<.001	.015	.015	.027	<.001
	r	.23 [.13; .32]	.29 [.19; .38]	.34 [.24; .42]	.43 [.35; .51]	.13 [.03; .23]	.13 [.03; .23]	.12 [.01; .22]	.24 [.14; .33]
SPIN	Z	4.93	6.15	7.03	8.82	4.01	4.03	3.79	6.09
	p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	adj-p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	r	.26 [.16; .35]	.33 [.23; .41]	.37 [.28; .45]	.47 [.39; .54]	.21 [.11; .31]	.21 [.11; .31]	.20 [.10; .30]	.32 [.23; .41]

Note: Z: z-value; p = *p*-value; adj-p: adjusted *p*-value following the Benjamini-Hochberg (2000)'s adaptive false discovery rate controlling procedure. SPS: Social Phobia Scale; SIAS: Social Interaction Anxiety scale; SPIN: Social Phobia Inventory; BDI: Beck Depression Inventory; BAI: Beck Anxiety Inventory; PSWQ: Penn State Worry Questionnaire; MOCQ: Maudsley Obsessional-Compulsive Questionnaire.

Finally, we tested the test-retest reliability of the LSAS-SR scales in a subsample of 55 community-dwelling participants who accepted to complete the LSAS-SR twice at a 4-week interval. The findings indicated that the LSAS Fear scores were consistent over the four-week interval (Intraclass Correlation Coefficient [ICC, two-way random effects model with a consistency definition (McGraw & Wong, 1996)] = .88 [.81; .93]; Time 1 α = .88 [.84; .93], Time 2 α = .89 [.85; .93]). Cronbach's alphas did not differ ($t(53) = 0.42$, $p = .677$, d = 0.11 [-0.26; 0.49]), and that the scores did not significantly differ (Time 1: $M = 25.18$, SD = 10.62; Time 2: $M = 24.44$, $SD = 10.49$; $t(54) = 1.08$, $p = .282$, $d = 0.15$ [-0.12; 0.41], Figure SM2.2). A similar pattern of results was found for the Avoidance subscale. The ICC was .80 [.68; .88] (Time 1 α = .91 [.88; .95], Time 2 α = .90 [.86; .94]). Cronbach's alphas did not differ ($t(53) = 1.03$, $p = .309$, $d = 0.28$ [-0.09; 0.66]), and neither did the scores (Time 1: M = 21.24, SD = 11.97; Time 2: $M = 21.51$, $SD = 11.04$; $t(54) = 0.28$, $p = .782$, $d = 0.04$ [-0.23; 0.30]). Taken together, these results suggested that in non-diagnosed participants the LSAS-SR scores and their reliability tends to remain stable in in 4-week interval.

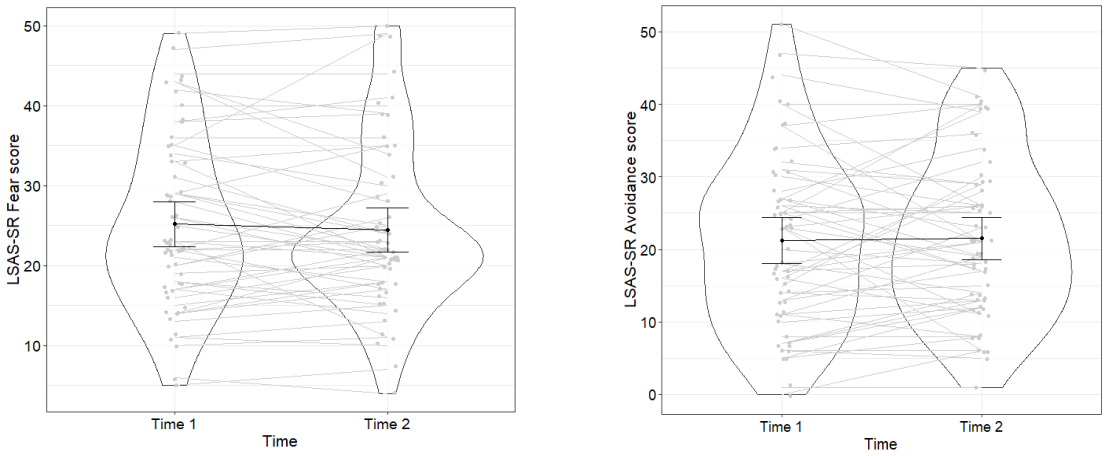


Figure SM2.2 Violin- and error-bar plots for the test of temporal stability of the Liebowitz Social Anxiety Scale- Self-Report (LSAS-SR) scales (Left: Fear scale; Right: Avoidance Scale) in a sample of 55 community dwelling-participants. Grey dots and lines represent individual data, black dots and lines represent sample means.

SM3 Frequency distribution of the Liebowitz Social Anxiety Scale-Self-Report item scores in this study

Table SM3.1 Proportions of item scores for the Fear scale in community-dwelling and SAD-diagnosed participants

		Community-dwelling (*n* = 356)					SAD-diagnosed (*n* = 257)				
Item	Text	0	1	2	3	M	0	1	2	3	M
fear01	Using a telephone in public	.72	.24	.03	.01	.00	.28	.37	.24	.11	.00
fear02	Participating in a small group activity	.49	.40	.10	.01	.00	.05	.32	.46	.16	.00
fear03	Eating in public	.77	.16	.07	.01	.00	.30	.30	.25	.15	.00
fear04	Drinking with others	.92	.05	.02	.00	.01	.47	.33	.15	.05	.00
fear05	Talking to someone in authority	.22	.43	.31	.04	.01	.04	.23	.39	.34	.00
fear06	Acting, performing, or speaking in front of an audience	.08	.28	.42	.22	.00	.02	.06	.16	.76	.00
fear07	Going to a party	.67	.26	.06	.01	.00	.08	.26	.35	.31	.00
fear08	Working while being observed	.40	.43	.14	.03	.00	.07	.24	.37	.32	.00
fear09	Writing while being observed	.62	.29	.06	.03	.00	.24	.34	.28	.14	.00
fear10	Calling someone you don't know very well	.32	.50	.15	.03	.00	.11	.30	.33	.26	.00
fear11	Talking face to face with someone you don't know very well	.47	.43	.10	.01	.00	.11	.32	.34	.23	.00
fear12	Meeting strangers	.38	.44	.15	.03	.01	.11	.21	.32	.36	.00
fear13	Urinating in a public bathroom	.81	.12	.05	.02	.00	.58	.23	.12	.07	.00
fear14	Entering a room when others are already seated	.45	.44	.09	.02	.00	.11	.33	.37	.19	.00
fear15	Being the centre of attention	.18	.42	.28	.12	.00	.02	.07	.32	.60	.00
fear16	Speaking up at a meeting	.18	.39	.31	.12	.00	.02	.11	.26	.60	.00
fear17	Taking a test of your ability, skill, or knowledge	.24	.40	.27	.09	.00	.27	.37	.24	.12	.00
fear18	Expressing disagreement or disapproval to someone you don't know very well	.28	.45	.23	.04	.00	.08	.31	.37	.24	.00
fear19	Looking someone who you don't know very well straight in the eyes	.53	.32	.13	.02	.00	.18	.28	.36	.18	.00
fear20	Giving a prepared oral talk to a group	.16	.34	.35	.14	.00	.03	.13	.29	.55	.00
fear21	Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship	.18	.35	.33	.13	.00	.05	.14	.26	.54	.00
fear22	Returning goods to a store for a refund	.54	.32	.11	.03	.00	.17	.33	.37	.13	.00
fear23	Giving a party	.46	.31	.17	.06	.00	.10	.18	.30	.41	.00
fear24	Resisting a high pressure sales person	.55	.31	.11	.03	.00	.28	.30	.28	.15	.00

Table SM3.2 Proportions of item scores for the Avoidance scale in community-dwelling and SAD-diagnosed participants

		Community-dwelling (*n* = 356)					SAD-diagnosed (*n* = 257)				
Item	Text	0	1	2	3	M	0	1	2	3	M
avoid01	Using a telephone in public	.55	.35	.07	.02	.00	.32	.40	.18	.11	.00
avoid02	Participating in a small group activity	.56	.37	.06	.01	.00	.14	.44	.29	.14	.00
avoid03	Eating in public	.74	.17	.06	.03	.00	.37	.36	.16	.11	.00
avoid04	Drinking with others	.86	.09	.03	.01	.00	.54	.25	.14	.06	.00
avoid05	Talking to someone in authority	.40	.44	.13	.02	.00	.14	.32	.34	.20	.00
avoid06	Acting, performing, or speaking in front of an audience	.28	.37	.21	.14	.01	.11	.11	.23	.55	.00
avoid07	Going to a party	.64	.27	.07	.02	.00	.16	.33	.26	.25	.00
avoid08	Working while being observed	.61	.27	.08	.04	.00	.23	.40	.22	.14	.00
avoid09	Writing while being observed	.71	.20	.05	.04	.00	.40	.35	.15	.10	.00
avoid10	Calling someone you don't know very well	.44	.39	.14	.03	.00	.18	.32	.27	.23	.00
avoid11	Talking face to face with someone you don't know very well	.50	.40	.08	.02	.01	.15	.35	.32	.18	.00
avoid12	Meeting strangers	.51	.35	.11	.03	.00	.16	.30	.25	.30	.00
avoid13	Urinating in a public bathroom	.73	.18	.05	.04	.00	.57	.26	.08	.09	.00
avoid14	Entering a room when others are already seated	.62	.30	.07	.01	.00	.21	.44	.22	.13	.00
avoid15	Being the centre of attention	.28	.42	.22	.08	.00	.07	.17	.26	.50	.00
avoid16	Speaking up at a meeting	.28	.43	.19	.10	.00	.11	.20	.22	.47	.00
avoid17	Taking a test of your ability, skill, or knowledge	.60	.27	.08	.04	.00	.48	.32	.11	.08	.00
avoid18	Expressing disagreement or disapproval to someone you don't know very well	.30	.52	.14	.04	.00	.12	.37	.29	.22	.00
avoid19	Looking someone who you don't know very well straight in the eyes	.55	.36	.07	.02	.00	.16	.39	.25	.20	.00
avoid20	Giving a prepared oral talk to a group	.37	.38	.17	.08	.00	.13	.22	.25	.40	.00
avoid21	Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship	.35	.33	.17	.15	.00	.13	.20	.23	.44	.00
avoid22	Returning goods to a store for a refund	.64	.25	.08	.03	.00	.28	.37	.20	.15	.00
avoid23	Giving a party	.64	.21	.09	.06	.00	.21	.19	.24	.37	.00
avoid24	Resisting a high pressure sales person	.58	.27	.10	.06	.00	.35	.33	.19	.13	.00

Table SM3.3 Proportions of item scores for the Fear scale in female and male participants

		Females (n = 390)					Males (n = 221)				
Item	Text	0	1	2	3	M	0	1	2	3	M
fear01	Using a telephone in public	.52	.29	.13	.06	.00	.56	.29	.10	.05	.00
fear02	Participating in a small group activity	.27	.41	.25	.08	.00	.37	.30	.26	.07	.00
fear03	Eating in public	.55	.21	.15	.09	.00	.60	.24	.14	.02	.00
fear04	Drinking with others	.71	.18	.09	.03	.01	.77	.16	.06	.01	.00
fear05	Talking to someone in authority	.10	.32	.39	.19	.01	.23	.37	.28	.13	.00
fear06	Acting, performing, or speaking in front of an audience	.05	.16	.30	.49	.00	.06	.24	.33	.37	.00
fear07	Going to a party	.42	.26	.17	.15	.00	.43	.26	.20	.11	.00
fear08	Working while being observed	.24	.36	.22	.19	.00	.30	.35	.27	.09	.00
fear09	Writing while being observed	.46	.30	.16	.08	.00	.45	.34	.15	.07	.00
fear10	Calling someone you don't know very well	.19	.43	.24	.14	.00	.29	.40	.20	.10	.00
fear11	Talking face to face with someone you don't know very well	.28	.39	.22	.11	.00	.37	.37	.16	.10	.00
fear12	Meeting strangers	.24	.36	.22	.18	.01	.31	.32	.21	.16	.00
fear13	Urinating in a public bathroom	.71	.17	.09	.03	.00	.70	.17	.07	.06	.00
fear14	Entering a room when others are already seated	.28	.41	.21	.10	.00	.34	.37	.20	.09	.00
fear15	Being the centre of attention	.08	.27	.33	.32	.00	.17	.28	.23	.32	.00
fear16	Speaking up at a meeting	.09	.27	.29	.35	.00	.17	.27	.29	.28	.00
fear17	Taking a test of your ability, skill, or knowledge	.22	.38	.27	.13	.00	.31	.39	.24	.06	.00
fear18	Expressing disagreement or disapproval to someone you don't know very well	.18	.38	.31	.13	.00	.21	.42	.26	.11	.00
fear19	Looking someone who you don't know very well straight in the eyes	.40	.29	.22	.09	.00	.35	.33	.24	.08	.00
fear20	Giving a prepared oral talk to a group	.09	.20	.37	.34	.00	.14	.33	.25	.28	.00
fear21	Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship	.12	.27	.30	.31	.00	.14	.24	.32	.30	.00
fear22	Returning goods to a store for a refund	.39	.31	.22	.07	.00	.37	.35	.22	.06	.00
fear23	Giving a party	.30	.26	.23	.22	.00	.32	.25	.24	.19	.00
fear24	Resisting a high pressure sales person	.44	.30	.19	.07	.00	.44	.32	.15	.09	.00

Table SM3.4 Proportions of item scores for the Avoidance scale in female and male participants

		Females (n = 390)					Males (n = 221)				
Item	Text	0	1	2	3	M	0	1	2	3	M
avoid01	Using a telephone in public	.44	.38	.12	.07	.00	.47	.37	.12	.05	.00
avoid02	Participating in a small group activity	.35	.43	.17	.06	.00	.45	.34	.13	.09	.00
avoid03	Eating in public	.57	.25	.10	.08	.00	.61	.25	.10	.05	.00
avoid04	Drinking with others	.72	.16	.07	.04	.00	.75	.15	.09	.02	.00
avoid05	Talking to someone in authority	.24	.43	.23	.11	.00	.38	.34	.19	.08	.00
avoid06	Acting, performing, or speaking in front of an audience	.18	.26	.23	.34	.00	.26	.26	.21	.27	.01
avoid07	Going to a party	.42	.30	.15	.13	.00	.46	.29	.16	.09	.00
avoid08	Working while being observed	.44	.32	.15	.09	.00	.47	.33	.14	.07	.00
avoid09	Writing while being observed	.58	.26	.09	.07	.00	.57	.28	.10	.05	.00
avoid10	Calling someone you don't know very well	.30	.36	.21	.13	.00	.38	.37	.17	.09	.00
avoid11	Talking face to face with someone you don't know very well	.34	.40	.17	.09	.00	.37	.35	.20	.08	.00
avoid12	Meeting strangers	.33	.35	.16	.16	.00	.42	.29	.18	.11	.00
avoid13	Urinating in a public bathroom	.65	.21	.08	.06	.00	.69	.23	.04	.05	.00
avoid14	Entering a room when others are already seated	.42	.38	.14	.06	.00	.48	.33	.12	.07	.00
avoid15	Being the centre of attention	.15	.33	.24	.28	.00	.26	.29	.23	.23	.00
avoid16	Speaking up at a meeting	.17	.32	.23	.27	.00	.26	.35	.16	.23	.00
avoid17	Taking a test of your ability, skill, or knowledge	.54	.29	.10	.07	.00	.58	.29	.09	.04	.00
avoid18	Expressing disagreement or disapproval to someone you don't know very well	.22	.45	.21	.12	.00	.23	.47	.20	.10	.00
avoid19	Looking someone who you don't know very well straight in the eyes	.41	.36	.12	.11	.00	.35	.40	.18	.07	.00
avoid20	Giving a prepared oral talk to a group	.25	.30	.22	.23	.00	.30	.33	.18	.19	.00
avoid21	Trying to make someone's acquaintance for the purpose of a romantic/sexual relationship	.23	.27	.20	.29	.00	.30	.27	.18	.24	.00
avoid22	Returning goods to a store for a refund	.49	.30	.14	.08	.00	.49	.30	.12	.08	.00
avoid23	Giving a party	.46	.19	.16	.19	.00	.46	.23	.14	.17	.00
avoid24	Resisting a high pressure sales person	.47	.30	.14	.09	.00	.51	.28	.13	.08	.00

SM4 The issue of assessing unidimensionality in psychological scales

Reise et al. (2015) proposed a method for testing the extent to which parameters of a unidimensional IRT model are biased when multidimensionality is not adequately modeled that relies on bifactor modeling. Specifically, it is assumed that while one common factor underlies the variance of all the scale items, a set of orthogonal group factors are specified in order to account for additional variation, typically assumed to be due to item parcels with similar content (Reise et al., 2015). Other methods have been proposed (see, e.g., Jennrich & Benter, 2011, 2012; Stucky & Edelen, 2015), but as pointed out by Reise (2012), bifactor modeling is not necessarily an appropriate analytic tool for any type of psychological measure. In fact, it is best suited for the psychometric analysis of those assessment tools where the researcher expects a response to primarily reflect a strong common trait, but there is multidimensionality caused by well-defined clusters of items from diverse subdomains. In other words, a bifactor model requires that the multidimensionality is well-structured, namely, each item loads in a general trait and one, and only one, subtrait. The presence of items displaying cross-loadings on the group factors, although admissible in exploratory solutions, leads to biased and untrustworthy item parameter estimates in bifactor solutions (Reise et al., 2010). This implies that measures that were not originally developed with a clear blueprint to include at least three items from at least three content domains and do not have a clear, replicable, and stable multidimensional structure (as it is the case of LSAS-SR), cannot be assessed with such method (Reise, 2012). More recently, Rodriguez, Reise, and Haviland (2016) showed how 50 measures that were reportedly "multidimensional" still had unit-weighted total scores clearly reflecting variance due to a single latent variable. In other words, they could be interpreted as univocal indicators of a single latent variable, despite the multidimensionality. They concluded that in many cases total scores are robust to the biasing effects of unmodeled multidimensionality, since when correlated items are aggregated

together, and they all share a single common factor, the more items that are grouped, the more the total score reflects that common latent variable, regardless of the dimensionality (Gustafsson & Aberg-Bengtsson, 2010).

SM5 Use and interpretation of results from dimensionality analyses

The scree-test (Cattell, 1966) is a visual inspection of a graph in which the eigenvalues obtained in the factor analysis are plotted against their associated components, in order of extraction, with a straight line connecting the points. As the number of factors increases, the slope of the line connecting two successive factors becomes less and less steep. The optimal number of factors is determined by the point at which the line begins to flatten out. In this case, unidimensionality would be supported by a scree-plot flattening out from the second factor. Due to its subjectivity in the definition of the cutoff point between the important and trivial factors, the scree-test has been often criticized, and other analytical, and thus objective, methods to explore the dimensionality of an item pool have been proposed.

Parallel analysis (Horn, 1965) is based on extracting eigenvalues from randomly generated correlation matrices (usually 1,000, Buja & Eyuboglu, 1992) with the same number of variables and participants of the original. The optimal number of factors corresponds to the number of real data eigenvalues larger than the 95th percentile (Longman et al., 1989) of the distribution of the corresponding randomly generated ones. As a result, unidimensionality would be supported if only the first real-data eigenvalue is larger than the corresponding randomly generated one.

The Minim Average Partial (MAP) Correlation statistic (Velicer, 1976) is based on the average partial correlations between the variables after successively removing the effect of the factors, one at the time following the decreasing order of their eigenvalues. After each step, the squared average partial correlation between the items is computed and the number of factors that minimizes this value indicates the optimal solution. Unidimensionality would be supported by having the first factor yielding the lowest squared average partial correlation.

SM6 Description of invariance models and DIF detection method

We initially specified a so-called *configural invariance model* (M1). Technically, this model is not really an "invariance" model, as there is no invariance of any parameter estimate (i.e., all parameters, slopes, and thresholds, are freely estimated), but it only evaluates the similarity of the overall pattern of parameters. However, it provides evidence of the ability of the a priori model to fit the data in each group without invariance constraints and sets a baseline for comparing the following models that actually impose equality constraints on the parameter estimates across samples. We then specified a model in which item slopes were constrained to be invariant across samples ("weak" or "metric" invariance, M2). This model assumes that each item has the same discrimination in either sample, i.e., the probability of endorsing a certain answer with a level of social anxiety near the item threshold changes equally fast for each item regardless of the sample. However, due to model identification issues, neither M1 nor M2 allow testing for differences in latent means. This is possible only with a model that imposes invariance also on item thresholds ("strong" or "scalar" invariance, M3). In the comparison of these invariance models, we used the same criteria listed earlier for the comparison of the GRM and the reduced GRM models.

M3 also allows the evaluation of differential item functioning (DIF), i.e., whether individuals in different samples but with the same level of social anxiety (as measured by the LSAS) have different probabilities of endorsing a certain answer to a certain item. The presence of DIF for all items would prevent us to consider the estimates of the difference in latent means as a valid test for differences in mean levels of social anxiety, while the absence of DIF would be the ideal outcome. More realistically, we could expect *partial* invariance of the thresholds, i.e., thresholds are invariant for some items, and not for others (M3p). For identification of the parameters of interest, partial invariance of at least two indicators per latent trait is considered sufficient (Byrne, Shavelson, & Muthén, 1989).

For detecting DIF, we used the method described in Meade and Wright (2012). We first tested a fully constrained model (slopes and thresholds invariant, factor means, and variances free) and we used likelihood ratio tests (LRTs, Thissen et al., 1988, 1993) for the testing test each of the items by freeing the parameters of each item, one at a time. The other items served as the anchor items (i.e., the so-called "all-others-as-anchors model"). We then examined the output and choose the five items with a non-significant LRT (hence, with no DIF) with the largest slope parameters as "anchor items" (the A5 method, Meade & Wright, 2012). We then tested a partial invariance model in order to correctly identify items with DIF and test latent mean differences. However, one drawback of the LRT is that it has high power to detect even very small differences in item functioning when sample sizes are large, as it is (almost) always the case when estimating IRT models. Hence, we evaluated DIF for each item using the Expected Score Standardized Difference (ESSD), which can be interpreted as the commonly used Cohen's *d* (i.e., $|d| < 0.20$ negligible effect; $0.20 \leq |d| < 0.50$: small effect; $0.50 \leq |d| < 0.80$: moderate effect; $|d| \geq 0.80$ large effect).

SM7 Results of the Exploratory Factor Analyses on the Liebowitz Social Anxiety Scale-Self-Report in all samples

Table SM7.1a Factor loadings and their 95% bootstrapped confidence interval for the single-factor exploratory factor analyses on the Fear scale

Item	Community-dwelling sample (*n* = 356)	SAD-diagnosed sample (*n* = 257)	Female sample (*n* = 390)	Male sample (*n = 221*)
fear01	.49 [.41; .57]	.55 [.45; .66]	.72 [.66; .79]	.64 [.54; .74]
fear02	.63 [.57; .70]	.60 [.51; .70]	.76 [.71; .81]	.83 [.77; .88]
fear03	.52 [.43; .59]	.40 [.27; .52]	.66 [.58; .74]	.65 [.55; .74]
fear04	.61 [.54; .68]	.45 [.30; .59]	.67 [.58; .77]	.78 [.68; .86]
fear05	.61 [.54; .68]	.55 [.45; .64]	.71 [.65; .77]	.67 [.58; .76]
fear06	.63 [.56; .70]	.53 [.38; .68]	.73 [.66; .81]	.76 [.67; .84]
fear07	.68 [.61; .74]	.58 [.47; .68]	.83 [.79; .88]	.81 [.75; .87]
fear08	.68 [.62; .74]	.59 [.49; .70]	.78 [.73; .83]	.78 [.72; .84]
fear09	.57 [.48; .64]	.50 [.38; .61]	.71 [.64; .78]	.60 [.50; .70]
fear10	.63 [.56; .70]	.71 [.63; .79]	.75 [.70; .80]	.73 [.66; .80]
fear11	.71 [.66; .77]	.70 [.62; .79]	.79 [.75; .84]	.83 [.78; .89]
fear12	.72 [.66; .78]	.73 [.65; .80]	.80 [.75; .85]	.82 [.76; .87]
fear13	.53 [.45; .61]	.58 [.46; .71]	.60 [.50; .71]	.59 [.46; .71]
fear14	.70 [.64; .76]	.68 [.59; .75]	.80 [.75; .84]	.78 [.71; .85]
fear15	.72 [.67; .78]	.65 [.56; .75]	.81 [.76; .85]	.86 [.81; .91]
fear16	.71 [.65; .76]	.59 [.47; .70]	.76 [.71; .82]	.82 [.76; .87]
fear17	.35 [.25; .44]	.36 [.23; .49]	.27 [.16; .37]	.17 [.02; .32]
fear18	.65 [.58; .71]	.65 [.56; .74]	.69 [.63; .75]	.74 [.67; .81]
fear19	.65 [.59; .72]	.68 [.60; .77]	.73 [.67; .79]	.78 [.72; .85]
fear20	.56 [.49; .64]	.49 [.37; .61]	.66 [.60; .73]	.68 [.59; .77]
fear21	.61 [.54; .68]	.50 [.40; .62]	.66 [.59; .73]	.73 [.64; .81]
fear22	.49 [.41; .58]	.58 [.49; .68]	.70 [.64; .76]	.64 [.55; .74]
fear23	.62 [.55; .69]	.52 [.41; .63]	.74 [.69; .80]	.72 [.63; .80]
fear24	.48 [.39; .57]	.50 [.38; .61]	.60 [.52; .68]	.57 [.44; .69]

Table SM7.1b Factor loadings and their 95% bootstrapped confidence interval for the single-factor exploratory factor analyses on the Avoidance scale

Item	Community-dwelling sample (*n* = 356)	SAD-diagnosed sample (*n* = 257)	Female sample (*n* = 390)	Male sample (*n = 221*)
avoid01	.44 [.35; .53]	.55 [.45; .64]	.65 [.58; .72]	.48 [.35; .61]
avoid02	.61 [.54; .68]	.68 [.60; .76]	.78 [.73; .83]	.81 [.75; .87]
avoid03	.47 [.38; .56]	.36 [.23; .49]	.62 [.53; .70]	.55 [.42; .67]
avoid04	.39 [.29; .49]	.32 [.15; .47]	.58 [.47; .68]	.53 [.39; .66]
avoid05	.60 [.53; .68]	.61 [.51; .70]	.73 [.68; .79]	.72 [.64; .81]
avoid06	.47 [.38; .57]	.60 [.48; .72]	.66 [.59; .74]	.66 [.57; .76]
avoid07	.55 [.47; .63]	.63 [.54; .73]	.76 [.71; .82]	.80 [.74; .86]
avoid08	.61 [.54; .69]	.59 [.50; .69]	.72 [.66; .78]	.74 [.66; .82]
avoid09	.56 [.47; .64]	.48 [.36; .59]	.69 [.61; .75]	.56 [.45; .67]
avoid10	.60 [.52; .68]	.73 [.66; .80]	.76 [.71; .81]	.75 [.67; .82]
avoid11	.69 [.63; .75]	.75 [.68; .82]	.82 [.78; .86]	.85 [.79; .90]
avoid12	.68 [.62; .74]	.74 [.66; .82]	.82 [.78; .87]	.81 [.75; .88]
avoid13	.41 [.32; .50]	.31 [.14; .47]	.47 [.36; .58]	.43 [.30; .56]
avoid14	.61 [.54; .68]	.59 [.49; .69]	.74 [.67; .80]	.76 [.69; .83]
avoid15	.60 [.52; .67]	.70 [.62; .79]	.74 [.68; .80]	.83 [.78; .88]
avoid16	.64 [.56; .71]	.75 [.67; .84]	.73 [.67; .79]	.84 [.79; .89]
avoid17	.45 [.36; .54]	.29 [.14; .43]	.44 [.34; .54]	.40 [.26; .54]
avoid18	.53 [.45; .61]	.67 [.58; .76]	.64 [.57; .72]	.77 [.70; .83]
avoid19	.58 [.50; .65]	.62 [.53; .71]	.73 [.67; .79]	.80 [.74; .87]
avoid20	.57 [.49; .65]	.63 [.53; .72]	.69 [.63; .76]	.75 [.67; .82]
avoid21	.45 [.36; .54]	.55 [.44; .66]	.57 [.49; .66]	.69 [.59; .79]
avoid22	.38 [.28; .48]	.52 [.40; .63]	.62 [.55; .70]	.63 [.53; .73]
avoid23	.54 [.46; .62]	.63 [.53; .73]	.75 [.69; .81]	.78 [.71; .85]
avoid24	.41 [.32; .50]	.43 [.31; .55]	.48 [.38; .57]	.58 [.46; .70]

Table SM7.2 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Fear scale, in the community-dwelling sample (*n* = 356)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
fear01	**.40**	.15	**.30**	.13	.15	.14	.15	.27	.08
fear02	**.35**	**.37**	.24	**.34**	.19	**.36**	.20	.17	.08
fear03	**.59**	-.04	**.56**	-.01	.05	.11	.00	-.02	**.69**
fear04	**.85**	-.20	**.78**	-.17	.11	-.06	.02	.04	**1.00**
fear05	.28	**.42**	.12	**.37**	.25	**.43**	.22	-.13	.26
fear06	-.08	**.89**	-.02	**.89**	-.04	**.91**	-.06	-.03	-.02
fear07	**.65**	.09	**.66**	.13	.01	.18	.05	**.44**	.26
fear08	**.53**	.23	**.50**	.25	.07	.30	.11	.27	.22
fear09	**.53**	.08	**.59**	.13	-.06	.20	-.01	.28	**.32**
fear10	**.48**	.22	.00	.04	**.71**	.01	**.72**	.14	-.11
fear11	**.59**	.20	-.01	-.03	**.90**	-.01	**.88**	-.08	.10
fear12	**.51**	.30	.03	.13	**.72**	.13	**.70**	.08	-.02
fear13	**.60**	-.04	**.48**	-.04	.18	-.04	.22	**.44**	.08
fear14	**.43**	**.36**	**.30**	**.32**	.21	**.35**	.23	.22	.10
fear15	.09	**.79**	.05	**.75**	.13	**.75**	.12	.08	-.04
fear16	.05	**.83**	.01	**.78**	.12	**.80**	.10	-.01	.01
fear17	.22	.18	.12	.15	.15	.21	.13	-.15	.27
fear18	**.34**	**.41**	-.01	.29	**.51**	.28	**.50**	.06	-.06
fear19	**.60**	.11	.21	-.01	**.58**	.02	**.57**	.04	.20
fear20	-.01	**.71**	.06	**.72**	-.05	**.74**	-.05	.03	.00
fear21	**.46**	.21	.30	.18	.25	.17	.27	**.34**	-.01
fear22	**.57**	-.05	**.56**	-.01	.02	-.05	.06	**.73**	-.06
fear23	**.58**	.10	**.72**	.17	-.16	.21	-.13	**.63**	.17
fear24	**.68**	-.18	**.50**	-.19	.24	-.19	.29	**.46**	.09
r with F2	.55		.48			.52			
r with F3			.61	.50		.41	.46		
r with F4						.31	.46	.44	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.3 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Fear scale, in the SAD-diagnosed sample (n = 257)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
fear01	**.56**	.03	**.53**	.04	.05	**.52**	.05	.04	.10
fear02	**.43**	.27	**.32**	.25	.21	**.30**	.26	.21	.06
fear03	**.34**	.10	-.01	-.03	**.75**	-.04	.00	**.74**	.03
fear04	**.34**	.16	-.03	.04	**.78**	-.06	.06	**.78**	.00
fear05	**.38**	.26	**.39**	.28	.00	**.37**	.26	-.04	**.31**
fear06	-.06	**.88**	-.02	**.91**	-.09	-.02	**.91**	-.10	.05
fear07	**.63**	-.04	**.52**	-.05	.22	**.52**	-.02	.28	-.27
fear08	**.48**	.18	**.34**	.15	.28	.29	.12	.28	**.41**
fear09	**.45**	.09	.26	.03	**.39**	.20	-.02	**.41**	**.47**
fear10	**.80**	-.07	**.85**	-.03	-.08	**.84**	-.02	-.09	.07
fear11	**.80**	-.08	**.84**	-.05	-.05	**.82**	-.04	-.06	.10
fear12	**.80**	-.05	**.79**	-.02	.03	**.77**	-.01	.04	.04
fear13	**.62**	-.02	**.47**	-.05	.30	**.45**	-.04	**.30**	.08
fear14	**.54**	.22	**.39**	.19	**.30**	**.37**	.21	**.31**	-.01
fear15	.17	**.72**	.09	**.70**	.16	.07	**.73**	.19	-.12
fear16	-.02	**.90**	-.04	**.89**	.05	-.04	**.89**	.04	.05
fear17	.18	.26	.10	.24	.16	.08	.23	.15	.13
fear18	**.49**	.26	**.53**	**.30**	-.08	**.52**	.30	-.09	.16
fear19	**.64**	.09	**.56**	.09	.17	**.54**	.11	.18	-.01
fear20	-.01	**.74**	.01	**.75**	-.04	.01	**.77**	-.03	-.08
fear21	**.56**	-.04	**.57**	-.02	-.02	**.60**	.02	.01	-.29
fear22	**.61**	.00	**.64**	.03	-.04	**.64**	.05	-.03	-.09
fear23	**.51**	.04	**.46**	.04	.11	**.47**	.08	.17	**-.36**
fear24	**.54**	-.03	**.57**	.01	-.05	**.57**	.03	-.04	-.07
r with F2	.45		.43			.42			
r with F3			.40	.31		.41	.30		
r with F4						.13	.11	.04	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.4 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Avoidace scale, in the community-dwelling sample (*n* = 356)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
avoid01	.25	**.39**	**.38**	.21	.11	**.34**	.19	.14	.04
avoid02	**.45**	**.39**	**.43**	**.46**	.04	**.40**	**.43**	.13	-.01
avoid03	-.05	**.84**	**.82**	-.04	.06	**.78**	-.07	.10	.05
avoid04	-.23	**1.01**	**1.01**	-.15	-.03	**.91**	-.19	.20	-.11
avoid05	**.51**	.29	.30	**.46**	.12	.22	**.43**	.30	-.06
avoid06	**.81**	-.25	-.15	**.87**	-.06	-.13	**.85**	-.01	-.02
avoid07	**.41**	**.38**	.21	.09	**.51**	.26	.08	.06	**.53**
avoid08	.20	**.71**	**.67**	.15	.14	**.71**	.13	-.02	.21
avoid09	.17	**.68**	**.69**	.18	.05	**.76**	.16	-.11	.18
avoid10	**.42**	**.40**	.20	.05	**.58**	-.01	.00	**.76**	.10
avoid11	**.43**	**.52**	**.35**	.10	**.52**	.09	.04	**.86**	-.02
avoid12	**.56**	**.35**	.18	.20	**.56**	-.01	.17	**.70**	.11
avoid13	.18	**.48**	**.38**	.02	.28	**.32**	.00	.25	.14
avoid14	**.43**	**.44**	**.51**	**.49**	-.03	**.53**	**.46**	-.01	.03
avoid15	**.77**	-.03	.03	**.77**	.04	.03	**.75**	.09	.01
avoid16	**.81**	-.02	.01	**.73**	.14	.04	**.71**	.06	.13
avoid17	**.45**	.19	.11	.26	**.31**	.13	.25	.09	.28
avoid18	**.53**	.15	.12	**.40**	.22	.11	**.39**	.14	.14
avoid19	**.32**	**.50**	**.39**	.10	**.36**	.26	.08	**.47**	.06
avoid20	**.80**	-.10	-.04	**.77**	.07	-.05	**.75**	.12	.01
avoid21	**.64**	-.08	-.30	.21	**.64**	-.29	.22	.27	**.45**
avoid22	.30	.26	-.03	-.21	**.76**	.01	-.22	.16	**.70**
avoid23	**.56**	.20	.04	.23	**.52**	.13	.22	-.06	**.63**
avoid24	**.38**	.18	-.03	-.01	**.60**	-.02	.00	.20	**.47**
r with F2	.33		.22			.18			
r with F3			.43	.53		.50	.41		
r with F4						.27	.44	.47	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.5 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Avoidance scale, in the SAD-diagnosed sample (*n* = 257)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
avoid01	**.37**	**.34**	**.52**	.00	.14	.06	**.31**	.14	.29
avoid02	**.61**	.18	.23	**.46**	.21	**.45**	.23	.20	.04
avoid03	-.01	**.64**	.01	.01	**.69**	-.03	.12	**.70**	-.11
avoid04	-.08	**.69**	-.13	.03	**.84**	.03	-.08	**.83**	-.05
avoid05	**.54**	.16	.27	**.37**	.14	**.38**	.19	.14	.13
avoid06	**.83**	-.28	.01	**.85**	-.11	**.79**	.15	-.12	-.13
avoid07	**.47**	**.32**	**.46**	.16	.18	.06	**.60**	.20	-.08
avoid08	**.32**	**.50**	.16	.24	**.52**	.24	.16	**.52**	.02
avoid09	.15	**.58**	.11	.10	**.59**	.10	.13	**.59**	.01
avoid10	**.60**	.28	**.87**	-.01	-.07	.00	**.71**	-.04	.27
avoid11	**.56**	**.38**	**.80**	-.01	.07	-.06	**.78**	.10	.13
avoid12	**.65**	.22	**.67**	.17	-.01	.04	**.85**	.00	-.08
avoid13	-.14	**.77**	**.33**	**-.35**	**.57**	-.26	.09	**.57**	**.30**
avoid14	**.41**	**.35**	.18	**.31**	**.37**	**.38**	.01	**.35**	.21
avoid15	**.83**	-.12	.14	**.75**	.00	**.67**	**.31**	-.01	-.15
avoid16	**.87**	-.09	.08	**.85**	.06	**.93**	-.05	.03	.15
avoid17	.12	**.31**	-.06	.19	**.39**	.23	-.13	**.38**	.08
avoid18	**.69**	.05	**.49**	**.34**	-.08	**.43**	.22	-.09	**.37**
avoid19	**.46**	**.32**	**.60**	.03	.10	.01	**.57**	.12	.11
avoid20	**.74**	-.10	-.08	**.84**	.11	**.86**	-.09	.08	.02
avoid21	**.59**	-.01	**.47**	.25	-.15	.22	**.48**	-.13	.05
avoid22	**.37**	.29	**.62**	-.07	.03	.08	.14	.00	**.67**
avoid23	**.62**	.08	**.48**	.28	-.05	.19	**.60**	-.04	-.06
avoid24	.29	.26	**.54**	-.09	.03	.06	.07	.00	**.64**
r with F2	.28		.52			.54			
r with F3			.38	.13		.13	.31		
r with F4						.19	.39	.26	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.6 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Fear scale, in the female sample (n = 390)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
fear01	**.62**	.13	**.51**	.14	.15	**.58**	.17	.08	-.16
fear02	**.58**	.22	**.42**	.25	.20	**.46**	.26	.15	-.11
fear03	**.76**	-.09	.02	.06	**.78**	.03	.05	**.77**	-.05
fear04	**.79**	-.12	.02	.03	**.82**	.03	.03	**.81**	-.05
fear05	**.35**	**.43**	**.40**	**.40**	-.01	**.32**	**.40**	.05	.18
fear06	-.06	**.94**	.03	**.92**	-.08	.06	**.92**	-.11	-.05
fear07	**.85**	.00	**.51**	.08	**.38**	**.59**	.09	**.31**	-.22
fear08	**.61**	.21	.22	**.31**	**.39**	.18	**.31**	**.43**	.04
fear09	**.67**	.06	.14	.20	**.53**	.03	.17	**.64**	.17
fear10	**.71**	.07	**.90**	-.02	-.09	**.91**	.01	-.12	-.02
fear11	**.75**	.07	**.90**	-.01	-.05	**.85**	.01	-.03	.10
fear12	**.71**	.12	**.77**	.08	.01	**.74**	.10	.02	.04
fear13	**.78**	-.18	**.51**	-.12	.30	**.40**	-.15	**.40**	.23
fear14	**.66**	.18	**.43**	.23	.25	**.40**	.23	.27	.04
fear15	.21	**.72**	.03	**.78**	.16	.06	**.77**	.15	-.06
fear16	.04	**.87**	-.01	**.88**	.04	.01	**.87**	.04	-.01
fear17	.20	.09	.11	.11	.10	-.04	.08	.23	**.33**
fear18	**.33**	**.43**	**.56**	**.37**	-.19	**.44**	**.36**	-.10	**.34**
fear19	**.74**	.01	**.62**	.02	.17	**.51**	.01	.27	.25
fear20	.04	**.74**	-.04	**.78**	.06	-.10	**.76**	.12	.14
fear21	**.60**	.08	**.57**	.08	.07	**.54**	.08	.10	.07
fear22	**.76**	-.06	**.59**	-.02	.21	**.61**	.00	.18	-.08
fear23	**.72**	.05	**.39**	.12	**.34**	**.48**	.14	.27	-.24
fear24	**.69**	-.08	**.62**	-.08	.12	**.58**	-.07	.14	.07
r with F2	.73		.70			.68			
r with F3			.61	.52		.63	.54		
r with F4						.16	.13	.01	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.7 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Fear scale, in the male sample (n = 221)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
fear01	**.55**	.14	.17	**.49**	.07	**.48**	.06	.21	-.06
fear02	.26	**.65**	**.65**	.22	.07	.21	**.47**	**.31**	-.01
fear03	.25	**.45**	**.47**	.22	.04	.05	.08	**.70**	-.01
fear04	.26	**.59**	**.59**	.19	.12	.05	.18	**.73**	.05
fear05	**.33**	**.40**	**.38**	.20	.21	**.30**	**.31**	.12	.13
fear06	-.13	**.99**	**.96**	-.19	.13	-.09	**.86**	.11	.14
fear07	**.48**	**.40**	**.43**	**.45**	.02	**.41**	.28	.28	-.11
fear08	**.58**	.25	.18	**.30**	**.52**	**.52**	.12	.17	.27
fear09	**.46**	.19	.03	.05	**.82**	**.40**	.01	.17	**.46**
fear10	**.91**	-.13	-.07	**.80**	.11	**.89**	-.02	-.09	.05
fear11	**.84**	.05	.09	**.69**	.20	**.74**	-.02	.18	.13
fear12	**.82**	.05	.10	**.69**	.16	**.75**	.04	.11	.09
fear13	**.61**	.02	.04	**.51**	.13	**.53**	-.04	.18	-.01
fear14	**.35**	**.49**	**.51**	**.31**	.05	.25	.26	**.43**	-.02
fear15	.15	**.80**	**.81**	.12	.04	.17	**.69**	.15	.03
fear16	.03	**.88**	**.89**	.03	.00	.06	**.78**	.13	.02
fear17	.00	.18	.13	-.17	**.31**	.08	**.31**	**-.33**	**.42**
fear18	**.64**	.16	.16	**.48**	.24	**.63**	.17	-.04	.20
fear19	**.63**	.22	.23	**.50**	.18	**.58**	.17	.10	.13
fear20	-.04	**.81**	**.85**	.02	-.14	.03	**.90**	-.08	-.12
fear21	**.60**	.18	.22	**.55**	.04	**.59**	.22	.01	-.03
fear22	**.66**	.02	.11	**.78**	-.23	**.73**	.16	-.09	**-.32**
fear23	**.48**	.30	**.34**	**.47**	-.03	**.47**	**.31**	.08	-.14
fear24	**.82**	-.22	-.16	**.77**	.04	**.75**	-.18	.08	-.11
r with F2	.70		.65			.62			
r with F3			.50	.51		.59	.61		
r with F4						.14	.17	.19	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.8 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Avoidace scale, in the female sample (*n* = 390)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
avoid01	**.45**	.29	**.51**	.10	.12	.14	**.37**	.15	.18
avoid02	**.54**	**.35**	.23	**.39**	**.34**	**.37**	**.31**	**.33**	-.08
avoid03	-.01	**.80**	.10	-.04	**.76**	-.07	.20	**.75**	-.09
avoid04	-.12	**.89**	-.07	-.04	**.93**	-.07	.02	**.91**	-.06
avoid05	**.66**	.15	**.31**	**.43**	.11	**.43**	**.33**	.11	.00
avoid06	**.93**	-.24	.02	**.90**	-.10	**.87**	.11	-.11	-.08
avoid07	**.48**	**.39**	**.49**	.15	.24	.18	**.42**	.26	.10
avoid08	.23	**.64**	.13	.17	**.62**	.17	.11	**.62**	.05
avoid09	.15	**.70**	.06	.13	**.71**	.14	-.02	**.71**	.13
avoid10	**.56**	.30	**.89**	-.05	-.03	.01	**.74**	.01	.18
avoid11	**.54**	**.39**	**.81**	-.02	.10	.00	**.82**	.12	.03
avoid12	**.64**	.28	**.73**	.13	.04	.15	**.73**	.06	.03
avoid13	-.05	**.65**	.26	-.19	**.52**	-.15	.04	**.55**	.27
avoid14	**.43**	**.42**	.11	**.37**	**.45**	**.38**	.05	**.45**	.10
avoid15	**.83**	-.04	.14	**.72**	.04	**.69**	.25	.03	-.10
avoid16	**.88**	-.10	.06	**.83**	.02	**.84**	-.03	.01	.12
avoid17	.20	**.31**	-.03	.23	**.36**	.28	-.25	**.39**	.26
avoid18	**.69**	.01	**.38**	**.41**	-.06	**.44**	.23	-.03	.18
avoid19	**.41**	**.43**	**.61**	.00	.22	.03	**.55**	.24	.08
avoid20	**.78**	-.03	-.08	**.84**	.13	**.83**	-.11	.12	.06
avoid21	**.64**	-.03	**.57**	.24	-.19	.29	**.38**	-.16	.22
avoid22	**.42**	.29	**.67**	-.04	.04	.04	.29	.09	**.49**
avoid23	**.64**	.20	**.58**	.23	.03	.26	**.49**	.05	.11
avoid24	**.36**	.18	**.66**	-.10	-.08	-.03	.19	-.05	**.66**
r with F2	.52		.66			.60			
r with F3			.59	.38		.37	.54		
r with F4						.35	.44	.32	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

Table SM7.9 Factor loadings of the two-, three-, and four-factor solutions from exploratory factor analysis (MINRES extraction, oblimin rotation) on the Liebowitz Social Anxiety Scale- Self-Report, Avoidance scale, in the male sample (*n = 221*)

	2-factor solution		3-factor solution			4-factor solution			
Item	F1	F2	F1	F2	F3	F1	F2	F3	F4
avoid01	.16	**.46**	.04	.19	**.40**	.06	.19	**.38**	.05
avoid02	**.68**	.22	**.61**	.07	**.31**	**.66**	.06	**.39**	-.06
avoid03	.01	**.77**	.02	-.05	**.84**	.04	-.01	**.75**	.15
avoid04	-.09	**.87**	-.13	.06	**.86**	-.11	.09	**.78**	.14
avoid05	**.52**	**.32**	**.38**	.18	**.32**	**.36**	.20	.17	.21
avoid06	**.93**	-.29	**.93**	-.04	-.12	**.90**	-.04	-.15	.06
avoid07	**.62**	.29	**.35**	**.40**	.20	**.35**	**.40**	.17	.06
avoid08	**.31**	**.62**	.23	.09	**.66**	.16	.11	.27	**.55**
avoid09	.17	**.56**	.14	.01	**.61**	.00	-.03	.02	**.95**
avoid10	**.56**	.30	.01	**.86**	-.04	.00	**.85**	-.09	.10
avoid11	**.56**	**.44**	.11	**.69**	.20	.10	**.69**	.12	.13
avoid12	**.67**	.24	.29	**.58**	.07	.29	**.57**	.05	.04
avoid13	-.01	**.62**	**-.30**	**.46**	**.41**	-.29	**.46**	**.34**	.10
avoid14	**.51**	**.39**	**.47**	.02	**.49**	**.50**	.02	**.50**	.02
avoid15	**.91**	-.04	**.83**	.08	.08	**.82**	.08	.05	.05
avoid16	**.90**	.00	**.80**	.11	.10	**.78**	.11	.03	.12
avoid17	**.33**	.13	.21	.17	.10	.12	.19	-.24	**.45**
avoid18	**.67**	.19	**.40**	**.38**	.11	**.41**	**.37**	.12	.02
avoid19	**.60**	**.33**	.27	**.49**	.19	.28	**.48**	.17	.06
avoid20	**.95**	-.19	**.91**	.01	-.04	**.89**	.01	-.05	.02
avoid21	**.77**	-.04	**.41**	**.53**	-.19	**.44**	**.50**	-.07	-.13
avoid22	**.44**	.30	-.09	**.83**	-.04	-.07	**.81**	.01	-.04
avoid23	**.76**	.08	**.56**	.26	.09	**.55**	.26	.02	.11
avoid24	**.40**	.27	.10	**.46**	.11	.13	**.45**	.17	-.05
r with F2	.46		.64			.63			
r with F3			.35	.56		.27	.46		
r with F4						.36	.46	.49	

Note: r with F1, F2, F3: correlation with F1, F2, and F3, respectively; Bolded loadings indicate loadings larger than |.30|; Greyed cells indicate items that do not conform to approximate simple structure (see manuscript), for ease of interpretation.

SM8 Parameter estimates for the Graded Response Model in all groups

Table SM8.1 Parameter estimates for the Graded Response Model (GRM) Liebowitz Social Anxiety Scale-Self-Report Fear scale. Bracketed values show the 95% confidence interval.

	Community-dwelling Sample (*n* = 356)				SAD-diagnosed Sample (*n* = 257)			
Item	Discrimination	Threshold 1	Threshold 2	Threshold 3	Discrimination	Threshold 1	Threshold 2	Threshold 3
fear01	0.97 [0.66; 1.28]	-1.11 [-1.40; -0.83]	-3.53 [-4.11; -2.95]	-4.96 [-5.99; -3.92]	1.17 [0.86; 1.49]	1.20 [0.86; 1.54]	-0.80 [-1.12; -0.47]	-2.58 [-3.06; -2.11]
fear02	1.55 [1.20; 1.90]	0.03 [-0.26; 0.33]	-2.87 [-3.35; -2.39]	-5.95 [-7.21; -4.70]	1.26 [0.93; 1.59]	3.48 [2.85; 4.12]	0.64 [0.31; 0.96]	-2.12 [-2.55; -1.69]
fear03	1.02 [0.69; 1.34]	-1.46 [-1.78; -1.15]	-2.97 [-3.45; -2.49]	-5.30 [-6.48; -4.12]	0.71 [0.45; 0.97]	0.91 [0.61; 1.20]	-0.48 [-0.75; -0.20]	-1.92 [-2.30; -1.55]
fear04	1.39 [0.86; 1.92]	-3.20 [-3.84; -2.55]	-4.51 [-5.41; -3.61]	-6.88 [-8.99; -4.77]	0.88 [0.59; 1.18]	0.15 [-0.14; 0.43]	-1.63 [-1.98; -1.27]	-3.30 [-3.90; -2.69]
fear05	1.44 [1.13; 1.75]	1.68 [1.34; 2.02]	-0.82 [-1.12; -0.53]	-4.12 [-4.79; -3.46]	1.14 [0.82; 1.45]	3.68 [3.00; 4.36]	1.25 [0.90; 1.59]	-0.84 [-1.16; -0.51]
fear06	1.84 [1.48; 2.20]	3.69 [3.10; 4.29]	0.86 [0.53; 1.19]	-1.88 [-2.26; -1.49]	1.16 [0.74; 1.57]	4.77 [3.70; 5.84]	3.00 [2.42; 3.58]	1.49 [1.09; 1.88]
fear07	1.56 [1.18; 1.94]	-1.01 [-1.33; -0.68]	-3.48 [-4.06; -2.90]	-5.64 [-6.77; -4.50]	1.27 [0.94; 1.61]	2.97 [2.43; 3.51]	0.84 [0.51; 1.18]	-1.03 [-1.38; -0.69]
fear08	1.67 [1.32; 2.02]	0.60 [0.29; 0.91]	-2.22 [-2.63; -1.81]	-4.52 [-5.27; -3.77]	1.25 [0.92; 1.59]	3.09 [2.53; 3.65]	0.96 [0.62; 1.29]	-1.02 [-1.36; -0.68]
fear09	1.21 [0.89; 1.53]	-0.58 [-0.85; -0.30]	-2.75 [-3.19; -2.30]	-4.18 [-4.90; -3.46]	0.99 [0.70; 1.28]	1.37 [1.03; 1.71]	-0.41 [-0.70; -0.11]	-2.13 [-2.54; -1.72]
fear10	1.40 [1.09; 1.72]	0.99 [0.69; 1.29]	-2.03 [-2.40; -1.65]	-4.42 [-5.16; -3.68]	1.89 [1.46; 2.32]	3.21 [2.59; 3.82]	0.63 [0.24; 1.02]	-1.57 [-2.01; -1.13]
fear11	1.83 [1.43; 2.22]	0.22 [-0.10; 0.53]	-3.09 [-3.62; -2.56]	-6.09 [-7.30; -4.89]	1.97 [1.52; 2.42]	3.26 [2.63; 3.90]	0.47 [0.07; 0.86]	-1.90 [-2.37; -1.42]
fear12	1.88 [1.49; 2.27]	0.77 [0.44; 1.10]	-2.28 [-2.72; -1.85]	-4.83 [-5.63; -4.02]	2.03 [1.56; 2.51]	3.31 [2.66; 3.96]	1.23 [0.79; 1.67]	-0.89 [-1.31; -0.48]
fear13	1.03 [0.68; 1.38]	-1.68 [-2.02; -1.34]	-2.88 [-3.35; -2.40]	-4.26 [-5.03; -3.48]	1.31 [0.94; 1.69]	-0.37 [-0.70; -0.05]	-1.85 [-2.26; -1.43]	-3.32 [-3.94; -2.70]
fear14	1.78 [1.40; 2.17]	0.32 [0.00; 0.63]	-3.04 [-3.55; -2.52]	-5.19 [-6.10; -4.29]	1.56 [1.19; 1.93]	2.89 [2.35; 3.43]	0.38 [0.03; 0.73]	-2.03 [-2.47; -1.58]
fear15	2.27 [1.83; 2.70]	2.48 [2.01; 2.96]	-0.70 [-1.06; -0.33]	-3.35 [-3.92; -2.77]	1.38 [0.98; 1.79]	4.75 [3.75; 5.75]	3.03 [2.45; 3.61]	0.55 [0.21; 0.90]
fear16	2.14 [1.73; 2.55]	2.41 [1.96; 2.86]	-0.42 [-0.77; -0.07]	-3.08 [-3.61; -2.56]	1.17 [0.80; 1.54]	4.34 [3.45; 5.22]	2.25 [1.80; 2.70]	0.48 [0.16; 0.80]
fear17	0.65 [0.43; 0.87]	1.26 [1.00; 1.52]	-0.61 [-0.85; -0.38]	-2.52 [-2.91; -2.13]	0.62 [0.36; 0.88]	1.06 [0.77; 1.36]	-0.62 [-0.90; -0.35]	-2.14 [-2.53; -1.74]
fear18	1.60 [1.26; 1.93]	1.32 [0.99; 1.65]	-1.38 [-1.72; -1.05]	-4.29 [-4.99; -3.60]	1.52 [1.15; 1.89]	3.23 [2.64; 3.83]	0.60 [0.25; 0.95]	-1.61 [-2.02; -1.21]
fear19	1.53 [1.18; 1.87]	-0.11 [-0.40; 0.18]	-2.26 [-2.67; -1.85]	-4.85 [-5.70; -4.00]	1.65 [1.27; 2.03]	2.16 [1.70; 2.62]	0.26 [-0.10; 0.62]	-2.17 [-2.64; -1.70]
fear20	1.45 [1.15; 1.75]	2.19 [1.80; 2.57]	-0.03 [-0.31; 0.25]	-2.33 [-2.72; -1.94]	0.88 [0.57; 1.20]	3.79 [3.04; 4.53]	1.88 [1.50; 2.26]	0.23 [-0.06; 0.52]
fear21	1.36 [1.07; 1.65]	1.93 [1.58; 2.29]	-0.20 [-0.47; 0.08]	-2.43 [-2.83; -2.04]	1.04 [0.71; 1.36]	3.37 [2.75; 3.99]	1.72 [1.34; 2.09]	0.21 [-0.09; 0.51]
fear22	0.96 [0.68; 1.23]	-0.19 [-0.44; 0.05]	-2.09 [-2.44; -1.74]	-3.96 [-4.63; -3.28]	1.25 [0.93; 1.57]	2.05 [1.63; 2.47]	-0.01 [-0.32; 0.31]	-2.44 [-2.90; -1.97]
fear23	1.31 [1.01; 1.61]	0.20 [-0.07; 0.48]	-1.52 [-1.85; -1.20]	-3.44 [-3.98; -2.90]	1.07 [0.75; 1.38]	2.55 [2.08; 3.02]	1.08 [0.76; 1.41]	-0.45 [-0.76; -0.15]
fear24	0.87 [0.60; 1.14]	-0.26 [-0.50; -0.01]	-2.08 [-2.42; -1.73]	-3.91 [-4.57; -3.24]	1.00 [0.71; 1.29]	1.17 [0.84; 1.49]	-0.35 [-0.64; -0.05]	-2.05 [-2.45; -1.64]

Table SM8.2 Parameter estimates for the Graded Response Model (GRM) Liebowitz Social Anxiety Scale-Self-Report Fear scale. Bracketed values show the 95% confidence interval.

	Female Sample (n = 390)				Male Sample (n = *221*)			
Item	Discrimination	Threshold 1	Threshold 2	Threshold 3	Discrimination	Threshold 1	Threshold 2	Threshold 3
fear01	1.74 [1.42; 2.07]	-0.08 [-0.38; 0.22]	-2.22 [-2.62; -1.83]	-4.01 [-4.60; -3.42]	1.47 [1.06; 1.89]	-0.32 [-0.69; 0.05]	-2.37 [-2.89; -1.85]	-3.92 [-4.70; -3.14]
fear02	2.04 [1.69; 2.40]	1.69 [1.31; 2.06]	-1.24 [-1.60; -0.88]	-3.97 [-4.55; -3.38]	2.63 [2.03; 3.23]	1.00 [0.48; 1.53]	-1.51 [-2.07; -0.94]	-4.68 [-5.61; -3.75]
fear03	1.43 [1.14; 1.72]	-0.31 [-0.59; -0.04]	-1.60 [-1.93; -1.28]	-3.03 [-3.47; -2.58]	1.46 [1.04; 1.87]	-0.58 [-0.96; -0.20]	-2.24 [-2.74; -1.74]	-4.64 [-5.64; -3.64]
fear04	1.58 [1.23; 1.92]	-1.31 [-1.65; -0.97]	-2.92 [-3.40; -2.45]	-4.64 [-5.38; -3.91]	2.30 [1.63; 2.98]	-2.13 [-2.79; -1.47]	-4.18 [-5.12; -3.23]	-6.45 [-8.03; -4.86]
fear05	1.79 [1.47; 2.10]	3.13 [2.66; 3.59]	0.49 [0.19; 0.80]	-2.24 [-2.63; -1.85]	1.60 [1.21; 1.99]	1.77 [1.32; 2.22]	-0.51 [-0.89; -0.13]	-2.64 [-3.18; -2.10]
fear06	2.25 [1.84; 2.66]	4.78 [4.05; 5.51]	2.38 [1.93; 2.83]	0.07 [-0.28; 0.42]	2.24 [1.73; 2.76]	4.59 [3.65; 5.52]	1.45 [0.95; 1.96]	-0.92 [-1.40; -0.45]
fear07	2.51 [2.08; 2.93]	0.70 [0.32; 1.07]	-1.45 [-1.86; -1.03]	-3.24 [-3.78; -2.70]	2.49 [1.91; 3.06]	0.48 [-0.01; 0.97]	-1.58 [-2.14; -1.03]	-3.63 [-4.38; -2.89]
fear08	2.13 [1.77; 2.49]	1.96 [1.57; 2.35]	-0.63 [-0.98; -0.29]	-2.49 [-2.93; -2.05]	2.11 [1.62; 2.59]	1.43 [0.95; 1.91]	-1.01 [-1.48; -0.55]	-3.62 [-4.33; -2.90]
fear09	1.65 [1.35; 1.96]	0.25 [-0.04; 0.54]	-1.71 [-2.06; -1.36]	-3.44 [-3.95; -2.93]	1.34 [0.97; 1.72]	0.25 [-0.11; 0.60]	-1.72 [-2.15; -1.29]	-3.29 [-3.93; -2.66]
fear10	2.05 [1.70; 2.40]	2.31 [1.90; 2.72]	-0.75 [-1.09; -0.41]	-2.90 [-3.37; -2.43]	1.89 [1.44; 2.34]	1.37 [0.92; 1.82]	-1.27 [-1.73; -0.82]	-3.27 [-3.93; -2.62]
fear11	2.32 [1.93; 2.71]	1.74 [1.34; 2.13]	-1.21 [-1.59; -0.82]	-3.63 [-4.20; -3.07]	2.83 [2.17; 3.49]	1.08 [0.52; 1.64]	-2.18 [-2.84; -1.52]	-4.26 [-5.16; -3.36]
fear12	2.42 [2.02; 2.83]	2.15 [1.72; 2.58]	-0.67 [-1.04; -0.29]	-2.80 [-3.29; -2.31]	2.42 [1.86; 2.98]	1.52 [0.99; 2.04]	-0.96 [-1.46; -0.46]	-2.98 [-3.65; -2.31]
fear13	1.27 [0.96; 1.58]	-1.14 [-1.43; -0.84]	-2.46 [-2.86; -2.06]	-4.18 [-4.84; -3.51]	1.30 [0.87; 1.73]	-1.17 [-1.57; -0.76]	-2.46 [-2.99; -1.93]	-3.43 [-4.12; -2.73]
fear14	2.33 [1.94; 2.72]	1.74 [1.34; 2.14]	-1.44 [-1.84; -1.04]	-3.89 [-4.48; -3.29]	2.24 [1.72; 2.76]	1.15 [0.67; 1.63]	-1.57 [-2.09; -1.05]	-3.83 [-4.60; -3.07]
fear15	2.56 [2.12; 3.00]	4.18 [3.54; 4.81]	1.29 [0.88; 1.69]	-1.44 [-1.86; -1.03]	3.04 [2.36; 3.73]	3.33 [2.55; 4.11]	0.47 [-0.10; 1.03]	-1.70 [-2.32; -1.08]
fear16	2.19 [1.81; 2.57]	3.76 [3.20; 4.32]	1.08 [0.72; 1.44]	-1.03 [-1.39; -0.68]	2.63 [2.04; 3.21]	3.00 [2.32; 3.68]	0.56 [0.05; 1.07]	-1.91 [-2.50; -1.32]
fear17	0.46 [0.26; 0.65]	1.31 [1.06; 1.56]	-0.44 [-0.65; -0.23]	-2.02 [-2.33; -1.71]	0.33 [0.07; 0.58]	0.83 [0.54; 1.13]	-0.85 [-1.14; -0.55]	-2.81 [-3.38; -2.25]
fear18	1.67 [1.37; 1.97]	2.18 [1.81; 2.55]	-0.33 [-0.63; -0.04]	-2.76 [-3.18; -2.33]	1.97 [1.52; 2.43]	2.01 [1.49; 2.52]	-0.91 [-1.35; -0.46]	-3.17 [-3.81; -2.53]
fear19	1.81 [1.49; 2.14]	0.73 [0.42; 1.04]	-1.14 [-1.47; -0.81]	-3.38 [-3.88; -2.87]	2.30 [1.77; 2.84]	1.06 [0.58; 1.55]	-1.34 [-1.84; -0.83]	-4.00 [-4.80; -3.21]
fear20	1.60 [1.30; 1.90]	3.20 [2.73; 3.67]	1.29 [0.98; 1.61]	-0.96 [-1.26; -0.65]	1.74 [1.33; 2.15]	2.62 [2.06; 3.17]	0.18 [-0.21; 0.57]	-1.42 [-1.86; -0.99]
fear21	1.53 [1.24; 1.82]	2.68 [2.28; 3.09]	0.65 [0.36; 0.93]	-1.15 [-1.45; -0.84]	1.86 [1.43; 2.29]	2.74 [2.16; 3.32]	0.79 [0.37; 1.21]	-1.27 [-1.71; -0.82]
fear22	1.63 [1.33; 1.93]	0.65 [0.36; 0.94]	-1.28 [-1.60; -0.96]	-3.51 [-4.02; -3.00]	1.49 [1.10; 1.87]	0.75 [0.37; 1.13]	-1.31 [-1.72; -0.90]	-3.63 [-4.33; -2.93]
fear23	1.86 [1.53; 2.19]	1.35 [1.01; 1.68]	-0.36 [-0.67; -0.05]	-2.03 [-2.41; -1.65]	1.82 [1.39; 2.26]	1.10 [0.68; 1.53]	-0.45 [-0.86; -0.05]	-2.15 [-2.65; -1.64]
fear24	1.25 [0.99; 1.51]	0.37 [0.11; 0.63]	-1.29 [-1.58; -1.00]	-3.16 [-3.62; -2.70]	1.18 [0.83; 1.53]	0.32 [-0.01; 0.66]	-1.37 [-1.76; -0.99]	-2.79 [-3.34; -2.24]

Table SM8.3 Parameter estimates for the Graded Response Model (GRM) Liebowitz Social Anxiety Scale-Self-Report Avoidance scale. Bracketed values show the 95% confidence interval.

	Community-dwelling Sample (n = 356)				SAD-diagnosed Sample (n = 257)			
Item	Discrimination	Threshold 1	Threshold 2	Threshold 3	Discrimination	Threshold 1	Threshold 2	Threshold 3
avoid01	1.04 [0.75; 1.33]	-0.27 [-0.53; -0.01]	-2.62 [-3.04; -2.21]	-4.24 [-4.99; -3.49]	1.07 [0.77; 1.37]	0.97 [0.64; 1.29]	-1.13 [-1.46; -0.80]	-2.51 [-2.97; -2.05]
avoid02	1.76 [1.36; 2.17]	-0.40 [-0.72; -0.08]	-3.62 [-4.23; -3.02]	-5.60 [-6.65; -4.55]	1.62 [1.24; 2.00]	2.62 [2.09; 3.15]	-0.49 [-0.85; -0.14]	-2.54 [-3.03; -2.04]
avoid03	1.21 [0.85; 1.56]	-1.34 [-1.67; -1.01]	-2.83 [-3.30; -2.36]	-4.07 [-4.76; -3.38]	0.56 [0.30; 0.82]	0.59 [0.32; 0.86]	-1.04 [-1.33; -0.75]	-2.17 [-2.57; -1.77]
avoid04	1.19 [0.76; 1.62]	-2.29 [-2.73; -1.85]	-3.56 [-4.18; -2.94]	-4.89 [-5.87; -3.91]	0.53 [0.25; 0.81]	-0.20 [-0.46; 0.07]	-1.45 [-1.77; -1.12]	-2.83 [-3.35; -2.30]
avoid05	1.72 [1.36; 2.09]	0.54 [0.23; 0.86]	-2.46 [-2.90; -2.02]	-5.00 [-5.86; -4.13]	1.34 [1.01; 1.66]	2.38 [1.91; 2.84]	0.17 [-0.16; 0.49]	-1.82 [-2.22; -1.42]
avoid06	1.20 [0.92; 1.47]	1.19 [0.90; 1.48]	-0.75 [-1.03; -0.48]	-2.21 [-2.58; -1.85]	1.35 [0.98; 1.72]	2.73 [2.20; 3.25]	1.62 [1.22; 2.02]	0.20 [-0.12; 0.53]
avoid07	1.50 [1.13; 1.86]	-0.80 [-1.12; -0.49]	-3.06 [-3.56; -2.56]	-4.87 [-5.73; -4.01]	1.39 [1.05; 1.74]	2.19 [1.74; 2.63]	-0.02 [-0.35; 0.31]	-1.53 [-1.91; -1.15]
avoid08	1.67 [1.28; 2.06]	-0.71 [-1.03; -0.38]	-2.79 [-3.26; -2.31]	-4.26 [-4.95; -3.56]	1.17 [0.86; 1.48]	1.52 [1.15; 1.89]	-0.73 [-1.06; -0.41]	-2.22 [-2.64; -1.79]
avoid09	1.57 [1.17; 1.98]	-1.26 [-1.61; -0.91]	-3.10 [-3.63; -2.58]	-4.15 [-4.84; -3.46]	0.83 [0.55; 1.12]	0.49 [0.21; 0.78]	-1.27 [-1.60; -0.95]	-2.46 [-2.91; -2.01]
avoid10	1.71 [1.34; 2.09]	0.29 [-0.02; 0.60]	-2.34 [-2.77; -1.91]	-4.63 [-5.40; -3.87]	1.89 [1.46; 2.32]	2.39 [1.87; 2.91]	-0.04 [-0.42; 0.34]	-1.86 [-2.32; -1.41]
avoid11	2.38 [1.87; 2.89]	-0.03 [-0.40; 0.35]	-3.81 [-4.49; -3.12]	-6.12 [-7.23; -5.00]	2.05 [1.59; 2.51]	2.80 [2.22; 3.38]	0.01 [-0.39; 0.41]	-2.47 [-3.01; -1.93]
avoid12	2.22 [1.75; 2.68]	-0.10 [-0.46; 0.26]	-3.01 [-3.56; -2.47]	-5.24 [-6.12; -4.35]	1.99 [1.54; 2.44]	2.70 [2.13; 3.27]	0.24 [-0.16; 0.63]	-1.42 [-1.85; -0.99]
avoid13	1.03 [0.71; 1.35]	-1.20 [-1.50; -0.90]	-2.75 [-3.19; -2.30]	-3.73 [-4.34; -3.11]	0.54 [0.25; 0.82]	-0.31 [-0.57; -0.05]	-1.69 [-2.04; -1.35]	-2.48 [-2.94; -2.03]
avoid14	1.81 [1.40; 2.23]	-0.73 [-1.07; -0.40]	-3.50 [-4.09; -2.91]	-5.65 [-6.71; -4.60]	1.21 [0.89; 1.53]	1.70 [1.31; 2.09]	-0.77 [-1.10; -0.45]	-2.32 [-2.76; -1.88]
avoid15	1.61 [1.27; 1.94]	1.31 [0.98; 1.65]	-1.23 [-1.56; -0.90]	-3.36 [-3.88; -2.83]	1.82 [1.38; 2.26]	3.81 [3.07; 4.55]	1.64 [1.20; 2.09]	-0.06 [-0.43; 0.31]
avoid16	1.71 [1.36; 2.06]	1.40 [1.05; 1.74]	-1.28 [-1.62; -0.94]	-3.05 [-3.54; -2.57]	1.98 [1.50; 2.45]	3.32 [2.66; 3.99]	1.25 [0.81; 1.68]	-0.23 [-0.62; 0.16]
avoid17	1.14 [0.83; 1.44]	-0.56 [-0.83; -0.29]	-2.39 [-2.78; -2.00]	-3.72 [-4.32; -3.13]	0.51 [0.24; 0.77]	0.07 [-0.19; 0.33]	-1.50 [-1.82; -1.17]	-2.52 [-2.99; -2.06]
avoid18	1.30 [1.00; 1.60]	1.09 [0.79; 1.38]	-1.95 [-2.31; -1.59]	-3.90 [-4.52; -3.27]	1.63 [1.26; 2.01]	2.84 [2.30; 3.39]	0.10 [-0.25; 0.45]	-1.79 [-2.22; -1.37]
avoid19	1.61 [1.24; 1.98]	-0.29 [-0.59; 0.02]	-3.12 [-3.63; -2.60]	-4.88 [-5.72; -4.03]	1.36 [1.03; 1.69]	2.18 [1.73; 2.62]	-0.31 [-0.64; 0.02]	-1.86 [-2.26; -1.45]
avoid20	1.47 [1.16; 1.79]	0.73 [0.44; 1.03]	-1.48 [-1.82; -1.15]	-3.21 [-3.71; -2.71]	1.37 [1.01; 1.72]	2.48 [1.99; 2.97]	0.76 [0.42; 1.10]	-0.62 [-0.95; -0.29]
avoid21	1.04 [0.78; 1.30]	0.71 [0.45; 0.97]	-0.92 [-1.19; -0.65]	-2.06 [-2.40; -1.71]	1.16 [0.84; 1.49]	2.33 [1.88; 2.79]	0.82 [0.50; 1.15]	-0.35 [-0.66; -0.04]
avoid22	0.84 [0.56; 1.12]	-0.68 [-0.93; -0.42]	-2.33 [-2.71; -1.96]	-3.95 [-4.64; -3.26]	1.04 [0.75; 1.34]	1.15 [0.82; 1.49]	-0.81 [-1.13; -0.50]	-2.09 [-2.50; -1.69]
avoid23	1.41 [1.07; 1.76]	-0.81 [-1.12; -0.50]	-2.32 [-2.73; -1.91]	-3.59 [-4.16; -3.02]	1.46 [1.09; 1.82]	1.76 [1.35; 2.18]	0.49 [0.16; 0.83]	-0.78 [-1.13; -0.43]
avoid24	0.95 [0.68; 1.23]	-0.36 [-0.61; -0.11]	-2.01 [-2.35; -1.67]	-3.20 [-3.69; -2.70]	0.79 [0.52; 1.07]	0.72 [0.43; 1.01]	-0.83 [-1.12; -0.54]	-2.10 [-2.50; -1.70]

Table SM8.4 Parameter estimates for the Graded Response Model (GRM) Liebowitz Social Anxiety Scale-Self-Report Avoidance scale. Bracketed values show the 95% confidence interval.

	Female Sample (n = 390)				Male Sample (n = *221*)			
Item	Discrimination	Threshold 1	Threshold 2	Threshold 3	Discrimination	Threshold 1	Threshold 2	Threshold 3
avoid01	1.46 [1.16; 1.75]	0.34 [0.07; 0.62]	-2.04 [-2.40; -1.68]	-3.42 [-3.92; -2.92]	0.90 [0.58; 1.22]	0.15 [-0.16; 0.46]	-1.86 [-2.27; -1.45]	-3.40 [-4.08; -2.72]
avoid02	2.21 [1.82; 2.61]	1.13 [0.77; 1.50]	-2.17 [-2.61; -1.73]	-4.49 [-5.16; -3.82]	2.39 [1.82; 2.95]	0.39 [-0.09; 0.87]	-2.42 [-3.04; -1.79]	-4.05 [-4.87; -3.23]
avoid03	1.27 [0.98; 1.55]	-0.37 [-0.64; -0.11]	-1.96 [-2.30; -1.62]	-3.10 [-3.55; -2.64]	1.02 [0.66; 1.38]	-0.55 [-0.89; -0.22]	-2.11 [-2.56; -1.66]	-3.45 [-4.14; -2.76]
avoid04	1.23 [0.91; 1.55]	-1.22 [-1.53; -0.92]	-2.52 [-2.93; -2.12]	-3.72 [-4.30; -3.15]	1.02 [0.62; 1.43]	-1.31 [-1.70; -0.92]	-2.52 [-3.04; -2.00]	-4.46 [-5.51; -3.41]
avoid05	1.91 [1.57; 2.25]	1.84 [1.47; 2.21]	-1.10 [-1.44; -0.76]	-3.28 [-3.77; -2.78]	1.88 [1.43; 2.33]	0.74 [0.32; 1.17]	-1.48 [-1.94; -1.01]	-3.51 [-4.20; -2.82]
avoid06	1.62 [1.32; 1.93]	2.21 [1.83; 2.58]	0.38 [0.09; 0.67]	-1.03 [-1.33; -0.72]	1.77 [1.34; 2.20]	1.63 [1.18; 2.09]	-0.04 [-0.43; 0.36]	-1.43 [-1.88; -0.98]
avoid07	2.04 [1.67; 2.41]	0.48 [0.16; 0.81]	-1.66 [-2.04; -1.28]	-3.08 [-3.57; -2.60]	2.28 [1.74; 2.82]	0.32 [-0.15; 0.78]	-2.01 [-2.57; -1.44]	-3.86 [-4.64; -3.09]
avoid08	1.66 [1.35; 1.98]	0.36 [0.06; 0.65]	-1.73 [-2.08; -1.37]	-3.21 [-3.68; -2.73]	1.72 [1.29; 2.16]	0.14 [-0.26; 0.54]	-2.14 [-2.65; -1.63]	-3.71 [-4.43; -3.00]
avoid09	1.54 [1.22; 1.86]	-0.47 [-0.76; -0.18]	-2.34 [-2.73; -1.94]	-3.46 [-3.97; -2.94]	1.08 [0.73; 1.43]	-0.34 [-0.68; -0.01]	-2.08 [-2.53; -1.63]	-3.32 [-3.97; -2.67]
avoid10	2.13 [1.75; 2.50]	1.39 [1.02; 1.75]	-1.18 [-1.54; -0.82]	-3.08 [-3.57; -2.59]	1.98 [1.51; 2.44]	0.83 [0.39; 1.27]	-1.67 [-2.16; -1.18]	-3.60 [-4.32; -2.87]
avoid11	2.70 [2.23; 3.18]	1.38 [0.96; 1.81]	-2.07 [-2.56; -1.59]	-4.39 [-5.08; -3.69]	2.73 [2.10; 3.37]	1.06 [0.51; 1.60]	-1.93 [-2.55; -1.31]	-4.54 [-5.49; -3.59]
avoid12	2.64 [2.18; 3.10]	1.44 [1.02; 1.86]	-1.47 [-1.90; -1.04]	-3.16 [-3.70; -2.61]	2.60 [2.00; 3.20]	0.66 [0.15; 1.17]	-1.78 [-2.36; -1.19]	-3.86 [-4.67; -3.06]
avoid13	0.89 [0.63; 1.15]	-0.72 [-0.97; -0.47]	-2.06 [-2.39; -1.73]	-3.05 [-3.50; -2.59]	0.76 [0.42; 1.10]	-0.88 [-1.21; -0.56]	-2.58 [-3.09; -2.06]	-3.18 [-3.83; -2.54]
avoid14	1.85 [1.51; 2.19]	0.47 [0.16; 0.79]	-2.17 [-2.57; -1.77]	-3.98 [-4.56; -3.40]	2.00 [1.51; 2.49]	0.15 [-0.28; 0.58]	-2.28 [-2.83; -1.72]	-3.84 [-4.59; -3.08]
avoid15	1.98 [1.64; 2.33]	2.65 [2.21; 3.08]	0.08 [-0.24; 0.40]	-1.61 [-1.97; -1.25]	2.88 [2.22; 3.54]	2.16 [1.53; 2.80]	-0.41 [-0.95; 0.14]	-2.59 [-3.27; -1.91]
avoid16	1.86 [1.53; 2.19]	2.36 [1.95; 2.76]	0.04 [-0.27; 0.34]	-1.52 [-1.87; -1.17]	3.04 [2.33; 3.74]	2.27 [1.60; 2.94]	-0.91 [-1.50; -0.32]	-2.61 [-3.33; -1.90]
avoid17	0.83 [0.59; 1.07]	-0.19 [-0.41; 0.04]	-1.82 [-2.12; -1.52]	-2.91 [-3.34; -2.49]	0.75 [0.44; 1.07]	-0.38 [-0.68; -0.07]	-2.08 [-2.51; -1.65]	-3.39 [-4.08; -2.69]
avoid18	1.52 [1.23; 1.81]	1.77 [1.43; 2.10]	-0.95 [-1.25; -0.66]	-2.67 [-3.08; -2.26]	2.07 [1.59; 2.54]	2.01 [1.48; 2.53]	-1.38 [-1.87; -0.89]	-3.41 [-4.11; -2.72]
avoid19	1.83 [1.49; 2.16]	0.61 [0.30; 0.92]	-1.84 [-2.22; -1.47]	-3.12 [-3.60; -2.65]	2.40 [1.84; 2.95]	1.14 [0.64; 1.65]	-1.99 [-2.56; -1.41]	-4.39 [-5.28; -3.50]
avoid20	1.65 [1.35; 1.95]	1.62 [1.29; 1.95]	-0.34 [-0.63; -0.04]	-1.83 [-2.18; -1.48]	2.22 [1.71; 2.74]	1.43 [0.94; 1.93]	-0.93 [-1.40; -0.45]	-2.47 [-3.06; -1.88]
avoid21	1.18 [0.93; 1.43]	1.49 [1.19; 1.78]	-0.08 [-0.33; 0.17]	-1.18 [-1.46; -0.91]	1.81 [1.37; 2.25]	1.27 [0.84; 1.71]	-0.43 [-0.83; -0.03]	-1.71 [-2.18; -1.24]
avoid22	1.30 [1.03; 1.58]	0.04 [-0.22; 0.31]	-1.74 [-2.06; -1.41]	-3.13 [-3.58; -2.67]	1.35 [0.98; 1.73]	0.03 [-0.32; 0.38]	-1.83 [-2.28; -1.39]	-3.12 [-3.73; -2.52]
avoid23	1.93 [1.57; 2.29]	0.26 [-0.06; 0.58]	-1.00 [-1.33; -0.66]	-2.26 [-2.66; -1.86]	2.26 [1.72; 2.80]	0.25 [-0.21; 0.72]	-1.47 [-1.98; -0.95]	-2.72 [-3.34; -2.10]
avoid24	0.95 [0.71; 1.19]	0.16 [-0.08; 0.39]	-1.42 [-1.70; -1.15]	-2.64 [-3.03; -2.25]	1.22 [0.86; 1.58]	-0.06 [-0.40; 0.28]	-1.67 [-2.08; -1.25]	-2.95 [-3.53; -2.38]

SM9 Item-fit statistics for the Graded Response Model

Table SM9.1a Orlando and Thissen's (2000, 2003) S-X^2 statistic for the evaluation of item fit in the Liebowitz Social Anxiety Scale-Self-Report, Fear scale

	Community-dwelling sample (n = 356)					SAD-diagnosed sample (n = 257)				
Item	S-X^2	df	RMSEA	p	p-adj	S-X^2	df	RMSEA	p	p-adj
fear01	45.05	29	.039	.029	.519	54.14	57	<.001	.583	.933
fear02	28.97	34	<.001	.713	.855	34.35	44	<.001	.852	.952
fear03	29.20	32	<.001	.609	.830	72.99	67	.019	.288	.687
fear04	14.49	12	.024	.271	.721	38.87	50	<.001	.873	.952
fear05	36.63	44	<.001	.777	.858	58.88	52	.023	.238	.687
fear06	45.18	45	.003	.465	.721	13.08	25	<.001	.976	.976
fear07	24.98	25	<.001	.464	.721	39.16	50	<.001	.866	.952
fear08	48.99	40	.025	.156	.692	59.69	50	.028	.164	.687
fear09	38.00	37	.009	.424	.721	56.81	51	.021	.268	.687
fear10	34.24	44	<.001	.855	.892	53.06	45	.026	.191	.687
fear11	30.74	31	<.001	.479	.721	48.08	45	.016	.349	.698
fear12	54.15	38	.035	.043	.519	43.79	46	<.001	.565	.933
fear13	20.93	31	<.001	.914	.914	48.30	44	.020	.303	.687
fear14	39.39	32	.026	.173	.692	42.24	47	<.001	.670	.952
fear15	37.22	41	<.001	.639	.830	31.05	28	.021	.315	.687
fear16	48.23	46	.012	.383	.721	43.70	38	.024	.242	.687
fear17	77.31	65	.023	.141	.692	53.98	52	.012	.399	.736
fear18	42.69	42	.007	.442	.721	39.79	47	<.001	.763	.952
fear19	36.82	41	<.001	.657	.830	39.56	48	<.001	.802	.952
fear20	48.33	57	<.001	.786	.858	55.75	48	.025	.206	.687
fear21	60.08	56	.014	.330	.721	36.68	52	<.001	.947	.976
fear22	58.44	49	.023	.167	.692	70.26	49	.041	.025	.596
fear23	49.62	47	.013	.369	.721	29.00	36	<.001	.790	.952
fear24	49.82	50	<.001	.481	.721	61.42	50	.030	.129	.687

Note: df: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; p: p-value; p-adj: p-values adjusted for false discovery rate using the Benjamini-Hochberg's (2000) procedure.

Table SM9.1b Orlando and Thissen's (2000, 2003) S-X^2 statistic for the evaluation of item fit in the Liebowitz Social Anxiety Scale-Self-Report, Fear scale

	Female sample (*n* = 390)					Male sample (*n = 221*)				
Item	S-X^2	df	RMSEA	p	p-adj	S-X^2	df	RMSEA	p	p-adj
fear01	64.50	55	.021	.178	.946	31.89	30	.017	.373	.749
fear02	58.44	57	.008	.422	.946	26.99	35	<.001	.832	.907
fear03	51.33	66	<.001	.908	.990	37.53	31	.031	.195	.667
fear04	34.29	47	<.001	.917	.990	17.35	18	<.001	.499	.749
fear05	55.64	61	<.001	.670	.946	52.81	41	.036	.102	.569
fear06	44.50	47	<.001	.577	.946	29.11	29	.004	.459	.749
fear07	38.88	58	<.001	.975	.990	36.57	36	.008	.442	.749
fear08	67.81	63	.014	.317	.946	35.45	36	<.001	.495	.749
fear09	66.15	65	.007	.437	.946	42.95	39	.021	.306	.749
fear10	58.84	61	<.001	.555	.946	41.71	38	.021	.313	.749
fear11	59.53	59	.005	.456	.946	39.32	30	.038	.119	.569
fear12	63.65	61	.011	.383	.946	45.62	31	.046	.044	.527
fear13	44.88	55	<.001	.833	.990	25.64	30	<.001	.693	.867
fear14	64.68	59	.016	.285	.946	24.65	38	<.001	.954	.954
fear15	64.27	51	.026	.100	.946	26.65	36	<.001	.872	.910
fear16	69.06	58	.022	.152	.946	49.54	41	.031	.169	.667
fear17	23.70	42	<.001	.990	.990	80.80	59	.041	.031	.527
fear18	63.25	67	<.001	.607	.946	39.68	38	.014	.395	.749
fear19	60.82	66	<.001	.657	.946	38.91	38	.010	.429	.749
fear20	60.73	64	<.001	.593	.946	37.91	45	<.001	.764	.873
fear21	63.06	74	<.001	.814	.990	42.61	44	<.001	.531	.750
fear22	86.70	65	.029	.037	.897	49.89	38	.038	.094	.569
fear23	70.13	73	<.001	.573	.946	46.46	50	<.001	.616	.822
fear24	66.39	75	<.001	.751	.990	39.01	45	<.001	.723	.867

Note: df: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; p: p-value; p-adj: p-values adjusted for false discovery rate using the Benjamini-Hochberg's (2000) procedure.

Table SM9.1c Orlando and Thissen's (2000, 2003) S-X^2 statistic for the evaluation of item fit in the Liebowitz Social Anxiety Scale-Self-Report, Avoidance scale

	Community-dwelling sample (n = 356)					SAD-diagnosed sample (n = 257)				
Item	S-X^2	df	RMSEA	p	p-adj	S-X^2	df	RMSEA	p	p-adj
avoid01	45.80	39	.022	.211	.791	58.46	54	.018	.315	.524
avoid02	26.83	27	<.001	.473	.797	60.44	42	.041	.032	.389
avoid03	36.66	38	<.001	.532	.797	34.80	44	<.001	.838	.838
avoid04	29.26	24	.025	.211	.791	53.68	55	<.001	.525	.630
avoid05	29.25	40	<.001	.895	.911	49.25	47	.014	.383	.524
avoid06	72.02	66	.016	.286	.791	52.97	44	.028	.167	.433
avoid07	24.39	35	<.001	.911	.911	61.39	50	.030	.130	.433
avoid08	35.80	41	<.001	.701	.911	62.54	52	.028	.150	.433
avoid09	28.65	36	<.001	.803	.911	58.48	55	.016	.349	.524
avoid10	50.80	40	.028	.118	.707	37.81	43	<.001	.695	.759
avoid11	25.06	27	<.001	.571	.806	44.62	41	.019	.322	.524
avoid12	38.62	35	.017	.309	.791	37.20	40	<.001	.597	.682
avoid13	35.78	37	<.001	.526	.797	54.03	52	.012	.397	.524
avoid14	43.30	32	.032	.088	.707	56.74	48	.027	.181	.433
avoid15	51.45	47	.016	.304	.791	72.62	44	.050	.004	.102
avoid16	48.83	47	.010	.399	.791	57.64	44	.035	.081	.433
avoid17	46.05	45	.008	.428	.791	60.06	50	.028	.156	.433
avoid18	32.74	43	<.001	.872	.911	50.62	46	.020	.296	.524
avoid19	35.35	34	.011	.404	.791	61.92	47	.035	.071	.433
avoid20	68.93	52	.030	.058	.707	50.48	49	.011	.415	.524
avoid21	75.93	72	.012	.353	.791	27.63	22	.032	.188	.433
avoid22	40.85	50	<.001	.818	.911	63.64	55	.025	.198	.433
avoid23	37.02	47	<.001	.851	.911	56.02	54	.012	.399	.524
avoid24	65.04	51	.028	.089	.707	43.46	52	<.001	.794	.829

Note: df: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; p: p-value; p-adj: p-values adjusted for false discovery rate using the Benjamini-Hochberg's (2000) procedure.

Table SM9.1d Orlando and Thissen's (2000, 2003) S-X^2 statistic for the evaluation of item fit in the Liebowitz Social Anxiety Scale-Self-Report, Avoidance scale

	Female sample (n = 390)					Male sample (n = 221)				
Item	S-X^2	df	RMSEA	p	p-adj	S-X^2	df	RMSEA	p	p-adj
avoid01	75.59	63	.023	.133	.755	62.88	38	.055	.007	.163
avoid02	49.37	48	.009	.418	.772	28.56	28	.010	.435	.652
avoid03	64.87	65	<.001	.481	.825	54.35	38	.044	.042	.451
avoid04	45.05	55	<.001	.828	.894	25.10	29	<.001	.673	.760
avoid05	54.64	58	<.001	.601	.894	43.98	36	.032	.169	.546
avoid06	67.71	79	<.001	.814	.894	40.75	43	<.001	.569	.706
avoid07	50.29	64	<.001	.894	.894	35.76	32	.023	.296	.546
avoid08	71.31	63	.018	.221	.755	31.49	32	<.001	.492	.695
avoid09	51.39	59	<.001	.749	.894	39.16	38	.012	.418	.652
avoid10	61.69	63	<.001	.523	.837	44.39	33	.040	.089	.451
avoid11	34.89	43	<.001	.806	.894	30.75	27	.025	.282	.546
avoid12	45.07	57	<.001	.873	.894	22.81	25	<.001	.589	.706
avoid13	66.98	62	.014	.310	.755	24.64	29	<.001	.697	.760
avoid14	61.50	55	.017	.255	.755	39.07	32	.032	.182	.546
avoid15	55.51	65	<.001	.793	.894	37.08	33	.024	.286	.546
avoid16	72.87	68	.014	.321	.755	46.46	34	.041	.075	.451
avoid17	82.39	72	.019	.189	.755	47.57	36	.038	.094	.451
avoid18	79.31	70	.018	.209	.755	38.48	34	.024	.274	.546
avoid19	90.46	62	.034	.011	.255	18.78	29	<.001	.927	.927
avoid20	87.65	80	.016	.261	.755	51.49	42	.032	.150	.546
avoid21	99.51	95	.011	.355	.755	51.35	47	.021	.307	.546
avoid22	82.33	66	.025	.084	.755	42.60	39	.020	.319	.546
avoid23	71.66	82	<.001	.786	.894	37.59	39	<.001	.534	.706
avoid24	81.27	78	.010	.378	.755	34.50	42	<.001	.788	.822

Note: df: degrees of freedom; RMSEA: Root Mean Square Error of Approximation; p: p-value; p-adj: p-values adjusted for false discovery rate using the Benjamini-Hochberg's (2000) procedure.

SM10 Details of ROC curve analyses on the Liebowitz Social Anxiety Scale-Self-Report

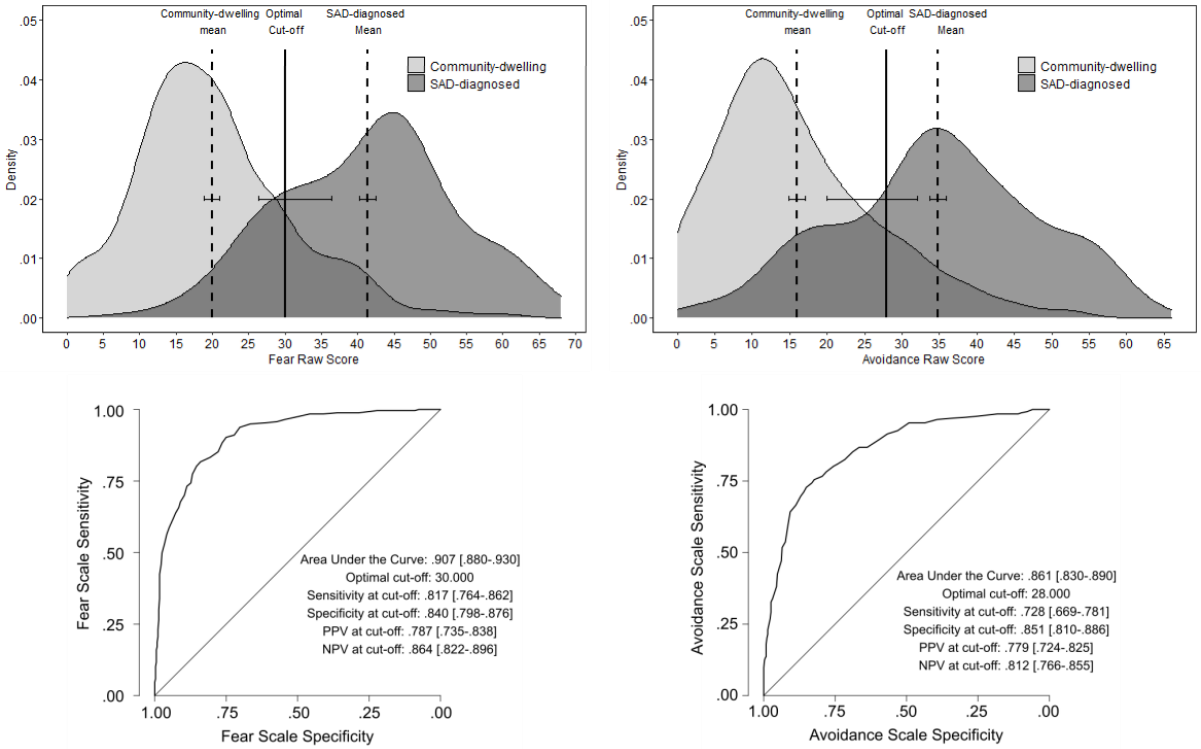


Figure SM10.1 Density plot of total scores (top) and Receiver Operating Characteristic curves (bottom) for the Liewobitz Social Anxiety Scale - Self-Report.

Table SM10.1 Sensitivity (Se), Specificity (Sp), Positive Predictive Value (PPP), Negative Predictive Value (NPV), and their 95% confidence intervals for the Fear scale (1 of 2)

Cut-off	Sp	Se	PPV	NPV
0	.00 [NC; .01]	1.00 [.99; NC]	.42 [NC; NC]	NC [.00; 1.00]
1	.02 [.01; .04]	1.00 [.99; NC]	.42 [.21; NC]	1.00 [.62; 1.00]
2	.02 [.01; .04]	1.00 [.99; NC]	.42 [.24; NC]	1.00 [.68; 1.00]
3	.04 [.02; .07]	1.00 [.99; NC]	.43 [.29; NC]	1.00 [.79; 1.00]
4	.05 [.03; .08]	1.00 [.99; NC]	.43 [.31; NC]	1.00 [.84; 1.00]
5	.06 [.04; .09]	1.00 [.99; NC]	.43 [.32; NC]	1.00 [.85; 1.00]
6	.06 [.04; .10]	1.00 [.99; NC]	.44 [.33; NC]	1.00 [.86; 1.00]
7	.08 [.05; .11]	1.00 [.99; NC]	.44 [.34; NC]	1.00 [.88; 1.00]
8	.09 [.06; .12]	1.00 [.98; 1.00]	.44 [.35; .97]	.97 [.85; .98]
9	.11 [.08; .14]	1.00 [.98; 1.00]	.45 [.36; .97]	.97 [.87; .98]
10	.13 [.10; .17]	1.00 [.98; 1.00]	.45 [.37; .97]	.98 [.89; .98]
11	.16 [.12; .20]	1.00 [.98; 1.00]	.46 [.39; .97]	.98 [.91; .99]
12	.19 [.15; .24]	1.00 [.98; 1.00]	.47 [.40; .97]	.99 [.92; .99]
13	.22 [.18; .27]	1.00 [.98; 1.00]	.48 [.42; .97]	.99 [.93; .99]
14	.29 [.24; .34]	.99 [.97; 1.00]	.50 [.44; .83]	.97 [.92; .98]
15	.32 [.27; .37]	.99 [.97; 1.00]	.51 [.45; .84]	.97 [.93; .98]
16	.36 [.31; .41]	.99 [.97; 1.00]	.53 [.47; .85]	.98 [.94; .98]
17	.41 [.36; .46]	.98 [.96; 1.00]	.55 [.49; .82]	.97 [.93; .98]
18	.46 [.41; .51]	.98 [.96; 1.00]	.57 [.51; .83]	.98 [.94; .98]
19	.49 [.44; .54]	.98 [.95; .99]	.58 [.53; .79]	.97 [.93; .97]
20	.54 [.49; .60]	.96 [.93; .98]	.60 [.55; .77]	.96 [.92; .96]
21	.57 [.52; .63]	.96 [.92; .98]	.62 [.57; .77]	.95 [.91; .96]
22	.61 [.56; .66]	.95 [.92; .98]	.64 [.59; .78]	.95 [.91; .96]
23	.67 [.61; .71]	.95 [.92; .97]	.67 [.62; .80]	.95 [.91; .96]
24	.70 [.65; .75]	.94 [.90; .96]	.69 [.64; .80]	.94 [.90; .95]
25	.72 [.67; .77]	.91 [.87; .94]	.70 [.65; .79]	.92 [.88; .93]
26	.75 [.70; .79]	.90 [.86; .94]	.72 [.67; .80]	.91 [.88; .93]
27	.76 [.72; .81]	.88 [.84; .92]	.73 [.68; .80]	.90 [.86; .92]
28	.78 [.73; .82]	.85 [.80; .89]	.73 [.68; .80]	.88 [.84; .90]
29	.81 [.76; .85]	.83 [.78; .88]	.76 [.70; .82]	.87 [.83; .90]
30	.84 [.80; .88]	.82 [.76; .86]	.79 [.73; .84]	.86 [.82; .90]
31	.85 [.81; .89]	.80 [.75; .85]	.80 [.75; .85]	.86 [.81; .89]
32	.87 [.83; .90]	.77 [.72; .82]	.81 [.76; .85]	.84 [.80; .88]
33	.87 [.83; .91]	.74 [.69; .80]	.81 [.76; .85]	.82 [.78; .87]
34	.89 [.85; .92]	.73 [.67; .78]	.82 [.77; .86]	.82 [.78; .87]
35	.90 [.86; .93]	.70 [.64; .76]	.83 [.78; .87]	.81 [.76; .86]
36	.91 [.88; .94]	.68 [.62; .73]	.84 [.79; .88]	.80 [.75; .85]
37	.92 [.88; .94]	.66 [.60; .72]	.85 [.79; .88]	.79 [.74; .85]
38	.93 [.89; .95]	.64 [.58; .70]	.86 [.81; .89]	.78 [.73; .85]

Table SM10.1 Sensitivity (Se), Specificity (Sp), Positive Predictive Value (PPP), Negative Predictive Value (NPV), and their 95% confidence intervals for the Fear scale (2 of 2)

Cut-off	Sp	Se	PPV	NPV
39	.94 [.90; .96]	.62 [.56; .68]	.87 [.82; .90]	.77 [.72; .84]
40	.95 [.92; .97]	.59 [.52; .65]	.89 [.84; .92]	.76 [.71; .84]
41	.96 [.93; .98]	.56 [.50; .63]	.91 [.85; .93]	.75 [.70; .85]
42	.97 [.94; .98]	.53 [.47; .59]	.92 [.86; .94]	.74 [.69; .85]
43	.97 [.95; .99]	.50 [.44; .56]	.93 [.88; .95]	.73 [.68; .86]
44	.98 [.96; .99]	.46 [.40; .52]	.94 [.88; .95]	.71 [.66; .85]
45	.98 [.96; .99]	.42 [.36; .49]	.95 [.89; .96]	.70 [.65; .87]
46	.98 [.96; .99]	.38 [.32; .44]	.94 [.88; .95]	.69 [.63; .86]
47	.98 [.96; .99]	.34 [.28; .40]	.94 [.87; .95]	.67 [.61; .85]
48	.99 [.97; 1.00]	.30 [.24; .36]	.94 [.87; .95]	.66 [.60; .86]
49	.99 [.97; 1.00]	.26 [.21; .32]	.93 [.85; .95]	.65 [.58; .85]
50	.99 [.97; 1.00]	.23 [.18; .29]	.94 [.85; .95]	.64 [.57; .87]
51	.99 [.97; 1.00]	.21 [.16; .26]	.93 [.84; .95]	.63 [.56; .86]
52	.99 [.98; 1.00]	.19 [.14; .24]	.94 [.84; .96]	.63 [.55; .89]
53	.99 [.98; 1.00]	.17 [.13; .22]	.94 [.83; .95]	.62 [.54; .89]
54	.99 [.98; 1.00]	.16 [.12; .21]	.95 [.85; .97]	.62 [.53; .93]
56	.99 [.98; 1.00]	.13 [.09; .18]	.94 [.82; .96]	.61 [.52; .93]
57	.99 [.98; 1.00]	.12 [.08; .16]	.94 [.80; .96]	.61 [.51; .93]
58	1.00 [.98; 1.00]	.10 [.07; .14]	.96 [.82; .98]	.61 [.50; .98]
59	1.00 [.98; 1.00]	.09 [.06; .14]	.96 [.81; .97]	.60 [.49; .98]
60	1.00 [.98; 1.00]	.09 [.05; .13]	.96 [.80; .97]	.60 [.48; .98]
61	1.00 [.98; 1.00]	.06 [.03; .09]	.94 [.73; .96]	.59 [.45; .98]
62	1.00 [.99; NC]	.05 [.02; .08]	1.00 [.76; 1.00]	.59 [.43; NC]
63	1.00 [.99; NC]	.04 [.02; .07]	1.00 [.71; 1.00]	.59 [.39; NC]
65	1.00 [.99; NC]	.02 [.01; .05]	1.00 [.62; 1.00]	.59 [.34; NC]
66	1.00 [.99; NC]	.02 [.01; .04]	1.00 [.57; 1.00]	.59 [.31; NC]
67	1.00 [.99; NC]	.01 [.00; .03]	1.00 [.45; 1.00]	.58 [.22; NC]
68	1.00 [.99; NC]	.00 [.00; .02]	1.00 [.21; 1.00]	.58 [.03; NC]

Note: NC: not computable

Table SM10.2 Sensitivity (Se), Specificity (Sp), Positive Predictive Value (PPP), Negative Predictive Value (NPV), and their 95% confidence intervals for the Avoidance scale (1 of 2)

Cut-off	Sp	Se	PPV	NPV
0	1.00 [.99; NC]	.00 [NC; .01]	.42 [NC; NC]	NC [.00; 1.00]
1	1.00 [.99; NC]	.04 [.02; .06]	.43 [.28; NC]	1.00 [.78; 1.00]
2	1.00 [.99; NC]	.06 [.04; .09]	.43 [.32; NC]	1.00 [.85; 1.00]
3	.99 [.97; 1.00]	.08 [.05; .11]	.44 [.34; .87]	.93 [.79; .95]
4	.99 [.97; 1.00]	.10 [.07; .13]	.44 [.35; .79]	.92 [.80; .94]
5	.98 [.96; 1.00]	.11 [.08; .15]	.44 [.36; .75]	.91 [.79; .93]
6	.98 [.96; 1.00]	.14 [.11; .18]	.45 [.38; .75]	.93 [.83; .94]
7	.98 [.96; 1.00]	.18 [.14; .23]	.47 [.40; .76]	.94 [.86; .96]
8	.98 [.96; .99]	.22 [.17; .26]	.47 [.41; .74]	.94 [.87; .95]
9	.98 [.95; .99]	.25 [.21; .30]	.48 [.42; .72]	.94 [.87; .95]
10	.97 [.94; .99]	.29 [.25; .34]	.50 [.44; .71]	.94 [.88; .95]
11	.97 [.94; .99]	.35 [.30; .40]	.52 [.46; .72]	.94 [.89; .95]
12	.96 [.93; .98]	.39 [.34; .45]	.53 [.48; .72]	.94 [.89; .95]
13	.95 [.92; .98]	.44 [.38; .49]	.55 [.50; .71]	.93 [.88; .94]
14	.95 [.92; .98]	.49 [.44; .54]	.58 [.52; .73]	.94 [.89; .95]
15	.93 [.89; .95]	.53 [.48; .58]	.59 [.54; .71]	.91 [.86; .92]
16	.91 [.87; .95]	.57 [.51; .62]	.60 [.55; .71]	.90 [.86; .92]
17	.89 [.85; .93]	.60 [.54; .65]	.61 [.56; .71]	.89 [.84; .91]
18	.87 [.82; .91]	.64 [.59; .69]	.63 [.58; .72]	.87 [.82; .89]
19	.87 [.82; .91]	.67 [.61; .71]	.65 [.60; .74]	.87 [.83; .90]
20	.85 [.80; .89]	.69 [.64; .74]	.66 [.61; .74]	.87 [.82; .89]
21	.82 [.77; .87]	.72 [.67; .76]	.68 [.62; .75]	.85 [.80; .88]
22	.81 [.76; .86]	.74 [.69; .79]	.69 [.64; .76]	.84 [.80; .87]
23	.80 [.75; .85]	.76 [.71; .80]	.70 [.65; .77]	.84 [.79; .87]
24	.78 [.73; .83]	.78 [.73; .82]	.72 [.67; .78]	.83 [.79; .87]
25	.77 [.71; .82]	.79 [.75; .84]	.73 [.68; .79]	.83 [.78; .86]
26	.75 [.70; .81]	.82 [.78; .86]	.75 [.70; .81]	.82 [.78; .86]
27	.74 [.69; .80]	.83 [.79; .87]	.76 [.71; .81]	.82 [.77; .86]
28	.73 [.67; .78]	.85 [.81; .89]	.78 [.72; .82]	.81 [.77; .86]
29	.71 [.65; .77]	.86 [.82; .89]	.79 [.73; .83]	.81 [.76; .85]
30	.70 [.64; .75]	.87 [.83; .90]	.80 [.74; .84]	.80 [.75; .85]
31	.67 [.60; .72]	.89 [.85; .92]	.81 [.75; .85]	.79 [.74; .84]
32	.64 [.58; .70]	.91 [.87; .94]	.83 [.78; .87]	.78 [.73; .84]
33	.59 [.53; .65]	.92 [.88; .94]	.84 [.78; .87]	.76 [.71; .82]
34	.54 [.47; .60]	.92 [.89; .95]	.84 [.78; .87]	.73 [.68; .81]
35	.52 [.45; .58]	.94 [.90; .96]	.85 [.79; .88]	.73 [.68; .81]
36	.47 [.41; .53]	.94 [.91; .96]	.85 [.78; .88]	.71 [.66; .80]
37	.44 [.38; .51]	.95 [.92; .97]	.86 [.79; .89]	.70 [.65; .80]
38	.42 [.36; .49]	.95 [.92; .97]	.87 [.80; .89]	.70 [.64; .80]
39	.38 [.32; .44]	.96 [.93; .97]	.86 [.79; .89]	.68 [.62; .79]

Table SM10.2 Sensitivity (Se), Specificity (Sp), Positive Predictive Value (PPP), Negative Predictive Value (NPV), and their 95% confidence intervals for the Avoidance scale (2 of 2)

Cut-off	Sp	Se	PPV	NPV
40	.34 [.28; .40]	.97 [.95; .98]	.89 [.81; .91]	.67 [.61; .80]
41	.33 [.27; .39]	.97 [.95; .99]	.90 [.83; .92]	.67 [.60; .82]
42	.30 [.25; .36]	.97 [.95; .99]	.90 [.82; .92]	.66 [.59; .81]
43	.29 [.24; .35]	.97 [.95; .99]	.89 [.81; .92]	.66 [.59; .81]
44	.27 [.22; .33]	.98 [.96; .99]	.90 [.81; .92]	.65 [.58; .81]
45	.23 [.18; .29]	.99 [.97; 1.00]	.92 [.83; .94]	.64 [.57; .85]
46	.21 [.17; .27]	.99 [.97; 1.00]	.92 [.82; .94]	.63 [.56; .84]
47	.20 [.15; .26]	.99 [.97; 1.00]	.93 [.83; .95]	.63 [.55; .86]
48	.18 [.13; .23]	.99 [.98; 1.00]	.94 [.84; .95]	.63 [.54; .89]
49	.16 [.12; .21]	.99 [.98; 1.00]	.93 [.82; .95]	.62 [.53; .89]
50	.15 [.11; .20]	.99 [.98; 1.00]	.93 [.81; .95]	.62 [.53; .89]
51	.14 [.10; .18]	.99 [.98; 1.00]	.92 [.80; .94]	.61 [.52; .89]
52	.13 [.09; .18]	.99 [.98; 1.00]	.94 [.82; .96]	.61 [.52; .93]
53	.12 [.09; .17]	1.00 [.98; 1.00]	.97 [.85; .98]	.61 [.51; .98]
54	.10 [.06; .14]	1.00 [.99; NC]	1.00 [.87; 1.00]	.61 [.49; NC]
55	.09 [.05; .13]	1.00 [.99; NC]	1.00 [.86; 1.00]	.60 [.48; NC]
56	.06 [.04; .10]	1.00 [.99; NC]	1.00 [.81; 1.00]	.60 [.45; NC]
57	.05 [.02; .08]	1.00 [.99; NC]	1.00 [.76; 1.00]	.59 [.43; NC]
58	.04 [.02; .07]	1.00 [.99; NC]	1.00 [.73; 1.00]	.59 [.41; NC]
59	.02 [.00; .04]	1.00 [.99; NC]	1.00 [.52; 1.00]	.58 [.28; NC]
61	.01 [.00; .03]	1.00 [.99; NC]	1.00 [.45; 1.00]	.58 [.22; NC]
62	.01 [.00; .03]	1.00 [.99; NC]	1.00 [.35; 1.00]	.58 [.14; NC]
66	.00 [.00; .02]	1.00 [.99; NC]	1.00 [.21; 1.00]	.58 [.03; NC]

Note: NC: not computable

SM11 Monte Carlo simulation results (1,000 replications) for estimating achieved power in parameter estimation in the sample of participants diagnosed with SAD (n = 257)

Table SM11.1 Fear scale (1/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
fear01_a	1.13	1.17	0.15	0.16	0.04	0.05	.94
fear01_t1	1.20	1.20	0.17	0.17	0.00	0.03	.96
fear01_t2	-0.81	-0.79	0.15	0.16	0.02	0.08	.97
fear01_t3	-2.62	-2.58	0.24	0.25	0.01	0.04	.96
fear02_a	1.22	1.26	0.16	0.17	0.03	0.02	.95
fear02_t1	3.56	3.48	0.34	0.34	0.02	0.00	.97
fear02_t2	0.64	0.63	0.15	0.17	0.00	0.07	.97
fear02_t3	-2.14	-2.12	0.21	0.22	0.01	0.06	.96
fear03_a	0.69	0.71	0.13	0.13	0.03	0.01	.95
fear03_t1	0.92	0.91	0.14	0.15	0.01	0.03	.96
fear03_t2	-0.48	-0.48	0.14	0.14	0.00	0.04	.96
fear03_t3	-1.94	-1.92	0.18	0.19	0.01	0.05	.97
fear04_a	0.85	0.88	0.14	0.15	0.03	0.03	.95
fear04_t1	0.14	0.14	0.13	0.14	0.03	**0.10**	.97
fear04_t2	-1.64	-1.63	0.17	0.18	0.01	0.07	.97
fear04_t3	-3.35	-3.30	0.33	0.32	0.02	0.01	.96
fear05_a	1.10	1.14	0.16	0.16	0.03	0.03	.95
fear05_t1	3.75	3.68	0.36	0.36	0.02	0.02	.97
fear05_t2	1.26	1.25	0.16	0.18	0.01	0.09	.97
fear05_t3	-0.84	-0.84	0.15	0.16	0.00	**0.11**	.98
fear06_a	1.12	1.16	0.21	0.21	0.03	0.02	.95
fear06_t1	4.90	4.77	0.61	0.61	0.03	0.01	.97
fear06_t2	3.02	3.00	0.36	0.30	0.01	**0.17**	.94
fear06_t3	1.50	1.49	0.20	0.20	0.01	0.01	.95
fear07_a	1.23	1.27	0.16	0.17	0.04	0.04	.95
fear07_t1	3.00	2.97	0.26	0.28	0.01	0.05	.96
fear07_t2	0.84	0.84	0.16	0.17	0.00	0.09	.97
fear07_t3	-1.04	-1.03	0.16	0.17	0.01	0.10	.96
fear08_a	1.20	1.25	0.16	0.17	0.04	0.02	.94
fear08_t1	3.13	3.09	0.28	0.29	0.01	0.03	.97
fear08_t2	0.96	0.96	0.15	0.17	0.01	**0.13**	.97
fear08_t3	-1.02	-1.02	0.15	0.17	0.01	**0.11**	.97
fear09_a	0.97	0.99	0.14	0.15	0.02	0.02	.95
fear09_t1	1.38	1.37	0.16	0.17	0.01	0.06	.97
fear09_t2	-0.41	-0.41	0.14	0.15	0.01	**0.11**	.97
fear09_t3	-2.16	-2.13	0.21	0.21	0.01	0.01	.96
fear10_a	1.84	1.89	0.20	0.21	0.03	0.07	.95
fear10_t1	3.26	3.21	0.30	0.32	0.02	0.06	.97
fear10_t2	0.64	0.63	0.17	0.20	0.01	**0.18**	.98
fear10_t3	-1.60	-1.57	0.20	0.22	0.02	**0.12**	.97
fear11_a	1.90	1.97	0.20	0.22	0.04	0.09	.95
fear11_t1	3.30	3.26	0.30	0.32	0.01	0.07	.97
fear11_t2	0.47	0.47	0.17	0.20	0.00	**0.18**	.98
fear11_t3	-1.91	-1.90	0.22	0.24	0.01	**0.10**	.98

Table SM11.1 Fear scale (2/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
fear12_a	1.97	2.03	0.22	0.23	0.03	0.07	.93
fear12_t1	3.35	3.31	0.31	0.33	0.01	0.07	.96
fear12_t2	1.24	1.23	0.18	0.22	0.00	**0.21**	.98
fear12_t3	-0.92	-0.89	0.17	0.21	0.03	**0.25**	.99
fear13_a	1.28	1.31	0.19	0.19	0.03	0.01	.94
fear13_t1	-0.38	-0.37	0.15	0.17	0.01	**0.15**	.98
fear13_t2	-1.86	-1.85	0.20	0.21	0.01	0.05	.97
fear13_t3	-3.37	-3.32	0.32	0.32	0.02	0.00	.97
fear14_a	1.50	1.56	0.17	0.18	0.04	0.09	.95
fear14_t1	2.92	2.89	0.26	0.28	0.01	0.07	.97
fear14_t2	0.38	0.38	0.15	0.18	0.00	**0.15**	.98
fear14_t3	-2.04	-2.03	0.21	0.23	0.01	0.10	.98
fear15_a	1.33	1.38	0.19	0.20	0.04	0.03	.95
fear15_t1	4.90	4.75	0.59	0.56	0.03	0.06	.96
fear15_t2	3.05	3.03	0.36	0.30	0.01	**0.18**	.95
fear15_t3	0.55	0.55	0.15	0.17	0.00	**0.14**	.97
fear16_a	1.13	1.17	0.17	0.18	0.03	0.08	.95
fear16_t1	4.47	4.34	0.49	0.49	0.03	0.00	.97
fear16_t2	2.27	2.25	0.22	0.23	0.01	0.03	.97
fear16_t3	0.48	0.48	0.15	0.16	0.00	0.08	.97
fear17_a	0.60	0.62	0.13	0.13	0.03	0.03	.95
fear17_t1	1.06	1.06	0.14	0.15	0.00	0.05	.97
fear17_t2	-0.63	-0.62	0.14	0.14	0.02	0.04	.96
fear17_t3	-2.16	-2.13	0.21	0.21	0.01	0.02	.94
fear18_a	1.47	1.52	0.17	0.18	0.03	0.05	.94
fear18_t1	3.28	3.23	0.30	0.31	0.01	0.01	.96
fear18_t2	0.59	0.60	0.15	0.18	0.00	**0.15**	.98
fear18_t3	-1.63	-1.61	0.19	0.21	0.01	0.09	.98
fear19_a	1.61	1.65	0.18	0.19	0.03	0.04	.94
fear19_t1	2.18	2.16	0.21	0.24	0.01	**0.12**	.97
fear19_t2	0.27	0.26	0.15	0.18	0.05	**0.18**	.98
fear19_t3	-2.18	-2.17	0.21	0.24	0.01	**0.13**	.98
fear20_a	0.86	0.88	0.15	0.16	0.03	0.05	.96
fear20_t1	3.86	3.79	0.42	0.40	0.02	0.05	.96
fear20_t2	1.90	1.88	0.19	0.20	0.01	0.02	.96
fear20_t3	0.23	0.23	0.14	0.15	0.01	0.05	.97
fear21_a	1.00	1.04	0.16	0.16	0.03	0.02	.95
fear21_t1	3.44	3.37	0.33	0.33	0.02	0.01	.96
fear21_t2	1.74	1.72	0.18	0.19	0.01	0.05	.97
fear21_t3	0.21	0.21	0.14	0.15	0.03	0.08	.97
fear22_a	1.21	1.25	0.15	0.16	0.03	0.06	.95
fear22_t1	2.08	2.05	0.20	0.21	0.01	0.06	.97
fear22_t2	-0.01	-0.01	0.14	0.16	0.07	**0.11**	.98
fear22_t3	-2.47	-2.44	0.23	0.24	0.02	0.02	.96
fear23_a	1.03	1.07	0.15	0.16	0.03	0.01	.94
fear23_t1	2.58	2.55	0.23	0.24	0.01	0.04	.96
fear23_t2	1.09	1.08	0.16	0.17	0.01	0.05	.97

Table SM11.1 Fear scale (3/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
fear23_t3	-0.46	-0.45	0.14	0.15	0.01	**0.11**	.97
fear24_a	0.97	1.00	0.14	0.15	0.03	0.09	.96
fear24_t1	1.18	1.17	0.15	0.17	0.01	0.08	.97
fear24_t2	-0.35	-0.35	0.14	0.15	0.00	0.09	.97
fear24_t3	-2.07	-2.04	0.19	0.21	0.01	0.09	.97

Note: _a: discrimination parameter; _t1, _t2, _t3: threshold parameters; bolded values indicate bias estimates larger than . 10;

The column Average gives the parameter estimate averages over the replications of the Monte Carlo study. The column labeled Starting gives the parameter values obtained in the sample at hand, that are considered as population values. Parameter bias is obtained by subtracting the Starting value from the Average value and by dividing it by the Starting value (Muthén & Muthén, 2002, p. 606).

The column labeled Std. Dev. gives the standard deviation of each parameter estimate over the replications of the Monte Carlo study. This is considered to be the population standard error when the number of replications is large. The column labeled S.E. Average gives the average of the estimated standard errors for each parameter estimate over the replications of the Monte Carlo study. Standard error bias is obtained by subtracting the Std. Dev.value from the S.E. Average value and by dividing it by the Std. Dev.value (Muthén & Muthén, 2002, p. 606).

Coverage is the proportion of replications for which the 95% confidence interval # contains the Starting parameter value (Muthén & Muthén, 2002, p. 606).

Table SM11.2 Avoidance scale (1/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
avo01_a	1.13	1.17	0.15	0.16	0.04	0.05	.94
avo01_t1	1.20	1.20	0.17	0.17	0.00	0.03	.96
avo01_t2	-0.81	-0.79	0.15	0.16	0.02	0.08	.97
avo01_t3	-2.62	-2.58	0.24	0.25	0.01	0.04	.96
avo02_a	1.22	1.26	0.16	0.17	0.03	0.02	.95
avo02_t1	3.56	3.48	0.34	0.34	0.02	0.00	.97
avo02_t2	0.64	0.63	0.15	0.17	0.00	0.07	.97
avo02_t3	-2.14	-2.12	0.21	0.22	0.01	0.06	.96
avo03_a	0.69	0.71	0.13	0.13	0.03	0.01	.95
avo03_t1	0.92	0.91	0.14	0.15	0.01	0.03	.96
avo03_t2	-0.48	-0.48	0.14	0.14	0.00	0.04	.96
avo03_t3	-1.94	-1.92	0.18	0.19	0.01	0.05	.97
avo04_a	0.85	0.88	0.15	0.15	0.03	0.03	.95
avo04_t1	0.14	0.14	0.13	0.14	0.03	**0.10**	.97
avo04_t2	-1.64	-1.63	0.17	0.18	0.01	0.07	.97
avo04_t3	-3.35	-3.30	0.33	0.32	0.02	0.01	.96
avo05_a	1.10	1.14	0.16	0.16	0.03	0.03	.95
avo05_t1	3.75	3.68	0.36	0.36	0.02	0.02	.97
avo05_t2	1.26	1.25	0.16	0.18	0.01	0.09	.97
avo05_t3	-0.84	-0.84	0.15	0.16	0.00	**0.11**	.98
avo06_a	1.12	1.16	0.21	0.21	0.03	0.02	.95
avo06_t1	4.90	4.77	0.61	0.61	0.03	0.01	.97
avo06_t2	3.02	3.00	0.36	0.30	0.01	**0.17**	.94
avo06_t3	1.50	1.49	0.20	0.20	0.01	0.01	.95
avo07_a	1.23	1.27	0.16	0.17	0.04	0.04	.95
avo07_t1	3.00	2.97	0.26	0.28	0.01	0.05	.96
avo07_t2	0.84	0.84	0.16	0.17	0.00	0.09	.97
avo07_t3	-1.04	-1.03	0.16	0.17	0.01	0.10	.96
avo08_a	1.20	1.25	0.16	0.17	0.04	0.02	.94
avo08_t1	3.13	3.09	0.28	0.29	0.01	0.03	.97
avo08_t2	0.96	0.96	0.15	0.17	0.01	**0.13**	.97
avo08_t3	-1.02	-1.02	0.15	0.17	0.01	**0.11**	.97
avo09_a	0.97	0.99	0.14	0.15	0.02	0.02	.95
avo09_t1	1.38	1.37	0.16	0.17	0.01	0.06	.97
avo09_t2	-0.41	-0.41	0.14	0.15	0.01	**0.11**	.97
avo09_t3	-2.16	-2.13	0.21	0.21	0.01	0.01	.96
avo10_a	1.84	1.89	0.20	0.21	0.03	0.07	.95
avo10_t1	3.26	3.21	0.30	0.32	0.02	0.06	.97
avo10_t2	0.64	0.63	0.17	0.20	0.01	**0.18**	.98
avo10_t3	-1.60	-1.57	0.20	0.22	0.02	**0.12**	.97
avo11_a	1.90	1.97	0.20	0.22	0.03	0.09	.95
avo11_t1	3.30	3.26	0.30	0.32	0.01	0.07	.97
avo11_t2	0.47	0.47	0.17	0.20	0.01	**0.18**	.98
avo11_t3	-1.91	-1.90	0.22	0.24	0.01	**0.10**	.98

Table SM11.1 Avoidance scale (2/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
avo12_a	1.97	2.03	0.22	0.23	0.03	0.07	.93
avo12_t1	3.35	3.31	0.31	0.33	0.01	0.07	.96
avo12_t2	1.24	1.23	0.18	0.22	0.00	**0.21**	.98
avo12_t3	-0.92	-0.89	0.17	0.21	0.03	**0.25**	.99
avo13_a	1.28	1.31	0.19	0.19	0.03	0.01	.94
avo13_t1	-0.38	-0.37	0.15	0.17	0.01	**0.15**	.98
avo13_t2	-1.86	-1.85	0.20	0.21	0.01	0.05	.97
avo13_t3	-3.37	-3.32	0.32	0.32	0.02	0.00	.96
avo14_a	1.50	1.56	0.17	0.18	0.04	0.09	.95
avo14_t1	2.92	2.89	0.26	0.28	0.01	0.07	.97
avo14_t2	0.38	0.38	0.15	0.18	0.00	**0.15**	.98
avo14_t3	-2.04	-2.03	0.21	0.23	0.01	0.10	.98
avo15_a	1.33	1.38	0.19	0.20	0.04	0.03	.95
avo15_t1	4.90	4.75	0.59	0.56	0.03	0.06	.96
avo15_t2	3.05	3.03	0.36	0.30	0.01	**0.18**	.95
avo15_t3	0.55	0.55	0.15	0.17	0.00	**0.14**	.97
avo16_a	1.13	1.17	0.17	0.18	0.03	0.08	.95
avo16_t1	4.47	4.34	0.49	0.49	0.03	0.00	.97
avo16_t2	2.27	2.25	0.22	0.23	0.01	0.04	.97
avo16_t3	0.48	0.48	0.15	0.16	0.00	0.08	.97
avo17_a	0.60	0.62	0.13	0.13	0.03	0.03	.95
avo17_t1	1.06	1.06	0.14	0.15	0.00	0.05	.97
avo17_t2	-0.63	-0.62	0.14	0.14	0.02	0.04	.96
avo17_t3	-2.16	-2.13	0.21	0.21	0.01	0.02	.94
avo18_a	1.47	1.52	0.17	0.18	0.03	0.05	.94
avo18_t1	3.28	3.23	0.30	0.31	0.01	0.01	.96
avo18_t2	0.59	0.60	0.15	0.18	0.00	**0.15**	.98
avo18_t3	-1.63	-1.61	0.19	0.21	0.01	0.09	.98
avo19_a	1.60	1.65	0.18	0.19	0.03	0.04	.94
avo19_t1	2.18	2.16	0.21	0.24	0.01	**0.12**	.97
avo19_t2	0.27	0.26	0.15	0.18	0.05	**0.18**	.98
avo19_t3	-2.18	-2.17	0.21	0.24	0.01	**0.13**	.98
avo20_a	0.86	0.88	0.15	0.16	0.03	0.05	.96
avo20_t1	3.86	3.79	0.42	0.40	0.02	0.05	.96
avo20_t2	1.90	1.88	0.19	0.20	0.01	0.02	.96
avo20_t3	0.23	0.23	0.14	0.15	0.01	0.05	.97
avo21_a	1.00	1.04	0.16	0.16	0.03	0.02	.95
avo21_t1	3.44	3.37	0.33	0.33	0.02	0.01	.96
avo21_t2	1.74	1.72	0.18	0.19	0.01	0.06	.97
avo21_t3	0.21	0.21	0.14	0.15	0.03	0.08	.97
avo22_a	1.21	1.25	0.15	0.16	0.03	0.06	.95
avo22_t1	2.08	2.05	0.20	0.21	0.01	0.06	.97
avo22_t2	-0.01	-0.01	0.14	0.16	**0.11**	**0.11**	.98
avo22_t3	-2.47	-2.44	0.23	0.24	0.02	0.02	.96
avo23_a	1.03	1.07	0.15	0.16	0.03	0.01	.94
avo23_t1	2.58	2.55	0.23	0.24	0.01	0.04	.96
avo23_t2	1.09	1.08	0.16	0.17	0.01	0.05	.97

Table SM11.1 Avoidance scale (3/3)

Item	Average	Starting	Std. Dev.	S.E. Average	Parameter bias	Standard error bias	Coverage
avo23_t3	-0.46	-0.45	0.14	0.15	0.00	**0.11**	.97
avo24_a	0.97	1.00	0.14	0.15	0.03	0.09	.96
avo24_t1	1.18	1.17	0.15	0.17	0.01	0.08	.97
avo24_t2	-0.35	-0.35	0.14	0.15	0.00	0.09	.97
avo24_t3	-2.07	-2.04	0.19	0.21	0.01	0.08	.97

Note: _a: discrimination parameter; _t1, _t2, _t3: threshold parameters; bolded values indicate bias estimates larger than . 10;

The column Average gives the parameter estimate averages over the replications of the Monte Carlo study. The column labeled Starting gives the parameter values obtained in the sample at hand, that are considered as population values. Parameter bias is obtained by subtracting the Starting value from the Average value and by dividing it by the Starting value (Muthén & Muthén, 2002, p. 606).

The column labeled Std. Dev. gives the standard deviation of each parameter estimate over the replications of the Monte Carlo study. This is considered to be the population standard error when the number of replications is large. The column labeled S.E. Average gives the average of the estimated standard errors for each parameter estimate over the replications of the Monte Carlo study. Standard error bias is obtained by subtracting the Std. Dev.value from the S.E. Average value and by dividing it by the Std. Dev.value (Muthén & Muthén, 2002, p. 606).

Coverage is the proportion of replications for which the 95% confidence interval # contains the Starting parameter value (Muthén & Muthén, 2002, p. 606).

References

Assunção, M. C., da Conceição Costa, D. L., de Mathis, M. A., Gedanke Shavitt, R., Ferrão, Y. A., do Rosário, M. C., Miguel, E. C., & Rodrigues Torres A. (2012). Social phobia in obsessive–compulsive disorder: Prevalence and correlates. *Journal of Affective Disorders*, *143*(1–3), 138–147. https://doi.org/10.1016/j.jad.2012.05.044.

Baldwin, D. S., Brandish, E. K., & Meron, D. (2008). The overlap of obsessive-compulsive disorder and social phobia and its treatment. *CNS Spectrums*, *13*(9 SUPPL. 14), 47–53. https://doi.org/10.1017/s1092852900026936.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897. https://doi.org/10.1037//0022-006x.56.6.893.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83. https://doi.org/10.3102/10769986025001060.

Borkovec, T. D., Robinson, E., Pruzinsky, T., & DePree, J. A. (1983). Preliminary exploration of worry: Some characteristics and processes. *Behaviour Research and Therapy*, *21*(1), 9–16. https://doi.org/10.1016/0005-7967(83)90121-3.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509–540. http://doi.org/10.1207/s15327906mbr2704_2.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456.

Carpita, B., Muti, D., Petrucci, A., Romeo, F., Gesi, C., Marazziti, D., Carmassi, C., & Dell'Osso, L. (2019). Overlapping features between social anxiety and obsessive-compulsive spectrum in a clinical sample and in healthy controls: Toward an integrative model. *CNS Spectrums*, *25*(4), 527–534. https://doi.org/10.1017/S109285291900138X.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In *Social Phobia: Diagnosis, Assessment and Treatment* (pp. 69–93). Guilford Press.

Connor, K. M., Davidson, J. R. T., Erik Churchill, L., Sherwood, A., Foa, E., & Weisler, R. H. (2000). Psychometric properties of the social phobia inventory (SPIN). New self-rating scale. *British Journal of Psychiatry*, *176*, 379–386. https://doi.org/10.1192/bjp.176.4.379.

Ghisi, M., Flebus, C.B., Montano, A., Sanavio, E., & Sica, C. (2006). *Beck Depression Inventory II. Manuale italiano.* Giunti-O.S.

Gori, A., Giannini, M., Socci, S., Luca, M., Dewey, D., Schuldberg, D., & Craparo, G. (2013). Assessing Social Anxiety Disorder: Psychometric properties of the Italian Social Phobia Inventory (I-SPIN). *Clinical Neuropsychiatry*, *10*(1), 37–42. https://psycnet.apa.org/record/2013-15314-005.

Gustafsson, J. E., & Aberg-Bengtsson, L. (2010). Unidimensionality and the interpretability of psychological instruments. In S. E. Embretson (Ed.), *Measuring psychological constructs* (pp. 97–121). Washington, DC: American Psychological Association.

Hearn, C. S., Donovan, C. L., Spence, S. H., & March, S. (2017). A worrying trend in Social Anxiety: To what degree are worry and its cognitive factors associated with youth Social Anxiety Disorder? *Journal of Affective Disorders*, *208*, 33–40. https://doi.org/10.1016/j.jad.2016.09.052.

Hodgson, R. J., & Rachman, S. (1977). Obsessional-compulsive complaints. *Behaviour Research and Therapy*, *15*(5), 389–395. https://doi.org/10.1016/0005-7967(77)90042-0

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447.

Huang, C., & Chen, J. H. (2015). Meta-analysis of the factor structures of the Beck Depression Inventory–II. *Assessment*, *22*(4), 459–472. https://doi.org/10.1177/1073191114548873.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*(4), 537–549. https://doi.org/10.1007/s11336-011-9218-4.

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*, *77*(3), 442–454. https://doi.org/10.1007/s11336-012-9269-1.

Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. *Multivariate Behavioral Research*, *24*(1), 59–69. http://doi.org/10.1207/s15327906mbr2401_4.

Lydiard, R. B. (2001). Social anxiety disorder: comorbidity and its implications. *Journal of Clinical Psychiatry*, *62*(Supp. 1), 17–24.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30.

Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, *97*(5), 1016–1031. https://doi.org/10.1037/a0027934.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6.

Morani, S., Pricci, D., & Sanavio, E. (1999). Penn State Worry Questionnaire and Worry Domains Questionnaire: Italian versions and reliability. *Psicoterapia Cognitiva e Comportamentale*, *5*, 195–209.

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50–64. https://doi.org/10.1177/01466216000241003.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*(4), 289–298. https://doi.org/10.1177/0146621603027004004.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555.

Reise, S. P., Cook, K. F., & Moore, T. M. (2015). Evaluating the impact of multidimensionality on unidimensional item response theory model parameters. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling. Applications to typical performance assessment* (pp. 13–40). New York, NY: Routledge.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477.

Revelle, W. (2015). psych: Procedures for Psychological, Psychometric, and Personality Research. In *R package version 1.5.4*. http://cran.r-project.org/package=psych

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249.

Ruscio, A. M., Brown, T. A., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2008). Social fears and social phobia in the USA: Results from the National Comorbidity Survey Replication. *Psychological Medicine*, *38*(1), 15–28. https://doi.org/10.1017/S0033291707001699.

Sanavio, E., & Vidotto, G. (1985). The components of the Maudsley Obsessional-Compulsive Questionnaire. *Behaviour Research and Therapy*, *23*(6), 659–662. https://doi.org/10.1016/0005-7967(85)90061-0.

Sica, C., Coradeschi, D., Ghisi, M., & Sanavio, E. (2006). *Beck Anxiety Inventory* - BAI. Giunti-O.S.

Sica, C., Musoni, I., Chiri, L. R., Bisi, B., Lolli, V., & Sighinolfi, C. (2007). Social Phobia Scale (SPS) and Social Interaction Anxiety Scale (SIAS): Italian translation and adaptation. *Bollettino di psicologia applicata*, *252*, 59–71.

Starcevic, V., Berle, D., Milicevic, D., Hannan, A., Lamplugh, C., & Eslick, G. D. (2007). Pathological worry, anxiety disorders and the impact of co-occurrence with depressive and other anxiety disorders. *Journal of Anxiety Disorders*, *21*(8), 1016–1027. https://doi.org/10.1016/j.janxdis.2006.10.015.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, *15*(3), 335–352. https://doi.org/10.1207/s15327906mbr1503_7.

Stucky, B. D., & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183–206). New York, NY: Routledge/Taylor & Francis Group.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147–172). Hillsdale, NJ: Erlbaum.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Erlbaum.

Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. Journal of Personality and Social Psychology, 84(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608.