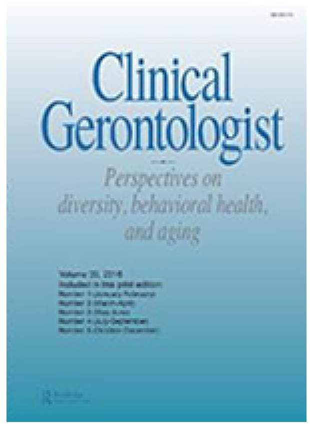


Development and Validation of a Short Form of the Geriatric Anxiety Scale (GAS-12) among Italian Older Adults

Journal:	*Clinical Gerontologist*
Manuscript ID	WCLI-2022-0044.R4
Manuscript Type:	Research Article
Keywords:	Geriatric Anxiety Scale, short form, late-life anxiety, anxiety < psychiatric disorders, aging, older adult, health assessment < psychological assessment, psychological assessment

Abstract

Objectives: We developed a new Italian short version of the Geriatric Anxiety Scale (GAS-12) and evaluated its psychometric properties. The GAS-12 specifically screens for anxiety symptoms in the Italian older adult population by identifying items that best discriminate anxiety in this population.

Methods: In Study 1, we administered the full-length Italian translation of the GAS to 517 older adults and used item response theory to identify the most discriminating items and to develop the short form used in Study 2. In Study 2, we evaluated the functioning of the new short form of the questionnaire in a new sample of 427 older adults using Confirmatory factor analysis.

Results: Analyses indicated 12 items that discriminated well between anxious and non-anxious participants and distributed along the latent continuum of each trait. The GAS-12 fits a three-factor structure. There was also evidence for convergent and divergent validity.

Conclusions: The Italian GAS-12 appears to be a useful instrument for the quantitative screening of anxiety in Italian older adults.

Clinical Implications: Anxiety imposes significant impairment thus making imperative the screening and assessment of anxiety symptoms. The GAS-12 is particularly indicated with limited time and many scales in a clinical assessment or research protocols.

Keywords: Geriatric Anxiety Scale; short form; item response theory; late-life anxiety; factor structure; measurement invariance; gender differences

Introduction

Many studies have investigated variables associated with successful aging (Vaillant and Mukamal, 2001; Dahany et al., 2014). Amongst these, anxiety seems to be a crucial factor mediating successful aging and is often associated with cognitive impairment and poorer life satisfactions, perceived health and quality of life (QoL) (e.g., de Beurs et al., 1999). In addition, anxiety is widespread among older adults with general prevalence estimates ranging from 3% to 14% (Wolitzky-Taylor et al., 2010; Baxter et al., 2013). More specifically, 7% of Italian older adults show symptoms of chronic anxiety (Istituto Nazionale di Statistica, 2018).

Nonetheless, anxiety is commonly underdiagnosed (Segal et al., 2018) and, more importantly, most screening measures were originally developed and validated in adult and younger adult samples, therefore lacking specific norms and psychometric evidence for use with older adults. Current measures of anxiety developed specifically for older adults include the Geriatric Anxiety Inventory (Pachana et al., 2017), the Adult Manifest Anxiety Scale-Elderly Version (Reynolds et al., 2003), the Anxiety in Cognitive Impairment and Dementia (ACID; Gerolimatos et al., 2015) and the Geriatric Anxiety Scale (GAS; Segal et al., 2010).

Amongst these measures, the GAS demonstrated excellent internal consistency of scale scores (α = 0.88–0.93) in a non-clinical population and significant convergent and discriminant validity (Balsamo et al., 2018; Segal et al., 2010; Yochim et al., 2013). The GAS includes 25 scorable items along three conceptually different subscales, including somatic, cognitive, and affective subscale scores (Segal et al., 2018). Importantly, somatic items were constructed to balance the importance of somatic symptoms without over-emphasizing somatic content of anxiety (Gottschling et al., 2016). The GAS has been translated in many languages including German, Persian, Arabic, Turkish, Chinese and

Italian (Bolghan-Abadi et al., 2013; Gottschling et al., 2016; Lin et al., 2017; Gatti et al., 2018) and shows good psychometric properties among Italian community-dwelling older adults (Picconi et al., 2018). Regarding the analysis of the Italian GAS factor structure, CFA confirmed the better fit for the three factors (Cognitive, Somatic, and Affective) from the original English version (Segal et al., 2010; Yochim et al., 2011, 2013) and results suggested good internal consistency (α range = 0.75-0.88) for the Italian GAS total and subscale scores. Cronbach's alpha values were comparable to the values of the original English version (Segal et al., 2010) and did not differ significantly except for the GAS total score (Feldt test = 0.5833, $p < 0.001$; Feldt et al., 1987) and Cognitive scale (Feldt test = 0.4167, $p < 0.001$) in which the original sample scored higher reliabilities values. Similar results were found when comparing the alpha values of the Italian version of the GAS with German version. Cronbach's alpha values not differ significantly (all p = ns; Feldt et al., 1987). Convergent validity of the Italian GAS was evidenced via significant and high correlations between the GAS total score, subscale scores and another measure of anxiety (GAI) (r range = 0.82-0.97). With respect to the discriminant validity of the Italian GAS, the relation between the GAS total score, subscale scores and depression (TDI) was lower than the correlation with anxiety measure (r range = 0.39-0.48) (Picconi et al., 2018).

Nonetheless, short forms are preferred in settings where time is limited and many scales are administered (Mueller et al., 2015) and developing brief assessment tools that reduce fatigue and facilitate individuals with cognitive problems are crucial (Gottschling et al., 2020; Mueller et al., 2015).

Mueller et al. (2015) developed a 10-item English version of the GAS (GAS-10) using the item response theory framework. Unidimensionality was reported. The GAS-10 had excellent internal consistency of scale scores (Cronbach's α = 0.89), and significantly and positively correlated with the GAS total scale (r = 0.96, p < 0.001) and subscales (Cognitive: r = 0.92, p < 0.001, Affective: r = 0.89, p < 0.001, Somatic: r = 0.82, p < 0.001). The authors combined groups for some analyses but acknowledged the possibility that differences in the data collection strategies may have impacted results and generalizability. More recently, Pifer et al. (2020) developed an adapted 10-item version of the GAS (GAS-LTC), specifically tailored for long-term care settings. They slightly modified items according to specific cognitive needs of long-term care residents and changed the response format from a Likert-type scale to a more simple dichotomous Yes/No response. Even though the initial validation was conducted in a small sample (N = 66), the scale (GAS-LTC total scale) showed good internal consistency and convergent validity. Gottschling et al. (2020) developed a German short version of the GAS (GAS-G-SF). Results confirmed that the GAS-G-SF is a good alternative for assessing anxiety in community and clinical populations in Germany. However, in the sample of community-dwelling older adults, the Somatic subscale was at the lower end of acceptability (McDonald's ω value = 0.64). The Cognitive and Affective subscales yielded acceptable omegas (ω =.71; ω = .81) and the total scale had high internal consistency (ω = .88). In the clinical sample, internal consistency was overall lower than the sample of community-dwelling older adults (Cognitive ω = 0.65 and Affective ω = 0.66). The total scale had high internal consistency (ω = .81), but the internal consistency of the Somatic subscale was further reduced in an unacceptable range (ω = 0.49). The substantial correlations between the GAS-G-SF with depression (r = .53), indicates a significant overlap between both constructs and highlights the fact that anxiety and depression often co-occur in older adults (Segal, June et al., 2010; Segal, Qualls & Smyer, 2010). Also, most items of the GAS-G-SF were skewed and not normally distributed in the validation sample, whereas most items were normally distributed in the clinical sample. Unlike the GAS-10, the GAS-G-SF is a multidimensional measure of geriatric anxiety and fits a three-factor model (Gottschling et. al., 2020).

In Study 1, we developed a new short form of the GAS for Italian older adults. In particular, we started from the Italian version of the GAS (Gatti et al., 2018; Picconi et al., 2018) and not from a translation of the English short version (GAS-10, Mueller et al., 2015) in order to select the best items for the Italian population and to control for cultural differences in anxiety symptoms (Friedman, 2002). Indeed, anxiety is a complex construct and its expression appears to differ across cultures and a series of studies has, in fact, confirmed cross cultural variations in the prevalence and presentation of anxiety disorders (Marques et al., 2011). In its simplest sense, anxiety is a descriptive label for how one feels and how one feels is critically influenced by ethnic, racial, and culturally dependent variations in beliefs about social context and norms (Hinton, 2012; Hofmann et al. 2010; Hofmann & Hinton, 2014). This may be even more evident in older adults when social context and norms are considered together with negative beliefs about aging. In Study 2, the new short form was tested in an independent sample to investigate factor structure using CFA and to assess internal consistency and convergent and discriminant validity.

Study 1. Developing the Geriatric Anxiety Scale (GAS-12)

We used item response theory (IRT; Embretson and Reise, 2000) to obtain a short Italian version of the GAS (Picconi et al., 2018; Segal et al., 2019) since IRT has been found to provide more sophisticated information and allows for improvement of the reliability of the scale compared to classical test theory (Petrillo et al., 2015). Specifically, IRT allows us to evaluate the amount of information provided by each item of the scale through the item information function (IIF). If the amount of information is large, the trait level can be estimated with precision. If the amount of information is small, the trait cannot be accurately estimated.

We evaluated the discrimination parameters and information curve peaks of each item from each subscale so that the GAS, although shorter, still captured Cognitive, Affective, and Somatic components of anxiety. Additionally, we assessed the measurement precision of the scale through the test information function (TIF) that evaluates the precision of the test at different levels of the measured construct (Embretson and Reise, 2000).

Methods

Participants and procedure

We collected data from 517 independent community-dwelling Italian older adults (49.5% females; mean age: 72.07; SD = 6.68 years), recruited from the general population through a snowball sampling procedure. All participants had normal or corrected to normal vision and no reports of severe head trauma, stroke, psychiatric or neurological disorders nor use of psychiatric drugs. Before completing the questionnaire packet, we screened all participants with the Mini Mental State Examination to exclude participants with cognitive impairment. Participants received no monetary reimbursement for participation and anonymously completed the questionnaire packet. The study was approved by the Departmental ethical committee at the University of Chieti and all participants provided written informed consent prior to inclusion.

Measures

Mini Mental State Examination (MMSE)

The Mini Mental State Examination (MMSE) is a simple pen and paper test of a 30 items designed to screen for cognitive impairment in older adults. Specifically, it examines spatial and temporal orientation, attention and calculation, recall, language, and executive functions. Scores range from 0 to 30 (MMSE; cut-off = 24; M = 27.71, SD = 1.3; Folstein et al., 1975). Advantages of the MMSE include rapid administration, availability of multiple language translations and high levels of acceptance as a screening instrument amongst health professionals and researchers (Nieuwenhuis-Mark, 2010). All participants scored above cut-off (M = 27.06, SD = 2.51).

Geriatric Anxiety Scale

The full-length Italian Geriatric Anxiety Scale (GAS; Picconi et al., 2018) is a 30-item self-report measure used to assess anxiety symptoms in older adults. It includes three theoretically derived subscales: Cognitive (8 items; e.g., 'I felt like I had no control over my life'), Somatic (9 items; e.g., 'I felt tired') and Affective (8 items; e.g., 'I felt restless, keyed up, or on edge'). Scores range from 0 to 75. Individuals indicate how often they have experienced each symptom during the immediately preceding week along a 4-point Likert scale ranging from 0 (not at all) to 3 (always), with higher scores indicating higher levels of anxiety. The GAS total score is based on the first 25 items. An additional 5 items, for clinical use alone, assess areas of anxiety often reported to be of concern for older adults. Cronbach's alphas for the GAS scores in our sample were: .87 for the Total score, .75 for the Somatic scale, .74 for the Cognitive scale, and .75 for the Affective scale.

Data Analysis

IRT analyses were used to shorten the scale. IRT posits that responses on a given item are a function of both person and item properties (Barbaranelli and Natali, 2005) so that a respondent with a certain level of ability, theta or θ (the underlying or latent trait) will have a probability p (θ) to respond correctly to an item in relation to item parameters. The first parameter is *bj* (*category threshold*), the difficulty of the item. Values range from – ∞ to + ∞. Item behavior along the ability scale is determined by item difficulty in reference to the median probability. On a characteristic item curve, items that are difficult to endorse are shifted to the right of the scale, while easier items are shifted to the left of the ability scale.

The second parameter is *aj* (*slope*) or item discrimination of different levels of ability. Values of the *slope* range from – ∞ to + ∞. Item discrimination determines the rate at which the probability of endorsing correct item changes given ability levels and is imperative in differentiating between individuals possessing similar levels of the latent construct.

IRT analyses were performed separately for each subscale. As responses to items are polytomous, with scoring ranging from zero to three, we used the graded response model of Samejima (GRM) to fit the item responses (Samejima, 1969). For this model, logistic curves, category response curves (CRC), are generated for each response option of each item showing the probability of a positive response to the option as a function of the underlying trait. Thus, threshold parameters (b_i), equal to the number of response options minus 1, for each item can be derived.

The GAS has four possible response categories (not at all, sometimes, most of the time, always), so we have three threshold parameters presented for each item (b_1, b_2, b_3). The first threshold parameter (scaled as a z-score, M = 0, SD = 1, lower values reflecting less anxiety) reflects how much anxiety is required to have a 50% chance of endorsing the 'sometimes' response category. The second threshold parameter reflects how much anxiety is needed to have a 50% chance of endorsing the 'most of the time' category, and the third threshold parameter reflects how much anxiety is needed to have a 50% chance of endorsing the 'always' category (see supplementary S5).

Additionally, the GRM provides one discrimination parameter (*a*) indicating the ability of an item to discriminate different levels of the underlying trait. TIF values and related standard errors of measurement (SEM) indicate the precision of the whole test (Baker, 2001; Embretson and Reise, 2000). A steeper slope indicates that the item provides more information about the latent trait, but over a more restricted range. A less steep slope indicates that the item provides less information over a broader range.

Discrimination parameter values ranging from 0.01 to 0.24 are considered very low, 0.25 to 0.64 low, 0.65 to 1.34 moderate, 1.35 to 1.69 high, and more than 1.7 very high (Baker, 2001). We used these criteria to determine which items discriminated anxiety levels best. The larger the value of TIF, the greater the accuracy with which the latent trait levels are measured. Thus, to shorten the original 25-item GAS, we selected the most informative items by examining the shape of each item information function (IIF).

We used a confirmatory factor analysis to assess unidimensionality, that ensures that a single construct accounts for the covariation among items, and local independence (items should not be correlated when the shared variance of the latent trait is removed), a fundamental assumption underlying the GRM (see supplementary S1). If unidimensionality is satisfied, local independence is met (Hambleton et al., 1991).

Prior to model testing, Mardia's test of normality (1974) was used to assess the normality of data (Mardia's normalized estimate = 792). The high Mardia's normalized estimate of kurtosis suggested non full normality of data. Data was then prepared for IRT (Paek and Han, 2013). We identified and retained items that provided the greatest information (12 items) and had the highest discrimination parameters while maintaining the integrity of the subscales to determine the short form (Segal et al., 2010).

Results

First, we calculated the cut-off of the GAS according to sensitivity (i.e. the percentage of true positives or positive results of an instrument in presence of disease) and specificity (i.e., the percentage of false positives or positive results of an instrument in absence of disease; Habibzadeh et al., 2016). The best cut-off of the long form of the instrument calculated through the Youden's Index (Youden, 1950) was equal to 18.5. Threshold parameter was based on cut-off score of >40 on the State Trait Inventory of Cognitive and Somatic Anxiety-State (STICSA-S; Balsamo et al., 2015). Using this cut-off, 35 participants (6.8%) presented elevated anxiety levels.

Confirmatory factor analysis - Preliminary analyses

Confirmatory Factor Analysis (CFA) was used to test unidimensionality. CFA run on the data of each of the three scales showed good fit indices confirming the unidimensionality of all subscales (see supplementary S1).

Estimated item parameters

In line with the factor analysis results, IRT analysis was performed under the unidimensional graded model. Item parameters are reported in Table 1, while information carried by each item is displayed graphically in appendix (see figure S1 to S3). An item typically offers more item information if it has a greater discriminating parameter (i.e., steeper slopes) and a broader range of threshold parameters along θ. We selected items for all subscales that offered more information by examining the shape of each item information function for each subscale.

For the nine items of the Somatic subscale, item discrimination values (*a*) ranged from 0.77 to 1.62 (M discrimination parameter = 1.26, SD = .26) with values in the moderate to high range for items 21, 22, 1 and 2 and moderate range for Items 17, 3, 9, 23, 8. Since the higher the *a,* the more an item differentiates between different levels of anxiety, we selected items 21, 22, 1 and 2 (Table 1).

Insert Table 1 here

For the Somatic subscale (see figure S1), threshold parameters of the chosen items revealed that items could be endorsed by individuals with varying amounts of anxiety severity (i.e., the threshold parameters were neither too high nor too low). We selected items 21, 22, 1 and 2. Regarding the eight items of the Cognitive scale, IRT analyses yielded discrimination parameters (*a*) ranging from 0.94 to 2.29 (M discrimination parameter = 1.64, SD = .45) reflecting a high to very high range for items 25, 24, 19, 18 and 5 and moderate to high for Items 16, 4 and 12 conveyed. We selected items 25, 24, 19, 18 and 5 (Table 1).

For the Cognitive subscale (see figure S2), we selected items 25, 24, 19 and 5. Regarding the eight items of Affective scale, item discrimination values (*a*) ranged from 0.80 to 1.81 (M discrimination parameter = 1.41, SD = .30) reflecting high to very high values for items 20, 7, 13, 11 and 6 and moderate for Items 10, 15 and 14. We selected items 20, 7, 13, 11 and 6 (Table 1). Regarding the trait measured by the Affective subscale (see figure S3), we selected items 20, 7, 13 and 6.

Items from the Somatic subscale tended to have lower discrimination parameters than items from the Cognitive and Affective subscales. This indicates that Somatic items were less informative than Cognitive and Affective items. Items were examined to exclude redundant or similar items, and threshold parameters were inspected to determine if parameters were reasonable in magnitude. Next, the TIF of the GAS was examined (see figure S4). The TIF provides test reliability estimations (information values) for each level of the latent trait. These values are equal to the inverse of the standard error, thus higher values indicate more reliable estimates. The GAS provided the greatest amount of information for individuals with average or higher levels of anxiety as indicated by the maximum Test Information Curve (TIC) and minimum standard error estimate (SEE). The TIC peak was at approximately 1.5 SD above the mean level of anxiety. The SEE was lower for average and higher levels of anxiety. The TIF also indicated that the GAS provides less information for levels below the mean level of anxiety (from -1 to -3 SD); the SEE was higher for lower levels of anxiety. This indicates that the GAS provides more information for individuals with average or higher levels of anxiety, whereas it is less precise for the lowest levels of anxiety.

Discussion

In this study, we developed a new Italian short version of the GAS with improved psychometric characteristics developed directly from Italian respondents that focuses on measures of somatic, cognitive, and affective symptoms of anxiety rather than worry symptoms as in short forms in the literature (Ferrari et al., 2017). IRT based statistics identified 12 items as well discriminating and well distributed along the latent trait, compared to the 10 items identified in the English short-form (Mueller et al., 2015). In addition, not all 10 items of the original English short-form are included amongst our 12 items. Our 12 items are well discriminating and well distributed along the latent trait yielding a clear picture of the good performance of the items and the global scale in measuring the construct of anxiety.

Items were selected upon their discrimination parameters and information curves while retaining the structure of the subscales (Edelen and Reeve, 2007). The choice of four items was a compromise between pragmatic considerations (e.g., time and cost of completion on large samples) and the need to obtain acceptable psychometric properties, especially concerning internal consistency coefficients which generally decrease as the number of items decreases (see study 2).

The Italian short form (GAS-12) differs regarding four items compared to the English version (GAS-10; Mueller et al., 2015; item 10. ‘I was irritable’, item 15. ‘I felt detached or isolated from others’, item 16. ‘I felt like I was in a daze’ and item 17. ‘I had a hard time sitting still'). Further, GAS-12 differs regarding two items from the German version (GAS-9; Gottschling et al., 2020; item 10. 'I was irritable' and item 23. 'I had back pain, neck pain, or muscle cramps'). These differences suggest that cultural differences may have contributed to the different item content.

Item response theory analyses indicated that the GAS-12 is most reliable in discriminating individuals at the average or higher end of the anxiety continuum versus people with very low levels of anxiety. Also, results are based on a general community sample of older adults, which limit the generalizability of these findings to older adults with diagnosed mental disorders. Finally, the discrimination parameter values in this study ranged from moderate to very high and as in the English version, the Somatic items provided less information than items from the Affective or Cognitive subscales, likely because they are endorsed frequently by individuals with comorbid medical conditions (Katon et al., 2007; Muller et al., 2015).

Study 2. The Italian version of the GAS-12 Short Form (Reliability and Convergent/Divergent Validity)

Study 1 developed a 12-item version of the GAS, using IRT. However, when developing a new instrument, researchers should provide evidence of its good psychometric properties (i.e., the items of the instrument are expected to be closely related as a group; the instrument is expected to measure what it claims to measure).

To date, information on the convergent and divergent validity (i.e., the degree of the association between the GAS-10 and other measures supposed to measure the same or divergent constructs) is limited. Mueller et al. (2015) provided results for item selection and age and gender differences for the original GAS-10. Carlucci et al. (2021) who examined associations between the GAS-10 and other psychological questionnaires, found strong, positive correlations with measures of anxiety and depression, and moderate, negative associations with mental health components of Quality of Life (QoL).

Since anxiety and depression are aspects of negative affectivity that play an important role in perceived quality of life (QoL) defined as an 'individuals' perception of their position in life in the context of the culture and values systems in which they live and in relation to their goals, expectations, standards and concerns (Ribeiro et al., 2020) they may constitute ideal dimensions with which to compare the GAS-12.

In addition, sleep quality also seems to be related to quality of life in older adults (Sella et al., 2021) and studies have shown associations between insomnia and anxious symptoms (Bolstad and Nadorff, 2020), with anxiety being more prevalent among individuals with onset insomnia (Bragantini et al., 2019) and this may also be an interesting dimension for evaluating.

Accordingly, Study 2 aimed to investigate the psychometric functioning of the GAS-12 in a new independent sample of older adults. Factor structure, reliability, and validity evidence were assessed by examining associations with commonly used measures of anxiety, depression, insomnia, and physical and mental health components of QoL.

Methods

Participants and Procedure

A total of 427 independent community-dwelling older adults (56% females; mean age: 73.88; SD = 7.63 years) from different Italian regions were recruited from the general population through a snowball sampling procedure. All participants had normal or corrected to normal vision and no reports of severe head trauma, stroke, psychiatric or neurological disorders nor use of psychiatric drugs. Before completing the questionnaire packet, we screened all participants with the Mini Mental State Examination to exclude participants with cognitive impairment. Participants received no monetary reimbursement for participation and anonymously completed the questionnaire packet. The study was approved by the Departmental ethical committee at the University of Chieti and all participants provided written informed consent prior to inclusion. To evaluate the construct validity of the GAS-12, participants additionally completed measures of anxiety, depression, insomnia, and QoL.

Measures

Mini Mental State Examination (MMSE)

The Mini Mental State Examination (MMSE) is a simple pen and paper test of 30 items designed to screen for cognitive impairment in older adults (see Study 1). All of the participants scored above cut-off (M = 27.70, SD = 2.37).

Geriatric Anxiety Scale - Short (GAS-12)

The GAS-12 is a 12-item self-report measure intended to assess and quantify anxiety symptoms in older adults (Study 1). It includes three theoretical subscales: Cognitive (e.g., 'I felt like I was losing control'), Somatic (e.g., 'My heart raced or beat strongly') and Affective (e.g., 'I was afraid of being judged by others'), each composed of four items. Individuals indicate how often they have experienced each symptom during the immediately preceding week along a 4-point Likert scale ranging from 0 (not at all) to 3 (always), with higher scores indicating higher levels of anxiety.

State Trait Inventory of Cognitive and Somatic Anxiety (STICSA)

The STICSA is a 21-item measure designed to assess cognitive and somatic symptoms of anxiety validated in a sample of Italian middle and older aged adults (Balsamo et al., 2015). In the trait anxiety subscale, individuals rate how often a statement is true in general, whereas in the state anxiety subscale they rate how they feel at the moment of assessment along a four-point Likert-type scale from 1 (not at all) to 4 (very much so). The overall scale includes four subscales: State-Somatic (SS), Trait-Somatic (TS), State-Cognitive (SC), and Trait-Cognitive (TC). Factor analysis supported the validity of both state-trait and cognitive-somatic distinctions underlying the STICSA. All dimensions exhibited excellent internal consistency of scale scores (Cronbach's α coefficients >= 0.86), and correlations with depression measures provided limited evidence for differentiation of anxious and depressive symptoms (Balsamo et al., 2015). The STICSA also showed evidence of discriminating anxious symptoms from physical health symptoms, a particularly relevant feature of a valid anxiety measure in older samples. In our study, all subscales of the STICSA had high internal consistency of scale scores, with Cronbach's α coefficients of 0.74 and 0.81, respectively for the STICSA-TS and the STICSA-TC, and 0.77 and 0.86, respectively for the STICSA-SS and the STICSA-SC. Cronbach's α coefficients were equal to 0.87 and 0.85, respectively for the STICSA-S and the STICSA-T.

Geriatric Depression Scale (GDS)

The GDS is a 30-item self-report questionnaire designed to measure depression in older adults (Yesavage et al., 1983). Items represent symptoms of depression common to older adults across affective and cognitive domains. Respondents rate each item on a dichotomous (yes/no) scale and the total score ranges between 0 and 30, with higher scores indicating more severe depression. The GDS has been shown to be internally consistent over time. In the original study, split-half and alpha coefficients were both 0.94 (Yesavage et al., 1983). Other studies have found similar results with diverse populations, showing alpha and split-half coefficients ranging from 0.80 to 0.99 (Agrell et al., 1989; Lesher, 1986). Among community-dwelling older adults, Yesavage et al. (1983) found an 84% sensitivity rate and a 95% specificity rate. In the present sample, the Cronbach's α was 0.88 (Kr20 = 0.88).

The Insomnia Severity Index (ISI)

The ISI is a 7-item self-report questionnaire assessing the nature, severity, and impact of insomnia in the last month (Bastien et al., 2001). Dimensions evaluated include the severity of sleep onset, sleep maintenance, and early morning awakening problems, sleep dissatisfaction, interference and distress caused by the sleep difficulties. A 5-point Likert scale is used to rate each item (e.g., 0 = no problem; 4 = very severe problem), yielding a total score ranging from 0 to 28 with (0-7) indicating absence of insomnia; (8-14) sub-threshold insomnia; (15-21) moderate insomnia; and (22-28) severe insomnia. In the original study (Bastien et al., 2001), internal reliability coefficients remained stable (0.76 at baseline, 0.78 at follow-up). Concurrent validity was documented by significant correlations with an equivalent clinician's version of the ISI and with sleep diary and polysomnographic measures. In the present sample, the Cronbach's α was 0.86.

The Short Form 36 Health Survey (SF-36)

The SF36 is a 36-item self-report measure of quality of life (McHorney et al., 1994; Ware et al., 1993; Ware et al., 1996; Apolone et al., 1998). One item is used to measure health transition (HT) while the remaining 35 items, grouped into scales, assess eight domains. These are: 1. physical functioning (PF); 2. role-physical (RP); 3. bodily pain (BP); 4. general health (GH); 5. Vitality (VT); 6. social functioning (SF); 7. role-emotional (RE); and 8. mental health (MH). These eight domains can be aggregated into two summary measures, the physical component summary measure (PCS) and the mental component summary measure (MCS). The PCS includes physical functioning, role-physical, bodily pain and general health scales and The MCS includes vitality, social functioning, role-emotional and mental health scales. Scores for each variable range from 0 to 100, with higher scores indicating better health. Ware et al. (1996) reported alpha internal consistency coefficients for the eight scales. The median alpha reliability for all scales exceeds 0.80. The SF-36 manual presents criterion validity information on the scales, comparing scale scores to ability to work, symptoms, utilization of care, and to a range of criteria for the mental health scale. In the present study, Cronbach's α coefficients were equal to 0.92 and 0.88, respectively for the PCS and the MCS.

Data Analysis

The factorial structure of the GAS-12 was examined through confirmatory factor analysis (CFA), allowing for correlation among error terms (see supplementary S2). In addition, a Multigroup Confirmatory Factor Analysis was performed to test measurement invariance of the GAS-12 with respect to gender on a set of nested models, that begin with the separate determination of a baseline model for each group. Finally, the invariance of latent factor means was examined in a CFA framework. Effect sizes were reported and interpreted according to guidelines (Cohen, 1992). Internal consistency was estimated by Cronbach's alpha (Cronbach, 1951), McDonald's omega (ω; Dunn et al., 2014), and mean corrected item-total correlations. Corrected item-total correlations were calculated to examine how each item contributed to the overall scale. Cronbach's alpha values below 0.60 are unacceptable, whereas item inter-correlation coefficients higher than 0.30 are adequate (Barbaranelli and D'Olimpio, 2007).

To assess convergent and discriminant validity, relationships between the GAS-12 total, its subscales, and all other measures were investigated using correlation coefficients (Pearson's *r*).

Results

First, we calculated the cut-off of the GAS-12 according to sensitivity (i.e. the percentage of true positives or positive results of an instrument in presence of disease) and specificity (i.e., the percentage of false positives or positive results of an instrument in absence of disease; Habibzadeh et al., 2016). The best cut-off of the Short form of the instrument calculated through the Youden's Index (Youden, 1950) was equal to 6.860. Threshold parameter was based on cut-off score of >40 on the State Trait Inventory of Cognitive and Somatic Anxiety-State (STICSA-S; Balsamo et al., 2015). Using this cut-off, 47 participants (11%) presented elevated anxiety levels.

The means of the 4-point Likert GAS-12 items were relatively low with values ranging from 0.24 (SD = .46; Item 5) to 0.96 (SD = .67; Item 9). Inspection of skewness and kurtosis indexes indicated that departures from normality were not severe, so no variable transformations were deemed necessary (West et al., 1995). The range values skewness is from 0.48 to 1.79 and kurtosis is from 0.31 to 2.40.

Confirmatory Factor Analysis, Invariance Measurement and Invariance of Latent Factors Means

Mardia's test of normality (1974) was used to assess the normality of data (Mardia's normalized estimate = 174.99). The high Mardia's normalized estimate of kurtosis suggested non full normality of data. Thus, all analyses were based on the robust maximum likelihood estimator (Satorra and Bentler, 1994).

CFA was used to validate both, the originally postulated three factor structure of the GAS-12 (First-order or lower order - Model 1: Cognitive, Affective and Somatic; Segal et al., 2010; Mueller et al., 2015; Gottschling et al, 2020), a one general anxiety factor solution (Model 2), and to test a single second-order factor (or higher order of each of these lower) of general anxiety (hierarchical - Model 3), with the first-order factors are explained by some high order structure. Goodness-of-fit statistics for all tested structural models were presented in Table 2 (see also the graphic representation in fig.1 to fig.3).

Insert Table 2 here and Insert fig.1 to fig.3 here

Results supported both, the one factor and the second-order solution implied by the GAS-12 item pool. However, Model 1 (three theoretical factor structure) demonstrated significantly better fit compared to Model 2 (one general anxiety factor solution; Satorra-Bentler Scaled Chi-Square Difference = 10.64; df = 2; p = 0.004) (Barbaranelli and Ingoglia, 2013). Further, Model 1 demonstrated a significantly better fit with respect to Model 3 (second-order structure with two modeled error covariances between items 4 and 5, and 9 and 10; Satorra-Bentler Scaled Chi-Square Difference = 5.635; df = 1; p = 0.01).

In Model 1, all factor loadings were statistically significant and ranged from 0.43 to 0.74, with an average standardized factor loading of 0.62. Squared multiple correlations ranged from 0.19 to 0.54, with an average SMC of 0.39 indicating that, on average, 39% of the variance in observed variables was accounted for by latent factors. The latent factor correlations were very high, ranging between 0.74 and 0.87. We also added structure coefficients, the correlations between the measured variables and the latent factors (see Table 3).

Insert Table 3 here

Measured variables are correlated with all factors when the factors are correlated, even for variables with CFA pattern parameters fixed to be zeroes.

A multiple-group approach was used to test measurement invariance across gender (Table 4). Measurement invariance across gender groups was entirely supported at the factorial structure and at the intercept level. The ΔCFIs were lower than 0.01 in all models, suggesting that invariance can be assumed. After establishing the full scalar invariance across gender, we compared the latent mean differences across these groups (see supplementary S3). Results indicated that whereas the means of Cognitive subscale and Somatic subscale for men were significantly different from those for women, the means for Affective subscale were not. More specifically, the psychometrics were not significantly different between men and women, but women reported more anxiety than men.

Insert here Table 4

Reliability

Internal consistency of the GAS-12 subscales was sufficient to good for all 3 subscales: Cognitive, $\alpha = 0.76$ ($\omega = 0.79$); Affective, $\alpha = 0.70$, ($\omega = 0.72$); Somatic, $\alpha = 0.68$ ($\omega = 0.69$). Internal consistency was excellent for the GAS-12 Total score: $\alpha = 0.85$ ($\omega = 0.86$; see supplementary S4).

Scale Intercorrelations

As expected, the GAS-12 Total scale positively and strongly correlated with the Cognitive, Affective, and Somatic subscales (Table 5). The correlation between the Cognitive and Affective subscales was stronger (large effect size) than the correlation between the Cognitive and Somatic subscale ($z = 1.95$, $p = 0.025$) and between the Affective and Somatic subscale ($z = 2.12$, $p = 0.017$; see supplementary S2).

Insert Table 5 here

Construct Validity

To investigate the convergent and discriminant validity of the Italian GAS-12, we computed correlations between the GAS-12 total and its subscales with measures of depression, anxiety, subjective health status, insomnia, education, gender, and age (Table 4). As expected, the GAS-12 total and its subscales were significantly and positively associated with other measures of depression, anxiety, insomnia, and significantly negatively correlated with measures of subjective health status, with medium to high effect sizes.

The Somatic subscale of the GAS-12 had higher correlation coefficients particularly with the physical component measure of SF-36 (PCS), which includes physical functioning, role-physical, bodily pain and general health, compared to the Cognitive and Affective scales, indicating good construct validity for the domains specifically designed to assess. Finally, the GAS showed good discriminant validity with demographic variables (i.e., education, gender, age). Correlations were low and only a few appeared significant.

Discussion

Study 2 examined reliability as well as evidence for convergent and discriminant validity of the Italian GAS-12 (Picconi et al., 2018) in an independent sample of older adults. The CFA analysis confirmed the better fit of the three-factor model (Cognitive, Somatic, and Affective) and the 12 items selected through IRT-based statistics as the most discriminating and well distributed items along the latent trait.

We also found strong positive relationships between the GAS-12 total score and each of the GAS-12 subscales. Convergent validity of the GAS-12 was evidenced via significant and high correlations between the GAS-12 total score, subscale scores and another measure of anxiety (STICSA), with medium and high effect size and mental health (MCS), confirming that the GAS-12 effectively measures anxiety in Italian older adults.

Regarding discriminant validity of the GAS-12, findings confirmed low relations with measures of constructs unrelated to anxiety (i.e., education, gender, age), whereby the relationship between the GAS-12 total score, subscale scores and depression was lower than the correlation with anxiety measure. Correlations indicate that those who reported more anxiety symptoms also reported more depressive symptoms and sleep difficulties and those with elevated anxiety symptoms also rated their health more poorly.

General Discussion

We developed a brief quantitative instrument of anxiety in Italian older adults that can be easily administered and used for rapid screening in older adults. Results are important because, although content validity may be lost due to the reduction of the number of items compared to the long form, results suggest that the GAS-12 is a good screening instrument and can detect anxiety in older adults. This is crucial since anxiety causes considerable subjective distress while its detection is often complicated by the high frequency of medical disorders present in this age group (Balsamo et al., 2015; Balsamo et al., 2018). Reliability and validity evidence suggests that the GAS-12 is a fast and easily administered instrument that can be used in clinical and research situations for screening anxiety in the Italian geriatric population.

Our study, however, is not without limitations. First, we did not investigate aspects of reliability of the questionnaire (e.g., test-retest). Second, the somatic scale is below threshold for good internal consistency (Barbaranelli and D'Olimpio, 2007). Third, our results are based on a general community sample of older adults, limiting the generalizability of findings to clinical samples (e.g., Balsamo et al., 2013). Also, regarding cultural differences, differences could be due to the way in which each instrument was constructed and how items were developed and selected. Future studies need to adopt comparable methods to confirm cultural differences. Future studies should also include older adults with anxiety disorders to extend findings to clinical populations and to increase generalizability and identify optimal cut-off scores. Future studies also need to address the lower discrimination parameters of the somatic subscale. Notwithstanding, the Italian GAS-12 appears to be a highly promising assessment measure of anxiety and should be further scrutinized and further developed in additional studies.

Clinical Implications

New epidemiological data show the deleterious impact of late-life anxiety in terms of mortality, suicide, and social isolation (Central Statistical Institute, 2018; Italian National Institute of Health, 2018). Furthermore, the anxiety is a problematic aspect, due to an overlap with processes such as cognitive decline or alcohol abuse. Consequently, literature underlines the urgency of an adequate screening in older adults to permit accurate clinical treatment (Balsamo et al., 2018; Canuto et al., 2018).

- The Italian GAS-12 allows the screening of anxiety in Italian older adults;
- The GAS-12 is particularly indicated with limited time and many scales in a clinical assessment or research protocol.

Conflict of interest

None.

References

Agrell, B., Dehlin, O. (1989). Comparison of six depression rating scales in geriatric stroke patients. *Stroke*, 20, 1190–94.

Apolone, G., Mosconi, P. (1998). The Italian SF-36 Health Survey: Translation, Validation and Norming. *Journal of Clinical Epidemiology*, 51, 1025-1036.

Baxter, A. J., Scott, K. M., Vos, T. and Whiteford, H. A. (2013). Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychological Medicine*, 43(5), 897. doi:10.1017/S003329171200147X

Baker, F. B. (2001). The basics of item response theory. College Park, MD: Eric Clearinghouse on Assessment and Evaluation.

Balsamo, M., Cataldi, F., Carlucci, L., Padulo, C. and Fairfield, B. (2018). Assessment of late-life depression via self-report measures: a review. *Clinical Interventions in Aging*, 13, 2021-2044. doi: 10.2147/CIA.S178943

Balsamo, M., Cataldi, F., Carlucci, L., and Fairfield, B. (2018). Assessment of anxiety in older adults: a review of self-report measures. *Clinical interventions in aging, 13*, 573. doi: 10.2147/CIA.S114100

Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *International Psychogeriatrics*, 27(6), 999. doi: https://doi.org/10.1017/S1041610214002634

Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L. and Saggino, A. (2013). The state-trait anxiety inventory: shadows and lights on its construct validity. *Journal of Psychopathology and Behavioral Assessment*, 35(4), 475-486.

Barbaranelli, C. and D'Olimpio, F. (2006). *Analisi dei dati con SPSS. I. Le analisi di base*. Milano: LED.

Barbaranelli, C. and Natali, E. (2005). *I test psicologici. Teorie e modelli psicometrici. Roma*: Carocci.

Barbaranelli, C. and Ingoglia, S. (2013). *I modelli di equazioni strutturali: temi e prospettive*. Milano: LED.

Bastien, C. H., Vallières, A. and Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, 2(4), 297-307. doi: 10.1016/s1389-9457(00)00065-4

Bolghan-Abadi, M., Segal, D. L., Coolidge, F. L. and Gottschling, J. (2013). Persian version of the Geriatric Anxiety Scale: translation and preliminary psychometric properties among Iranian older adults. *Aging & Mental Health*, 17(7), 896-900. doi.org/10.1080/13607863.2013.788999

Bolstad, C. J. and Nadorff, M. R. (2020). What types of insomnia relate to anxiety and depressive symptoms in late life? *Heliyon*, 6(11), e05315. https://doi.org/10.1016/j.heliyon.2020.e05315

Bragantini, D., Sivertsen, B., Gehrman, P., Lydersen, S. and Güzey, I. C. (2019). Differences in anxiety levels among symptoms of insomnia. The HUNT study. *Sleep Health*, 5(4), 370-375. https://doi.org/10.1016/j.sleh.2019.01.002

Canuto, A., Weber, K., Baertschi, M., Andreas, S., Volkert, J., Dehoust, M. C. and Härter, M. (2018). Anxiety disorders in old age: psychiatric comorbidities, quality of life, and prevalence according to age, gender, and country. *The American Journal of Geriatric Psychiatry, 26(2)*, 174-185. https://doi.org/10.1016/j.jagp.2017.08.015

Carlucci, L., Balestrieri, M., Maso, E., Marini, A., Conte, N., and Balsamo, M. (2021). Psychometric properties and diagnostic accuracy of the short form of the geriatric anxiety scale (GAS-10). *BMC Geriatrics*, 21(1), 1-12. https://doi.org/10.1186/s12877-021-02350-3

Central Statistic Institute. (2018). *La salute mentale nelle varie fasi della vita*. Retrieved from https://www.istat.it/it/files/2018/07/Report_Salute_mentale.pdf

Cohen, J. (1992). Statistical power analysis. Current Directions in *Psychological Science*, 1(3), 98-101. doi.org/10.1111/1467-8721.ep10768783

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

de Beurs, E., Van Balkom, A. J. L. M., Dyck, R. V. and Lange, A. (1999). Long-term outcome of pharmacological and psychological treatment for panic disorder with agoraphobia: a 2-year naturalistic follow-up. *Acta Psychiatrica Scandinavica*, 99(1), 59-67. doi: 10.1111/j.1600-0447.1999.tb05385.x

Dahany, M. M., et al. (2014). Factors associated with successful aging in persons aged 65 to 75 years. *European Geriatric Medicine*, *5*(6), 365-370.

Dunn, T. J., Baguley, T. and Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412. doi.org/10.1111/bjop.12046

Edelen, M. O. and Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(1), 5-18. doi: 10.1007/s11136-007-9198-0

Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ.: Lawrence Erlbaum Associates.

Feldt, L.S., Woodruff, D.J., and Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93–103. doi: 10.1177/014662168701100107

Ferrari, S., Signorelli, M. S., Cerrato, F., Pingani, L., Massimino, M., Valente, S., & Atti, A. R. (2017). Never too late to be anxious: validation of the Geriatric Anxiety Inventory, Italian version. *Clinical Therapeutics*, 168(2), e120-e127

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state: a practical method for grading the state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198. doi.org/10.1016/0022-3956(75)90026 -6

Friedman, S. (2002). Cultural issues in the assessment of anxiety disorders. In *Practitioner's guide to empirically based measures of anxiety* (pp. 37-41). Springer, Boston, MA.

Gatti, A. et al. (2018). An investigation of the psychometric properties of the Geriatric Anxiety Scale (GAS) in an Italian sample of community-dwelling older adults. *Aging & Mental Health*, 22(9), 1176-1184. doi.org/10.1080/13607863.2017.1347141

Gerolimatos, L. A., Ciliberti, C. M., Gregg, J. J., Nazem, S., Bamonti, P.M., Cavanagh, C. E., & Edelstein, B. A. (2015). Development and preliminary evaluation of the Anxiety in Cognitive Impairment and Dementia (ACID) scales. *International Psychogeriatrics*, 27(11), 1825-1838

Gottschling, J., Segal, D. L., Häusele, C., Spinath, F. M. and Stoll, G. (2016). Assessment of anxiety in older adults: translation and psychometric evaluation of the German version of the Geriatric Anxiety Scale (GAS). *Journal of Psychopathology and Behavioral Assessment*, 38(1), 136-148.

Gottschling, J. et al. (2020). Measuring anxiety in older adults: Development, diagnostic accuracy, and preliminary validation of a short-form of the German Geriatric Anxiety Scale (GAS-G-SF). *Journal of Personality Assessment*, 102*(2)*, 196-204. doi: 10.1080/00223891.2019.1677679

Istituto Nazionale di Statistica (2018). *La salute mentale nelle varie fasi della vita*. Retrieved from www.istat.it/it/files//2018/07/.

Italian National Institute of Health (2017). *Salute mentale epidemiologia in Italia*. Retrieved from https://www.epicentro.iss.it/mentale/epidemiologia-italia

Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia medica, 26*(3), 297-307. https://doi.org/10.11613/BM.2016.034

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Thousand Oaks, Calif.: Sage.

Hinton, D. E. (2012). Multicultural challenges in the delivery of anxiety treatment. *Depression and Anxiety*, 29(1), 1–3. https://doi.org/10.1002/da.20889

Hofmann, S. G., Anu Asnaani, M. A., & Hinton, D. E. (2010). Cultural aspects in social anxiety and social anxiety disorder. *Depression and Anxiety, 27(12)*, 1117-1127.

Hofmann, S. G., & Hinton, D. E. (2014). Cross-cultural aspects of anxiety disorders. *Current Psychiatry Reports, 16(6)*, 1-5.

Hussenoeder, F. S. et al. (2021). Depression and quality of life in old age: A closer look. *European Journal of Ageing*, 18, 75-83. https://doi.org/10.1007/s10433-020-00573-8

Katon, W., Lin, E. H. and Kroenke, K. (2007). The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *General Hospital Psychiatry*, 29(2), 147-155. doi: 10.1016/j.genhosppsych.2006.11.005

Kiely, K. M., Brady, B. and Byles, J. (2019). Gender, mental health and ageing. *Maturitas*, 129, 76-84. https://doi.org/10.1016/j.maturitas.2019.09.004

Lesher, E.L. (1986). Validation of the Geriatric Depression Scale among nursing home residents. *Clinical Gerontologist*, 4, 21–8.

Lin, X. L., Lu, D. L., Gottschling, J., Segal, D. L. and Tang, S. Y. (2017). Validation of a Chinese version of the Geriatric Anxiety Scale among community-dwelling older adults in mainland China. *Journal of Cross-Cultural Gerontology*, 32(1), 57-70. doi: 10.1007/s10823-016-9302-4

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhyā: *The Indian Journal of Statistics*, Series B, 36(2), 115-128.

Marques, L., Robinaugh, D. J., LeBlanc, N. J., & Hinton, D. (2011). Cross-cultural variations in the prevalence and presentation of anxiety disorders. *Expert Review of Neurotherapeutics*, 11(2), 313-322.

McHorney, C. A., Ware Jr, J. E., Lu, J. R. and Sherbourne, C. D. (1994). The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*, 32(1), 40-66. doi: 10.1097/00005650-199401000-00004

Mueller, A. E. et al. (2015). Geriatric Anxiety Scale: item response theory analysis, differential item functioning, and creation of a ten-item short form (GAS-10). *International Psychogeriatrics*, 27(7), 1099-1111. doi:10.1017/S1041610214000210

Nieuwenhuis-Mark, R.E. (2010). The death knoll for the MMSE: has it outlived its purpose? *Journal of Geriatric Psychiatry and Neurology*, 23(3):151-7.

Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E. and Arnold, E. (2007). Development and validation of the Geriatric Anxiety Inventory. *International Psychogeriatrics*, 19(1), 103-114. doi: 10.1017/S1041610206003504

Paek, I., & Han, K. T. (2013). IRTPRO 2.1 for Windows (item response theory for patient-reported outcomes). *Applied Psychological Measurement*, 37(3), 242-252.

Petrillo, J., Cano, S. J., McLeod, L. D. and Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value in Health*, 18(1), 25-34. doi.org/10.1016/j.jval.2014.10.005

Picconi, L., Balsamo, M., Palumbo, R. and Fairfield, B. (2018). Testing factor structure and measurement invariance across gender with Italian Geriatric Anxiety Scale. *Frontiers in Psychology*, 9, 1164. doi.org/10.3389/fpsyg.2018.01164

Pifer, M.A., and Segal D.L. (2020). Geriatric anxiety scale: development and preliminary validation of a long-term care anxiety assessment measure. *Clinical Gerontologist*, *43(3)*, 295–307. https://doi.org/10.1080/07317115.2020.1725793.

Reynolds, C. R., Richmond, B. O. and Lowe, P. A. (2003). *The Adult Manifest Anxiety Scale-College Version*. Los Angeles, CA: Western Psychological Services.

Ribeiro, O., Teixeira, L., Araújo, L., Rodríguez-Blázquez, C., Calderón-Larrañaga, A. and Forjaz, M. J. (2020). Anxiety, depression and quality of life in older adults: Trajectories of influence across age. International Journal of Environmental *Research and Public Health*, 17(23), 9039. https://doi.org/10.3390/ijerph17239039

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 1-100.

Satorra, A. and Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In von Eye and C. C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research* (pp. 399-419). Thousand Oaks, CA: Sage.

Segal, D. L., June, A., Payne, M., Coolidge, F. L. and Yochim, B. (2010). Development and initial validation of a self-report assessment tool for anxiety among older adults: The Geriatric Anxiety Scale. *Journal of Anxiety Disorders*, 24(7), 709-714. doi.org/10.1016/j.janxdis.2010.05.002

Segal, D. L., Granier, K. L., Pifer, M. A. and Stone, L. E. (2019). *Geriatric Anxiety Scale*. Encyclopedia of gerontology and population aging. Cham, Switzerland: Springer.

Segal, D. L., Qualls, S. H. and Smyer, M. A. (2018). *Aging and mental health (3rd edition)*. Hoboken, NJ: Wiley/Blackwell.

Sella, E., Miola, L., Toffalini, E. and Borella, E. (2021). The relationship between sleep quality and quality of life in aging: a systematic review and meta-analysis. *Health Psychology Review*, 1-58. https://doi.org/10.1080/17437199.2021.1974309

Therrien, Z., and Hunsley, J. (2012). Assessment of anxiety in older adults: a systematic review of commonly used measures. *Aging & Mental Health* 16, 1–16. doi: 10.1080/13607863.2011.602960

Vaillant, G. E. and Mukamal, K. (2001). Successful aging. *American Journal of Psychiatry*, *158*(6), 839-847.

Ware Jr, J. E., Kosinski, M. and Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3), 220-233.

Ware, J. E., Snow, K. K., Kosinski, M. and Gandek, B. (1993). *Health survey manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Center.

West, S. G., Finch, J. F. and Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Wolitzky-Taylor, K. B., Castriotta, N., Lenze, E. J., Stanley, M. A. and Craske, M. G. (2010). Anxiety disorders in older adults: a comprehensive review. *Depression and Anxiety*, 27(2), 190-211. doi: 10.1002/da.20653

Yesavage, J. A. et al. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*, 37-49.

Yochim, B. P., Mueller, A. E. and Segal, D. L. (2013). Late life anxiety is associated with decreased memory and executive functioning in community dwelling older adults. *Journal of Anxiety Disorders*, 27(6), 567-575. doi.org/10.1016/j.janxdis.2012.10.010

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35. https://doi.org/10.1002/1097-0142

The World Health Organization Quality of Life assessment (WHOQOL) (1982): position paper from the World Health Organization. (1995). *Social Science & Medicine*, *41(10)*, 1403–1409. https://doi.org/10.1016/0277-953, 6(95)00112-k

Table 1. Item Response Theory Parameters of the full Graded Response Model

	Item Response Theory Parameters			
GAS Item - *Somatic scale*	*a*	b_1	b_2	b_3
1[a]. My heart raced or beat strongly	1.37 (.17)	-0.35 (.09)	2.60 (.26)	-
2[a]. My breath was short	1.32 (.16)	-0.29 (.09)	2.38 (.24)	-
3[a]. I had an upset stomach	1.09 (.14)	0.05 (.09)	2.90 (.33)	-
8. I had difficulty falling asleep	1.31 (.15)	-0.78 (.11)	1.25 (.13)	2.91 (.30)
9. I had difficulty staying asleep	1.11 (.14)	-0.36 (.10)	1.62 (.19)	3.81 (.46)
17[a]. I had a hard time sitting still	0.77 (.13)	1.35 (.23)	4.20 (.70)	-
21. I felt tired	1.62 (.17)	-1.54 (.14)	0.97 (.10)	2.79 (.25)
22. My muscles were tense	1.57 (.18)	-0.15 (.08)	2.02 (.18)	3.32 (.35)
23. I had back pain, neck pain, or muscle cramps	1.21 (.14)	-1.64 (.17)	0.86 (.11)	2.78 (.28)
GAS Item – *Cognitive Scale*	*a*	b_1	b_2	b_3
4[a]. I felt like things were not real or like I was outside of myself.	1.16 (.20)	1.78 (.23)	-	-
5[a]. I felt like I was losing control	1.67 (.22)	0.85 (.09)	3.48 (.42)	-
12[a]. I had difficulty concentrating	0.94 (.13)	-0.47 (.12)	3.64 (.48)	-
16[a]. I felt like I was in a daze	1.35 (.18)	0.89 (.11)	3.57 (.44)	-
18. I worried too much	1.82 (.21)	-0.73 (.09)	1.24 (.10)	2.82 (.27)
19. I could not control my worry	1.93 (.22)	0.09 (.07)	1.66 (.13)	2.60 (.23)
24[a]. I felt like I had no control over my life	1.96 (.31)	1.32 (.12)	-	-
25[a]. I felt like something terrible was going to happen to me	2.29 (.32)	0.91 (.08)	2.83 (.28)	-
GAS Item - *Affective Scale*	*a*	b_1	b_2	b_3
6[a]. I was afraid of being judged by others	1.46 (.18)	0.05 (.08)	2.57 (.26)	-
7[a]. I was afraid of being humiliated or embarrassed	1.60 (.21)	0.68 (.09)	2.68 (.27)	-
10. I was irritable	1.32 (.16)	-0.69 (.10)	2.20 (.22)	4.31 (.55)
11. I had outbursts of anger	1.52 (.18)	0.20 (.08)	2.34 (.22)	4.36 (.60)
13[a]. I was easily startled or upset	1.54 (.18)	0.36 (.08)	2.62 (.25)	-
14[a]. I was less interested in doing something I typically enjoy	0.80 (.13)	0.59 (.14)	4.15 (.63)	-
15[a]. I felt detached or isolated from others	1.22 (.16)	0.75 (.11)	3.20 (.37)	-
20[a]. I felt restless, keyed up, or on edge	1.81 (.21)	-0.29 (.07)	1.77 (.15)	4.22 (.63)

Notes. Discrimination (*a*) and category threshold (*b*) estimates with standard errors (in parentheses). Items included on GAS-12 are highlighted in bold. [a]Items without b_3 parameter estimates are collapsed due to sparse cells.

Table 2. Fit indices for the structural models (*N = 427*)

Model	SB χ^2	*df*	CFI	NNFI	SRMR	RMSEA	90% CI	AIC
Three theoretical factor (Model 1)	105.11*	49	0.95	0.94	0.041	0.052	0.038/0.065	7.11
One factor (Model 2)	114.99*	51	0.94	0.93	0.042	0.054	0.041/0.067	12.99
Second-order (Model 3)	111.33*	50	0.94	0.93	0.043	0.054	0.040/0.067	11.33

Notes. *p < 0.001. SB χ_2, Satorra and Bentler chi-squared test; df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; NNFI, non normed fit index, RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of RMSEA; AIC, Akaike's information criterion used in the comparison of two or more models with smaller values representing a better fit of the hypothesized model.

Table 3. Factor Loadings, Standardized solution and Factor Structure Coefficients (R_s) – Model 1

	Cognitive pattern (R_s)	Affective pattern (R_s)	Somatic pattern (R_s)
Item 7. I could not control my worry.	0.602 (0.524)	0	0 (0.446)
Item 12. I felt like something terrible was going to happen to me.	0.721 (0.628)	0	0 (0.534)
Item 3. I felt like I was losing control.	0.718 (0.626)	0	0 (0.532)
Item 11. I felt like I had no control over my life.	0.735 (0.640)	0	0 (0.545)
Item 4. I was afraid of being judged by others.	0	0.433 (0.377)	0 (0.358)
Item 5. I was afraid of being humiliated or embarrassed.	0	0.590 (0.514)	0 (0.487)
Item 6. I was easily startled or upset.	0	0.657 (0.572)	0 (0.543)
Item 8. I felt restless, keyed up, or on edge.	0	0.656 (0.572)	0 (0.543)
Item 9. I felt tired.	0 (0.388)	0 (0.432)	0.523
Item 10. My muscles were tense.	0 (0.395)	0 (0.440)	0.533
Item 1. My heart raced or beat strongly.	0 (0.498)	0 (0.555)	0.672
Item 2. My breath was short.	0 (0.413)	0 (0.459)	0.556

Notes. Pattern coefficients constrained and not estimated in the model are presented as '0'; the structure coefficients are added in parentheses next to the pattern coefficients.

Table 4. Test for measurement invariance of the GAS-12 across gender: Summary of Goodness of Fit Statistics

Model	SB χ^2	df	CFI	RMSEA	90% CI	Model Comparison	ΔCFI
Baseline model males	77.48	50	0.931	0.054	0.028/0.077	----------	-------
Baseline model females	81.90	49	0.953	0.053	0.032/0.073	----------	-------
M1	159.10	99	0.937	0.038	0.026/0.048	----------	-------
M2*	163.94	108	0.941	0.035	0.023/0.045	2 vs. 1	0.004
M3	197.76	120	0.936	0.039	0.029/0.048	3 vs. 2	0.005

Notes. M1, model for configural invariance, no constraints; M2, model for full metric invariance with all factor loadings constrained equal. M3, model for scalar invariance with all intercepts constrained equal.
*We included the correlation between errors. Error covariance involving GAS8 and GAS7, unique for males and GAS5, GAS4 and GAS1, GAS9, unique for females.

Table 5. Gas-12 inter-scale correlations (n = 427), correlations with convergent (STICSA, n = 426; SF-36, n = 424) and discriminant scales (GDS, n = 427; ISI, n = 426; Education, n = 425; Sex, n = 427; Age, n = 425)

	C	**A**	**S**	**Total GAS-12**
Cognitive (C)	1			0.86*** (p = 0.000)
Affective (A)	0.64*** (p = 0.000)	1		0.85*** (p = 0.000)
Somatic (S)	0.55*** (p = 0.000)	0.54*** (p = 0.000)	1	0.84*** (p = 0.000)
STICSA-Trait	0.58*** (p = 0.000)	0.58*** (p = 0.000)	0.56*** (p = 0.000)	0.68*** (p = 0.000)
STICSA-Trait, Somatic	0.47*** (p = 0.000)	0.47*** (p = 0.000)	0.62*** (p = 0.000)	0.62*** (p = 0.000)
STICSA-Trait, Cognitive	0.55*** (p = 0.000)	0.55*** (p = 0.000)	0.40*** (p = 0.000)	0.59*** (p = 0.000)
STICSA-State	0.47*** (p = 0.000)	0.48*** (p = 0.000)	0.51*** (p = 0.000)	0.57*** (p = 0.000)
STICSA-State, Somatic	0.32*** (p = 0.000)	0.32*** (p = 0.000)	0.50*** (p = 0.000)	0.45*** (p = 0.000)
STICSA-State, Cognitive	0.48*** (p = 0.000)	0.50*** (p = 0.000)	0.41*** (p = 0.000)	0.54*** (p = 0.000)
ISI	0.37*** (p = 0.000)	0.36*** (p = 0.000)	0.36*** (p = 0.000)	0.43*** (p = 0.000)
GDS	0.43*** (p = 0.000)	0.40*** (p = 0.000)	0.32***(p = 0.000)	0.45*** (p = 0.000)
SF-36, PCS	-0.37***(p = 0.000)	-0.32*** (p = 0.000)	-0.50*** (p = 0.000)	-0.47*** (p = 0.000)
SF-36, MCS	-0.50*** (p = 0.000)	-0.49*** (p = 0.000)	-0.51*** (p = 0.000)	-0.59*** (p = 0.000)
Education	-0.12* (p = 0.014)	-0.12* (p = 0.012)	-0.16** (p = 0.001)	-0.16** (p = 0.001)
Sex	0.16** (p = 0.001)	0.08	0.16** (p = 0.001)	0.16** (p = 0.001)
Age	0.11* (p = 0.021)	0.08	0.09	0.11* (p = 0.023)

Notes. C, Cognitive Subscale; A, Affective Subscale; S, Somatic Subscale; STICSA, State Trait Inventory of cognitive and somatic anxiety; ISI, The Insomnia severity index; GDS, Geriatric Depression Scale; SF-36, The Short Form 36 health Survey; PCS, SF-36 physical composite score; MCS, SF-36 mental composite score; *p < 0.05; **p < 0.01; ***p < 0.001; actual p-values are in parentheses.

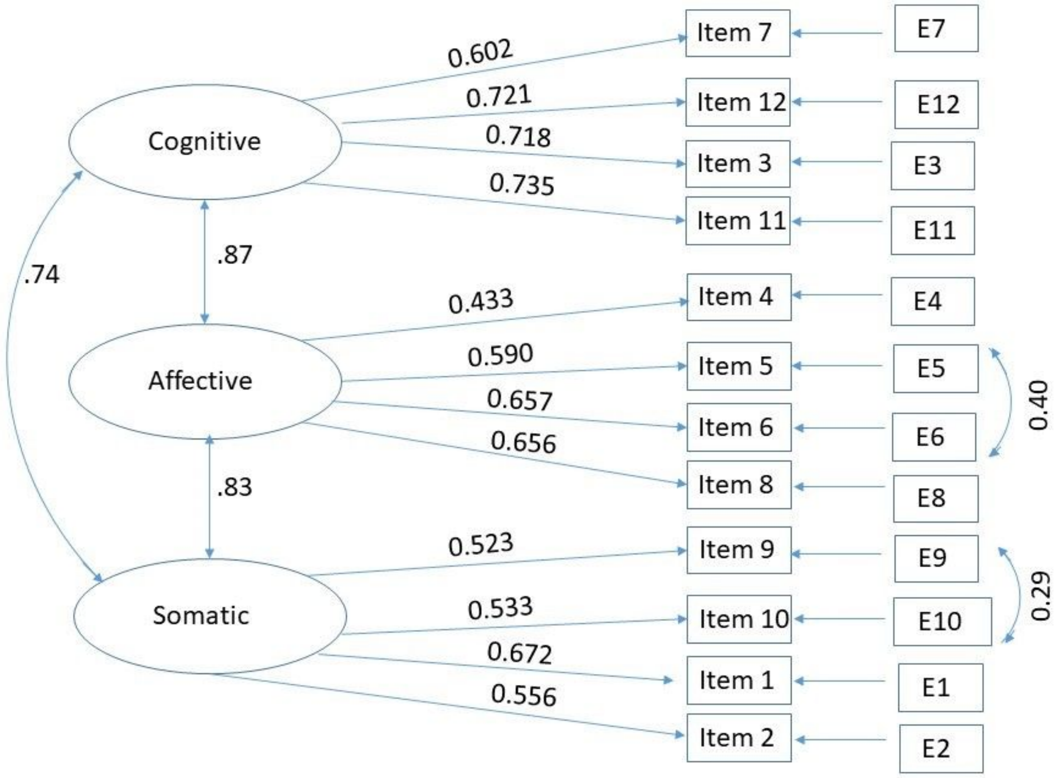


Fig.1. FACTOR LOADINGS AND STANDARDIZED SOLUTION – MODEL 1
E:: Measurement error of observed variable; Arrows (←): Standardized regression weights; Arrow (↔): Covariance

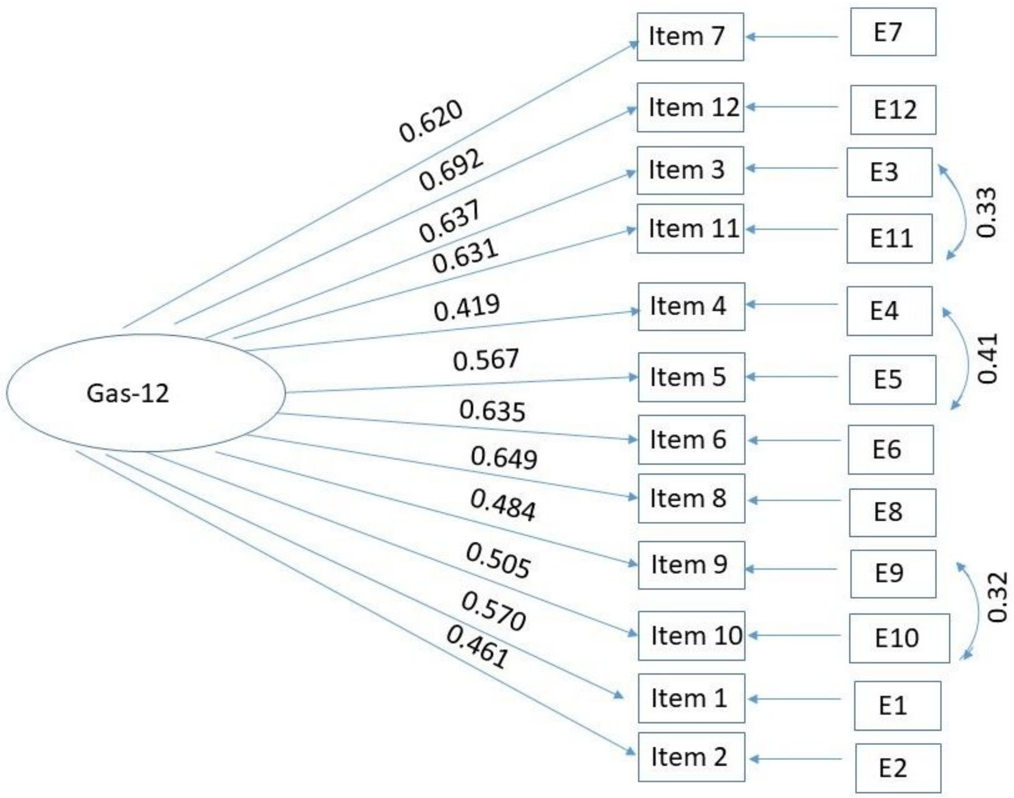


Fig.2. FACTOR LOADINGS AND STANDARDIZED SOLUTION – MODEL 2
E: Measurement error of observed variable; Arrows (←): Standardized regression weights; Arrow (↔): Covariance

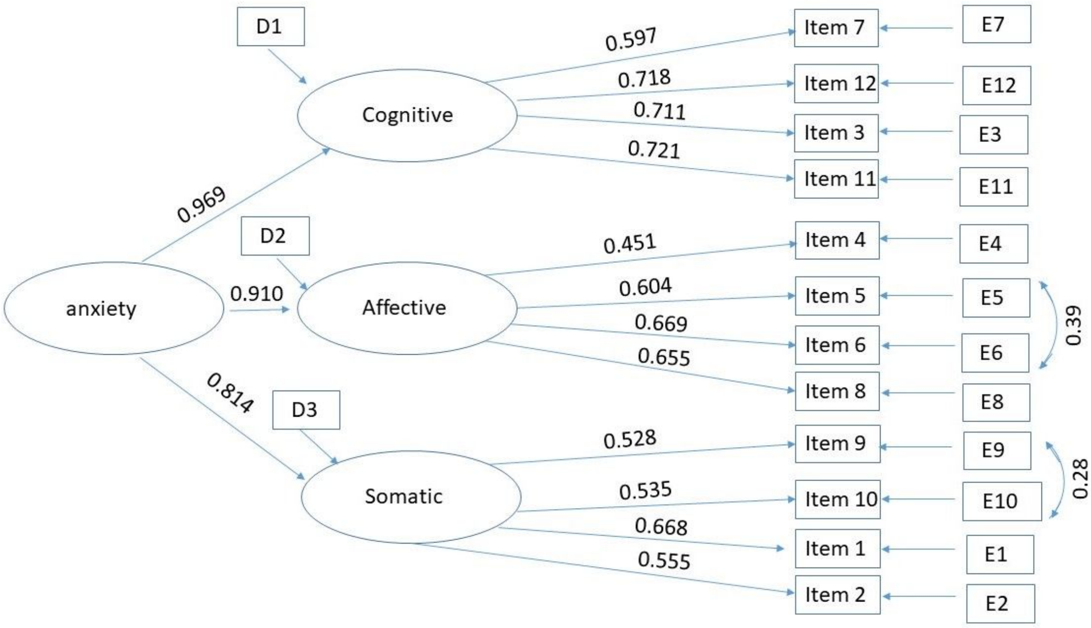


Fig.3. FACTOR LOADINGS AND STANDARDIZED SOLUTION – MODEL 3
E: Measurement error of observed variable; D: disturbance term; Arrows (←): Standardized regression weights; Arrow (↔): Covariance

APPENDIX

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY S1.

Factorial structure was examined within the framework of structural equation modeling (CFA) analyzed by EQS 6.0 (Bentler, 2006).

The unidimensionality of *Somatic*: SB $\chi^2_{(25)} = 38.29$, $p < 0.05$; CFI = 0.98; RMSEA = 0.032; 90% CI = (0.006; 0.051)**;** SRMR = 0.03. All factor loadings were statistically significant, ranging from .31 to .63, with an average standardized factor loading of 0.48 (with the presence of two error covariances between the items (GAS9 and GAS8, GAS2 and GAS1), suggested by Lagrange multiplier test (MI) and by the EPC. Squared multiple correlations ranged from 0.10 to 0.40, with an average SMC of 0.24 indicating that, on average, 24% of the variance in observed variables was accounted for by latent factor;

The unidimensionality of *Cognitive*: SB $\chi^2_{(18)} = 41.46$, $p < 0.01$; CFI = 0.96; RMSEA = 0.050; 90% CI = (0.030; 0.070)**;** SRMR =0.04. All factor loadings were statistically significant, ranging from .38 to .62, with an average standardized factor loading of 0.51 (with the presence of two error covariances between the items (GAS19 and GAS18, GAS25 and GAS24), suggested by Lagrange multiplier test (MI) and by the EPC. Squared multiple correlations ranged from 0.15 to 0.38, with an average SMC of 0.26 indicating that, on average, 26% of the variance in observed variables was accounted for by latent factor;

The unidimensionality of *Affective*: SB $\chi^2_{(18)} = 46.66$, $p < 0.001$; CFI = 0.95; RMSEA = 0.056; 90% CI = (0.036; 0.075)**;** SRMR =0.04. All factor loadings were statistically significant, ranging from .31 to .68, with an average standardized factor loading of 0.51 (with the presence of two error covariances between the items (GAS7 and GAS6, GAS15 and GAS14), suggested by Lagrange multiplier test (MI) and by the EPC. Squared multiple correlations ranged from 0.10 to 0.46, with an average SMC of 0.27 indicating that, on average, 27% of the variance in observed variables was accounted for by latent factor.

In the local independence, the item characteristic function expresses the association between the latent trait and observed responses.

SUPPLEMENTARY S2.

The analyses (CFA; MG-CFA) were performed on covariance matrices, since SEM statistical theory relies on the distributional properties of the elements of a covariance matrix, using the software EQS 6.0 (Bentler, 2006). The method of estimation used in all models was the robust maximum likelihood estimator, which yields corrected standard errors using the Satorra-Bentler method (Satorra & Bentler, 1994). Accordingly, we reported the Satorra-Bentler chi square statistic, with the following robust indices: robust comparative fit index (CFI), robust non normed fit index (NNFI), robust root mean square error of approximation (RMSEA), and robust standardized root-mean-square residual (SRMR). The following heuristic labels were used to describe model fit: acceptable when CFI and NNFI was 0.90-0.94, RMSEA was 0.08 and SRMR was 0.08, while good when CFI and NNFI is equal to or above 0.95, RMSEA is 0.06 or below and SRMR is 0.05 (Byrne, 2006). Lagrange multiplier test (LM) was used to identify which fixed parameters, if freely estimated, would lead to a significantly better fitting model. The LM test operates multivariately in determining misspecified parameters in a model. EQS produces univariate and multivariate χ^2 statistics that permit evaluation of the appropriateness of the specific restrictions; it also yields a parameter change statistic that represents the value that would be obtained if a particular fixed parameter were freely estimated in a future run. Statistically significant LM χ^2 values would argue for the presence of factor cross-loadings and error covariances, respectively. Decisions regarding possible misspecification followed by respecification of the model are based on the incremental univariate statistics. The user tipically looks for parameters whose χ^2 values stand apart from the rest and probabilities <0.05 (Byrne, 2006). We used the Expected Parameter Change (EPC) in combination with the Modification Index (MI) (Saris, Satorra, & van der Veld, 2009). For each parameter tested via the LM Test, the parameter change statistic represents its estimated value if this parameter was freely estimated in a subsequent test of the model. If the EPC is rather small, one concludes that there is no serious misspecification. However, when the EPC is large, for example larger than 0.2, it is concluded that there is a relevant misspecification in the model.

A Multigroup Confirmatory Factor Analysis (MG-CFA; van de Schoot, Lugtig, & Hox, 2012) was performed to test measurement invariance of the GAS-12 with respect to gender on a set of nested models, that begin with the separate determination of a baseline model for each group. Estimation was based on the robust statistics (ML, robust; the SB χ^2) and analyses were based on the covariance matrix.

We modeled variances, covariances as well intercepts. Associated with each constraint was a cumulative multivariate LM Test χ^2 value, and an incremental univariate χ^2 value, along with their probability values. To locate parameters that are noninvariant across groups, we looked for probability values associated with the incremental univariate χ^2 values that are <0.05. Invariance was tested for configural (M1), metric (M2) and scalar (M3) invariance. According to Cheung and Rensvold (2000), the CFI was a robust statistic for testing the between-group invariance of CFA models. They recommended that invariance can be assumed when this value was 0.01 or less, in absolute values.

The invariance of Latent Factor Means was to be examined in a CFA framework. We used the value of the critical ratio (CR) to assess latent mean differences. CR was calculated by parameter estimate divided by its standard error, which tests whether the coefficient is significantly different from 0. A CR value larger than 1.96 indicates statistically significant differences in the latent means (Byrne, 2006).

The application of the Fisher r-to-z transformation (Cohen and Cohen, 1983) is used to examine one-tailed differences in the magnitude of the correlation coefficients to determine whether correlations were significantly different from each other.

SUPPLEMENTARY S3.

To obtain an estimate of this difference, the female group was chosen as a reference group. Thus, since the female group was designated as the reference group, their factor means were fixed to zero, and we concentrated solely on estimates as they relate to the male group. Because analyses were based on the robust statistics, these estimates are interpreted in terms of robust standard errors and the resulting z-statistics. A positive CR implies that the comparison group has higher latent mean than the reference group. Conversely, a negative CR suggests that the comparison group's latent mean is smaller than the reference group (Byrne, 2006).

The means of F1 (Cognitive; females = 1.75; males = 1.21; CR = −3.085; small effect size, Cohen's d = −0.32) and F3 (Somatic; females = 2.83; males = 2.22; CR = −3.256; small effect size, Cohen's d = −0.33) for males were significantly different from those for females, the means for F2 (Affective; females = 1.78; males = 1.50; CR = −1.799; zero o near zero effect, Cohen's d = −0.17) were not.

SUPPLEMENTARY S4.

Analysis using Feldt's test (see Feldt, Woodruff, & Salih, 1987) indicating that the Cognitive scale scored higher reliabilities values (vs. Somatic: Feldt test = 0.7500, $p < 0.01$; vs. Affective: Feldt test = 0.8000, $p < 0.05$).

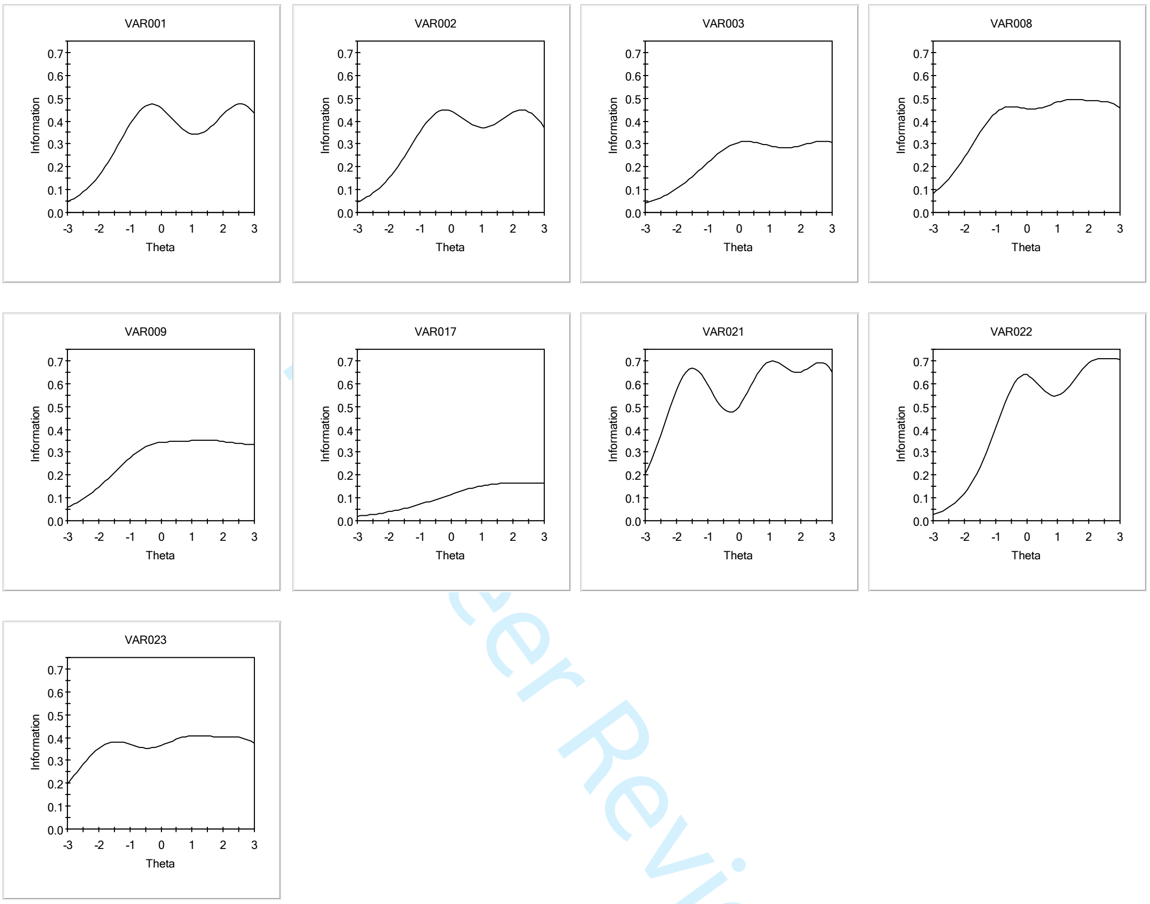
No item correlations with the scale, excluding the item itself, fall in the low range of 0.0-0.3, and discriminated well (Barbaranelli & Natali, 2005). The inter-correlations mean of items within each scale ranged from 0.58 (Cognitive) to 0.47 (Somatic).

SUPPLEMENTARY S5.

Response options with very low or negative threshold parameters are considered less useful items in measuring anxiety, as very low levels of anxiety would be needed to endorse these response options. In contrast, response options with extremely high threshold parameters would also be less useful, as extreme or atypical levels of anxiety would be needed to endorse these response options.

SUPPLEMENTARY FIGURE S1.

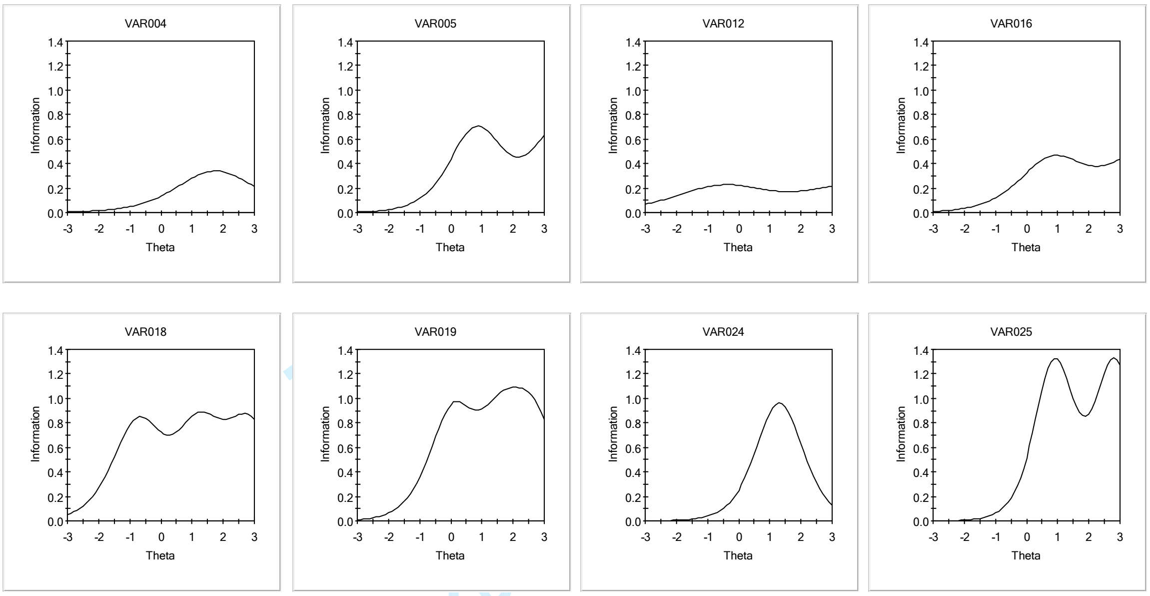
Item Information Function of the 9 items of the Somatic scale



Notes. Latent trait (Theta) was shown on the horizontal axis, and the amount of information yielded by the item at each trait level was shown on the vertical axis.

SUPPLEMENTARY FIGURE S2.

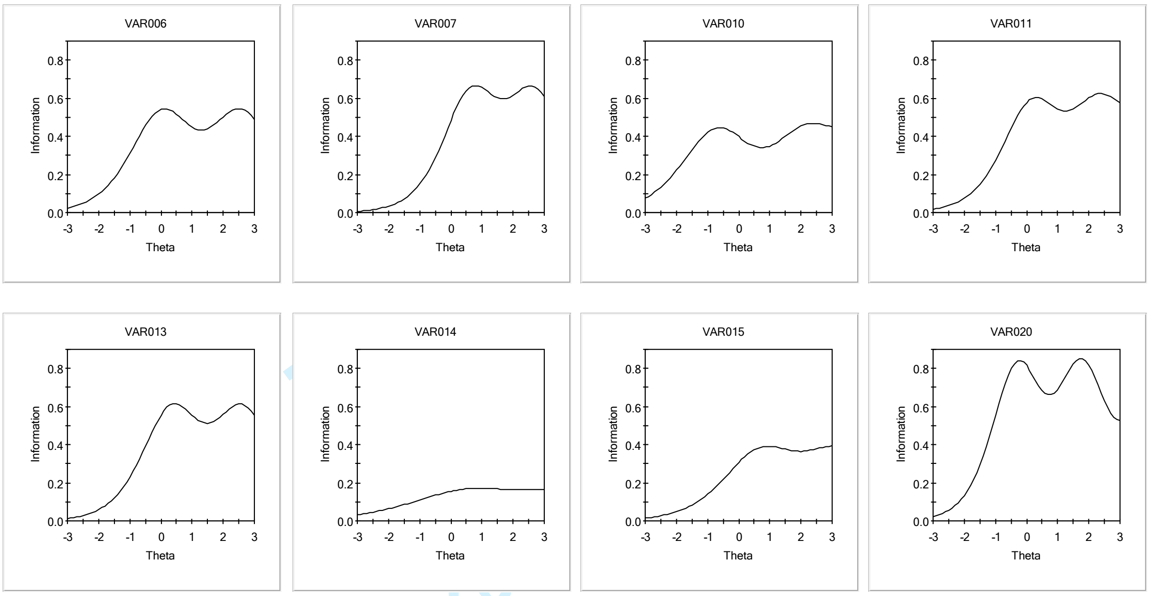
Item Information Function of the 9 items of the Cognitive scale



Notes. Latent trait (Theta) was shown on the horizontal axis, and the amount of information yielded by the item at each trait level was shown on the vertical axis.

SUPPLEMENTARY FIGURE S3.

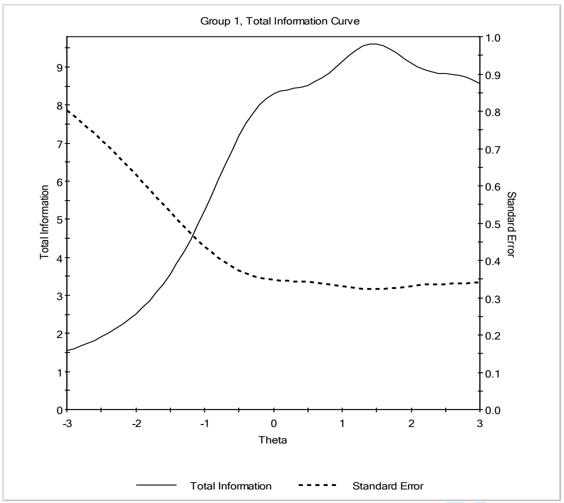
Item Information Function of the 8 items of the Affective scale



Notes. Latent trait (Theta) was shown on the horizontal axis, and the amount of information yielded by the item at each trait level was shown on the vertical axis.

SUPPLEMENTARY FIGURE S4.

Test Information Function



Notes. Test information function (TIF) for 25-item GAS. The dotted line represents standard error estimate (SEE), and the solid line represents the summative test information.

SUPPLEMENTARY TABLE S1.

Cut-off - GAS	Cut-off The likelihood of an anxiety disorder	Sensitivity (True positive)	1 – Specificity (False Positive)	AUC
Total score	18.5000	.723	.175	.825
Somatic	7.50	.795	.346	.794
Cognitive	3.50	.807	.313	.790
Affective	5.50	.651	.228	.760
Cut-off - GAS-12				
Total score	6.860	.809	.317	.790
Somatic	3.8202	.681	.248	.761
Cognitive	1.5000	.723	.332	.742
Affective	2.5000	.574	.206	.745

Table S1. Cut-off of likelihood of anxiety disorder; Sensitivity; 1- Specificity; AUC: Area under the curve