CrossMark

Factorial Validity of the Toronto Alexithymia Scale (TAS-20) in Clinical Samples: A Critical Examination of the Literature and a Psychometric Study in Anorexia Nervosa

Sandra Torres[1,2] · Marina P. Guerra[1,2] · Kylee Miller[3] · Patrício Costa[1,4,5] · Inês Cruz[1] · Filipa M. Vieira[1] · Isabel Brandão[6,7] · António Roma-Torres[6,7] · Magda Rocha[8]

Abstract

There is extensive use of the 20-item Toronto Alexithymia Scale (TAS-20) in research and clinical practice in anorexia nervosa (AN), though it is not empirically established in this population. This study aims to examine the factorial validity of the TAS-20 in a Portuguese AN sample ($N = 125$), testing four different models (ranging from 1 to 4 factors) that were identified in critical examination of existing factor analytic studies. Results of confirmatory factor analysis (CFA) suggested that the three-factor solution, measuring difficulty identifying (DIF) and describing feelings (DDF), and externally oriented thinking (EOT), was the best fitting model. The quality of measurement improves if two EOT items (16 and 18) are eliminated. Internal consistency of EOT was low and decreased with age. The results provide support for the factorial validity of the TAS-20 in AN. Nevertheless, the measurement of EOT requires some caution and may be problematic in AN adolescents.

✉ Sandra Torres
storres@fpce.up.pt

1 Faculty of Psychology and Education Sciences, University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal

2 Center for Psychology at University of Porto, Porto, Portugal

3 Carolina Institute for Developmental Disabilities, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA

4 Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga, Portugal

5 ICVS/3B's, PT Government Associate Laboratory, Braga/Guimarães, Portugal

6 Psychiatry Department, Hospital S. João, Porto, Portugal

7 Faculty of Medicine, University of Porto, Porto, Portugal

8 Center for Research in Health Technologies and Services - CINTESIS, Porto, Portugal

Introduction

Alexithymia is a well reported disturbance that affects the way individuals experience and express their emotions. It is characterized by a marked difficulty in consciously experiencing, identifying, and describing emotions, as well as reduced introspection. This multifaceted construct is currently accepted as a deficit in the cognitive processing of emotions with two major components: an affective component, which encompasses reduced emotional awareness, and a cognitive component, which encompasses a concrete and reality-based thinking style (Taylor, Bagby, & Parker, 1997). Alexithymia interferes with interpersonal relationships, as individuals exhibit deficits in understanding and relating to the feelings of others, in addition to their own. However, the consequences of this emotional deficit extend beyond interpersonal difficulties. The presence of alexithymia has been considered both to have a harmful effect on health and to play a key role in the expression of various psychosomatic and mental pathologies (De Gucht & Heiser, 2003; Taylor et al., 1997).

Currently, the most widely used method to assess alexithymia is the 20-item Toronto Alexithymia Scale (TAS-20), which was developed by Bagby, Parker, and Taylor (1994).

The English language version has a three-factor structure that corresponds to the theoretical construct of alexithymia: difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and externally oriented thinking (EOT). DIF refers to problems distinguishing between emotions and bodily sensations, as well as difficulties distinguishing between different emotions. DDF concerns the inability to verbally express emotions to others. EOT denotes an impoverished imaginative life marked by concrete and poor introspective thinking.

The TAS-20 has been cross-validated in different languages and cultures (cf. Taylor, Bagby, & Parker, 2003) and the three-factor structure has been replicated often, though support has not been universal. There are several studies which have failed to replicate this factor solution, suggesting that there might be alternative factor structures. These findings limit the understanding of whether the factorial structure of the TAS-20 varies across samples and, in particular, if patient populations have a different factor structure than non-patient populations. This issue is of major interest because the TAS-20 is widely used in clinical research and practice, though its factorial validity is not well-established in medical and psychological disorders (Kooiman, Spinhoven, & Trijsburg, 2002).

Examining the construct validity of the TAS-20 has an increased significance in the particular context of anorexia nervosa (AN). First, because alexithymia is a core symptom in this disorder and has been implicated in the severity of eating symptoms (Courty, Godart, Lalanne, & Berthoz, 2015). Being considered a negative prognostic indicator (Speranza, Loas, Guilbaud, & Corcos, 2011), it is assumed that assessing alexithymia can help the clinicians to scrutinize emotion regulation deficits and determine a better-suited treatment (Lumley, Neely, & Burger, 2007; Speranza et al., 2011). In addition, evidence for the validity of the TAS-20 has a potential impact on improving treatment by making it possible to determine how far each alexithymia dimension might be linked to the severity of eating symptoms and other clinical outcomes. It is noteworthy that alexithymia dimensions can act differently when compared with the measure as a whole (Taylor et al., 1997). Finally, we do not know if emotional avoidance and control that typically characterizes AN patients might influence the ability to self-assess the different facets of alexithymia (Oldershaw, Lavender, Sallis, Stahl, & Schmidt, 2015), thus giving inaccurate self-report.

Critical Examination of Factor Analytic Studies on the TAS-20 in Clinical Samples

The literature on factorial validity of the TAS-20 in specific populations with medical conditions or psychiatric problems (hereinafter referred to as clinical samples) is increasing in recent years. However, these studies have not been summarized and critically examined in a manner that allows us to draw conclusions and make recommendations for future research. Here we update findings of the factor analytic studies of the TAS-20, last done formally in 2003 (Taylor et al., 2003); and consider only those studies which used clinical samples (alone or in combination with non-clinical). In addition to this overview, we critically examine some methodological issues that can impact the quality and comparability of the data. More specifically, our aims here are to critically reflect upon the data preparation process, factorial analysis procedures, and study sample characteristics.

The studies using clinical samples are briefly described in the *supplementary material* (Tables S1 and S2), divided into those which confirm ($n = 10$) or contradict ($n = 10$) the three-factor structure proposed by Bagby et al. (1994). The first conclusion to be drawn is related precisely to the balance between these two categories. The relative variety of factor solutions stemmed from these studies makes the original structure questionable in clinical groups. Some studies found a combination of the factors DIF and DDF (Cleland, Magura, Foote, Rosenblum, & Kosanke, 2005; Guillén et al., 2014; Haviland & Reise, 1996; Kooiman et al., 2002) which can be influenced by the fact that communication of feelings to others requires introspection and the ability to identify those feelings in oneself (Bagby et al., 1994). Other researchers found a factor structure where the EOT dimension comprised two (Müller, Bühner, & Ellgring, 2003) or three factors (Pinaquy, Chabrol, & Barbe, 2002). No stable solutions were found by other studies (Koch et al., 2015; Richards, Fortune, Griffiths, & Main, 2005). Moreover, most reported a low internal consistency in EOT factor ($\alpha < .70$). Thus, the three main questions to be clarified are whether (1) DIF and DDF comprise two independent dimensions, (2) clinical data exhibit the EOT factor, and (3) difficulties in the EOT structure and consistency are related to some problematic items.

At the methodological level, there are some issues that should be highlighted. In terms of data preparation, only 30% ($n = 6$) of the studies have clearly reported that the TAS-20 had been examined for normality. Generally, the failure to meet the assumption of multivariate normality can lead to an overestimation of the chi-square statistic, being recommended in this case, the use of distribution-free methods (Powell & Schafer, 2001). We also found a lack of information regarding the extent of missing data in 75% ($n = 15$) of the studies. Not mentioning how missing data were addressed limits the relevance of the findings (Jackson, Gillaspy Jr, & Purc-Stephenson, 2009).

In terms of factorial analysis procedures, studies revealed different configurations varying between the use of exploratory factor analysis (EFA; $n = 4$), confirmatory factor analysis (CFA; $n = 10$), or both ($n = 6$). Within EFA,

we found that researchers used the component analysis extraction method ($n = 6$) and varimax rotation ($n = 3$), despite the use of these methods being discouraged (Costello & Osborne, 2005). Within CFA, choices have to be made concerning estimation method, fit indices, and cutoff criteria. In this overview we concluded that different cutoff criteria for fit indices were adopted. Interpretation of results can vary greatly depending on the stringency of criteria employed. In addition, only three studies that used CFA compared the proposed three-factor model with alternative solutions published in literature (Koch et al., 2015; Meganck, Vanheule, & Desmet, 2008; Müller et al., 2003), despite the recommendation of including alternative and theoretically plausible models (Jackson et al., 2009).

When looking at the characteristics of the tested samples we observed that several studies used mixed diagnostic groups instead of specific patient groups, reinforcing the perspective that there is a lack of studies that analyze the factor structure of the TAS-20 in selective clinical samples. More research using specific psychiatric diagnoses is needed since emotional instability can vary across psychiatric disorders. Particularly as emotional liability affects awareness and expression of emotions, thus making the most commonly utilized methodology of self-report unsuitable for assessing alexithymia facets (Koch et al., 2015).

Assessment of Alexithymia in Anorexia Nervosa

Alexithymia is a predominant factor in AN (Bourke, Taylor, Parker, & Bagby, 1992; Torres et al., 2015). There is a robust body of literature documenting that alexithymia levels are elevated in individuals with AN compared to healthy controls (for review see Caglar-Nazali et al., 2014; Nowakowski, McFarlane, & Cassin, 2013; Oldershaw et al., 2015; Torres, Guerra, Lencastre, Vieira et al., 2011; Westwood, Kerr-Gaffney, Stahl, & Tchanturia, 2017). The prevalence rates (established cutoff for alexithymic cases: TAS-20 ≥ 61; Taylor et al., 1997) in this eating disorder (ED) are high, reaching up to 50% in the majority of studies (Nowakowski et al., 2013). Based on these findings is the assumption that maladaptive eating behaviors can be used as a strategy to avoid or cope with feelings (e.g., Brockmeyer et al., 2012; Clinton, 2006; Wildes, Ringham, & Marcus, 2010). In fact, research data suggest that ED behaviors developed as a means to control emotional experiences (for review see Haynos & Fruzzetti, 2011), limiting the regulation of emotions through mental processes, as seen in alexithymia (Veríssimo, 2003). This evidence is in line with etiological models, which posit that emotional dysregulation

is implicated in AN development (e.g., Hatch et al., 2010; Southgate, Tchanturia, & Treasure, 2005; Treasure, Corfield, & Cardi, 2012).

Alexithymia also appears as a critical factor in the persistence of disorders in AN. Functional abnormalities associated with alexithymia might foster broader emotion processing deficits that are often observed in AN patients. Neuroimaging studies found functional differences in brain systems during the processing of unpleasant body image-related words (Miyake et al., 2009) and negative words concerning interpersonal relationships (Miyake et al., 2012), as a function of alexithymia in AN. Alexithymia proved to be closely linked to social difficulties, playing a key role in the relational isolation of these patients (Courty et al., 2015). Past research in AN also suggested that alexithymia is strongly related to anxiety and depression (Li, Zhang, Guo, & Zhang, 2015; Lulé et al., 2014). Although depression is probably the variable that best accounts for the variance in alexithymia (Parling, Mortazavi, & Ghaderi, 2010), it is not a complete explanation for cognitive–affective disturbances in this ED (Torres et al., 2015). In AN, alexithymia seems to be a trait feature or a consolidated change that is a consequence of the illness. Specifically, no association was found between alexithymia and several state variables such as age, body mass index (BMI), medication status, illness duration, treatment duration (Torres et al., 2015), and weight restoration (Beadle, Paradiso, Salerno, & McCormick, 2013).

This stability in alexithymic characteristics may have implications in AN prognosis including a poor response to psychological treatments, particularly those focusing on insight, emotional awareness, and a close alliance with a therapist (Lumley et al., 2007). Cognitive limitations in emotion regulation may also predispose one for the use of maladaptive eating behaviors in stressful situations, limiting the AN recovery (Speranza, Loas, Wallier, & Corcos, 2007).

Given the critical role of alexithymia in AN development and maintenance, the assessment of this construct is a common procedure in both clinical and research settings. However, we do not know if the TAS-20, as a self-report measure that requires insight, provides valid data in individuals exhibiting avoidance of emotions and poor reflective functioning, as is the case with AN (Oldershaw et al., 2015). By ascertaining the factorial validity of the TAS-20, we will better understand the validity of self-assessment of alexithymia as a multidimensional construct.

Until now, three studies are known to have investigated the validity of the TAS in samples with ED. The first published study was carried out by Troop, Schmidt, and Treasure (1995) with a mixed sample ($N = 127$) of patients with AN and bulimia nervosa (BN). These authors used a previous version of the instrument—the TAS-26 (Taylor, Ryan, & Bagby, 1985)—and found a four-factor structure: Inability to Identify Feelings, Paucity of Fantasy, Non-communication

of Feelings, and Concrete Thinking. In turn, Loas, Braun, Delhaye, and Linkowski (2001) used the most recent version of TAS, the TAS-20, and their results outlined the proposed three-factor model. Two factors, however, revealed low internal consistency: DDF ($\alpha = .61$) and EOT ($\alpha = .56$). The clinical sample assessed in this study was composed of a mixed group of patients with ED or addictive behaviors ($N = 659$). Lastly, Guillén et al., (2014) investigated the performance of the TAS-20 in patients with ED ($N = 103$) and found a set of 13 clinically interpretable items composed of DID and DDF items, plus item 8 from the EOT subscale, with a one-dimensional structure.

None of these studies conducted factor analysis by ED types. Despite a recently proposed transdiagnostic view of ED based on cross-diagnostic commonalities (Fairburn, Cooper, & Shafran, 2003), there are reasons to support the study of construct validation of the TAS-20 by diagnosis, including: (a) the nascent nature of the transdiagnostic model necessitating further investigation; (b) previous studies having observed differences in emotion processing between individuals with AN and BN (Gilboa-Schechtman, Avnon, Zubery, & Jeczmien, 2006; Pascual, Etxebarria, & Cruz, 2011; Sexton, Sunday, Hurt, & Halmi, 1998); (c) starvation and low BMI may impact emotional functioning (Westwood et al., 2017); and (d) individuals with AN usually seek treatment at younger ages (American Psychiatric Association, 2013), resulting in very different profiles in terms of cognitive and social maturity.

An additional issue to be explored is the factorial validity of the TAS-20 in the light of the new AN diagnostic criteria. The transition from DSM-IV (American Psychiatric Association, 2000) to DSM-5 (American Psychiatric Association, 2013) has resulted in several changes including removal of the amenorrhea criteria, broadening of the weight criteria, and greater emphasis on clinical impressions of patients' fear of weight gain if behaviors interfering with weight gain can be observed. This broader definition of the disorder has allowed the inclusion of atypical or subthreshold presentations (Brown, Holland, & Keel, 2014; Dahlgren & Wisting, 2016), whose differences in clinical picture and prognosis appear to increase phenotypic heterogeneity (Mustelin et al., 2016). This creates the need to investigate whether the underlying putative structure of alexithymia construct is reflected in the TAS-20, when applied to this more heterogeneous diagnostic category.

In sum, several reasons justify the relevance of determining the factorial solution of the TAS-20 in AN. Keeping in mind that impaired emotional functioning is an essential element in the AN genesis and prognosis, the use of this measure can be very useful to increase understanding of the mechanisms by which alexithymia influences clinical aspects of the condition. This line of research might benefit from the use of subscale scores rather than a global score,

attending to the multidimensional nature of the construct. Given the lack of such research in AN, this study sought to examine whether the original three-factor model of the TAS-20 could be replicated in a Portuguese sample of treatment-seeking patients with AN, in light of the new DSM-5 diagnostic criteria. In addition, we aimed to assess whether this factor solution provides a better fit to the scale compared with alternative factor structures that have been proposed in the literature (Koch et al., 2015; Meganck et al., 2008; Müller et al., 2003; Tsaousis et al., 2010; Zhu et al., 2007). This study predicted that the original three-factor solution was the model that best fits the data.

Method

Participants

A total of 125 female participants with AN, ranging between 13 and 40 years old ($M = 19.73$ years; $SD = 5.97$), were recruited from six public hospitals and two private clinics. Exclusion criteria included past or present psychotic disorders and illicit substance use or alcohol abuse. At the time of data collection they were all in active treatment, inpatient (29.6%; $n = 37$) or outpatient (70.4%; $n = 88$). Illness duration varied between 3 and 168 months ($M = 35.77$; $SD = 34.37$; Median = 24). Seventy-six patients were diagnosed as having AN restrictive subtype (AN-R) and 49 as having AN binge–purge subtype (AN-B). The participant's BMI ranged between 13.4 and 18.5 kg/m^2 ($M = 15.81$; $SD = 1.77$).

Materials

Interview for the Diagnosis of Eating Disorders-IV

Participants' diagnosis was established through the application of the Interview for the Diagnosis of Eating Disorders-IV (IDED-IV; Kutlesic, Williamson, Gleaves, Barbin, & Murphy-Eberenz, 1998; Torres et al., 2008). The IDED-IV is a semi-structured interview that was developed for the purpose of differential diagnosis of ED, based on DSM-IV-TR criteria (American Psychiatric Association, 2000). With the transition to DSM-5, the diagnostic threshold for AN was lowered, enabling inclusion of atypical or subthreshold cases previously diagnosed as eating disorder not otherwise specified (EDNOS). Thus, in order to adhere to the new definition of AN, we reanalyzed and recoded all the interviews of cases diagnosed with EDNOS, using the DSM-5 criteria (American Psychiatric Association, 2013), post hoc. The IDED-IV begins with a semi-structured overview of the participant's history of ED symptoms and current eating pattern. Descriptions of current eating patterns are obtained for both a typical

day and days in dieting. Information is also obtained regarding medical problems associated with ED (Kutlesic et al., 1998). This overview, together with the questions included in diagnostic section, allows for exploration of the reasons for maintaining a low weight other than fear of weight gain (e.g., somatic complaints, extreme need for control) and the behaviors intended to avoid weight gain (e.g., skipping meals, substantial caloric restriction). To recode diagnoses, we used this information to infer fear of weight gain without needing to re-interview participants, and thus expand the Criterion B as defined in DSM-5. In addition, we readjusted the maximum BMI threshold for determining low weight (Criterion A) to 18.5 kg/m², in accordance with the WHO definition of underweight (WHO, 1995), and removed the amenorrhea criteria. No changes have been made to Criterion C (body image disturbance, undue influence of weight or shape on self-evaluation, or the denial of seriousness of low weight). Two members of the team experienced in the diagnosis and treatment of ED recoded the interviews of EDNOS participants ($n = 31$). Agreement between AN diagnosis (yes/no) was evaluated using Cohen's kappa coefficient (κ). We found high interrater agreement ($\kappa = .83$). The new sample of AN participants included all DSM-IV-TR cases ($n = 103$) as well as new cases that were previously diagnosed as EDNOS ($n = 22$) now encompassed under the DSM-5 criteria.

20-Item Toronto Alexithymia Scale (TAS-20)

The TAS-20 is a self-report scale comprising 20 items that respondents rate on a five-point Likert type response format ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The first factor (DIF) in the three-factor model for the TAS-20 consists of seven items assessing the difficulty in identifying feelings and distinguishes them from the somatic sensations that accompany emotional arousal. Factor 2 (DDF) consists of five items assessing the difficulty in describing feelings to others. Factor 3 (EOT) consists of eight items assessing externally oriented thinking. Five items are negatively keyed (Bagby et al., 1994). We used a validated Portuguese version of the TAS-20, with the same number of items as the original version and good psychometric properties (Prazeres, Parker, & Taylor, 2000). This version underwent the gold-standard translation and back-translation process to establish cross-language equivalence, involving eight translators (three of them bilingual) and the authors of the original English version. Factor structure cross-validation was tested in two non-clinical samples (normal adults and university students), replicating the three-factor model (Prazeres et al., 2000). This model was also confirmed by Veríssimo (2001) in Portuguese clinical samples (outpatients attending a routine general practice consultation and patients with inflammatory bowel disease).

Procedure

The study received ethical approval from all the hospitals and clinics involved in sample recruitment. Patients participated voluntarily after signing informed consent. For subjects younger than 18, informed written consent was provided both by a parent and assent of the participant. All participants agreed to participate in the study.

Data Analysis and Tested Models

No missing data were found and no data transformations were carried out. Confirmatory Factor Analyses (CFA) were accomplished using Structural Equation Modeling with EQS version 6.1. The four models tested were as follows:

(a) Model I: A one-factor model, where it is assumed that all items reflect a one-dimensional construct (Lambert et al., 1999);

(b) Model II (DIDF–EOT): A two-factor model with DIF and DDF forming one factor (DIDF: Items 1, 2, 3, 4, 6, 7, 9, 11, 12, 13, 14, and 17) and EOT as the second factor (EOT: Items 5, 8, 10, 15, 16, 18, 19, and 20; Cleland et al., 2005; Erni, Lötscher, & Modestin, 1997; Loas, Otmani, Verrier, Fremaux, & Marchand, 1996);

(c) Model III (DIF–DDF–EOT): the common, three-factor solution reported by Bagby et al. (1994): DIF (Items 1, 3, 6, 7, 9, 13, and 14), DDF (Items 2, 4, 11, 12, and 17), and EOT (Items 5, 8, 10, 15, 16, 18, 19, and 20);

(d) Model IV (DIF–DDF–PT–LIE): a four-factor solution, with DIF, DDF, and EOT split into two factors: "pragmatic thinking" (PT; refers to a concrete and reality-based cognitive style; items 5, 8, and 20) and "lack of importance of emotions" (LIE; concerns the low importance placed on emotional experiences; items 10, 15, 16, 18, and 19; Müller et al., 2003).

In evaluating the model fit, the following indices were considered: the Satorra–Bentler scaled chi-square (Satorra & Bentler, 2001) and the corresponding chi-square to degrees of freedom ratio, the Comparative Fit Index (CFI; Bentler, 1990), the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), and the standardized root mean residual (SRMR; Bentler, 1995). The following criteria were used as standards of acceptable fit: non-significant chi-squared (χ^2) test ($p > .05$; Barrett, 2007), χ^2/df < 2 (Wheaton, 1987), CFI > .90 (Byrne, 2010), RMSEA < .06, and SRMR < .08 (Hu & Bentler, 1999). Models were compared using the Akaike's information criterion (AIC; Akaike, 1987) and the consistent Akaike's information criterion (CAIC; Bozdogan, 1987; Dayton, 2003), the latter because it is based on the sample size in order to compensate for the overestimation of AIC (Acquash, 2013). The

model that presents the lowest values of AIC and CAIC is the one that is more representative of the true model or "the best approximation model among those being considered" (Dayton, 2003, p. 284).

The Cronbach's alpha coefficient was calculated to assess the internal consistency of each factor. Item-to-scale homogeneity was also evaluated by calculating mean inter-item correlations (MIIC), but only in the best fitting model. A value > 0.7 for the Cronbach's α coefficient and between 0.2 and 0.4 for the MIIC were required (Briggs & Cheek, 1986).

In all analyses, p values of < .05 were considered statistically significant.

Results

Preliminary Analyses

Descriptive statistics and item intercorrelations were calculated (supplementary material, Tables S3 and S4, respectively). To test the distribution of data we used the Kolmogorov–Smirnov test. A significant p value was found but a close inspection of items' skewness and kurtosis indicated that the deviation from normality was not problematic since absolute values were lower than 2.0 (Schumacker & Lomax, 2016).

The sample was composed of 59 adolescents (ages 13–17; 47.2%) and 66 adults (ages 18–40; 58.2%). We defined the cutoff point between age groups as age 18, based on the definition of adolescence—a stage of physical and psychological development that occurs during the period from puberty to legal adulthood (Dahl, 2004). The prevalence of alexithymia in adolescents was 57.6% ($n = 34$) and in adults was 66.7% ($n = 44$). According to t tests, there were no significant differences between age groups in TAS-20 mean scores, $t(123) = -1.80$, $p = .074$, $d = -0.32$. No mean differences were found in BMI, $t(123) = 1.17$, $p = .234$, $d = 1.30$. There was no significant difference between the two AN subtypes (AN-R and AN-B) on the TAS-20, $t(123) = 0.77$, $p = .444$, $d = 1.19$, and BMI, $t(123) = 1.15$, $p = .252$, $d = 1.32$.

Confirmatory Factor Analyses

Despite the high probability of normally distributed data, the conservative robust maximum likelihood estimation method was used (Ullman, 2006). Tested models were specified as presenting reflective latent factors correlated among them (exception made to the one-dimensional factor), without allowing for correlations between indicator's errors. For establishing the scale factor (metric), the first indicator of each latent factor was set to one and all

the other indicators were freely estimated. There were no cross-loadings in models. The fit indices of the four models are presented in Table 1 (see "Original models" data). All models showed poor robust fit indices. Model I provided the worst fits, followed by Model II. In Models III and IV the values were more acceptable, although below the cutoff values for the presented fit indexes. Despite the very similar results between these two models, the AIC and CAIC values indicated that Model III offered a better fit to the data.

As none of the four tested models met values for acceptable fit, we hypothesized that this poor performance might be related to some problematic items, as stated in the critical examination of factor analytic studies presented in this paper. When analyzing the standardized factor loadings we verified that items 16 and 20 repeatedly showed low factor loadings (bellow .06). Thus, to clarify if problematic items have a significant impact on the structure of the instrument, we retested the models removing the zero-parameter variables detected by Wald test. In Model I, the Wald test indicated that eight items could be removed without affecting the chi-square in a meaningful way (items 5, 8, 10, 15, 16, 18, 19, and 20; χ^2 between .21 and 1.00, $p > .05$). Following the same procedure, items 16 and 20 were eliminated from the other three models (item 16: χ^2 between .05 and .25; item 20: χ^2 between .10 and 3.83; $p > .05$).

The standardized factor loadings of final models are presented in Table 2. Items 5 and 15 had very low factor loadings in every model. In Models II and III, item loadings in the EOT latent factor were relatively unbalanced (between .24 and .65), and five out six items showed loadings lower than .50, indicating that < 25% of the result of those items were explained by the latent dimension. Globally, the removal of items indicated by Wald test resulted in an improvement in the fit of all models (see Table 1, "Final models" data). Despite this improvement, only Models III and IV had sufficient fit with the data. In what concerns Model IV, it should be noted, however, that the PT factor is only composed of two indicators, below the number of items per factor required in CFA analysis, which may be a primary limitation to the model adequacy. The combined fit indices for the CFA supported the three-factor structure tested in final Model III (SBχ^2/df = 1.40; CFI = .90; RMSEA [HI95%] = .06 [.08]; SRMR = .08). This model also presented the lowest values of AIC and CAIC, which indicates that it is the one that best served the data of this sample.

For final models, intercorrelations among factors were examined using Pearson's r (Table 3). All correlation coefficients were positive and significant ($p < .05$). Strong associations were found between DIF and DDF ($r = .72$) and PT and LIE ($r = .83$). All other correlations ranged from weak to moderate (r between .27 and .33).

Table 1 Model fit indexes for confirmatory factor analysis ($N = 125$)

	Tested parameters	$SB\chi^2$	$SB\chi^2$/df	CFI	RMSEA	90% C.I._RMSEA	SRMR	AIC	CAIC
Original models[a]									
Model I (1-Factor)	61	284.31***	1.67	.78	.07	[.06, .09]	.09	− 55.69	− 706.50
Model II (DIDF–EOT)	63	262.34***	1.55	.82	.07	[.05, .08]	.09	− 75.66	− 722.64
Model III (DIF–DDF–EOT)	66	233.27**	1.40	.87	.06	[.04, .07]	.09	− 100.73	− 740.06
Model IV (DIF–DDF–PT–LIE)	70	232.85***	1.42	.87	.06	[.04, .07]	.09	− 95.15	− 722.99
Final models (Wald test)[b]									
Model I (1-Factor)	47	100.90***	1.87	.89	.08	[.06, .11]	.08	− 7.10	− 213.33
Model II (DIDF–EOT)	57	209.69***	1.56	.85	.07	[.05, .08]	.08	− 58.31	− 209.69
Model III (DIF–DDF–EOT)	60	180.53**	1.37	.90	.05	[.03, .07]	.08	− 83.47	− 588.81
Model IV (DIF–DDF–PT–LIE)	64	180.48**	1.40	.90	.06	[.04, .08]	.08	− 77.52	− 571.38

DIDF difficulty in identifying and describing feelings, *EOT* externally oriented thinking, *DIF* difficulty in identifying feelings, *DDF* difficulty in describing feelings to others, *PT* pragmatic thinking, *LIE* lack of importance of emotions, $SB\chi^2$ Satorra–Bentler scaled chi-squared, $SB\chi^2$/df chi-squared to degrees of freedom ratio, *CFI* comparative fit index, *RMSEA* root mean square error of approximation, *90% C.I._RMSEA* root mean square error of approximation at 90% confidence interval, *SRMR* standardized root mean residual, *AIC* Akaike's information criterion, *CAIC* consistent Akaike's information criterion

**p < .01; ***p < .001

[a]Original models without removing or adding correlation between errors

[b]Models with eliminated parameters suggested in Wald test

Table 2 Confirmatory factor analysis: standardized factor loadings and measurement error for single items of the TAS-20 for final models

	Model I 1-Factor β	ε	Model II DIDF β	ε	EOT β	ε	Model III DIF β	ε	DDF β	ε	EOT β	ε	Model IV DIF β	ε	DDF β	ε	PT β	ε	LIE β	ε
1	.59	.81	.59	.81			.55	.83					.55	.83						
2	.60	.80	.60	.80					.67	.74					.67	.74				
3	.59	.81	.59	.81			.60	.80					.60	.80						
4	.53	.85	.53	.85					.65	.76					.64	.77				
5					**.24**	.97					**.24**	.97					**.28**	.96		
6	.54	.84	.54	.84			.55	.83					.55	.83						
7	.52	.86	.52	.85			.55	.84					.55	.84						
8					.35	.94					.35	.94					.40	.92		
9	.76	.65	.75	.66			.75	.66					.75	.66						
10					.44	.90					.42	.91							.41	.91
11	.58	.81	.58	.81					.71	.71					.71	.71				
12	.48	.88	.48	.88					.53	.85					.53	.85				
13	.82	.57	.83	.56			.87	.50					.87	.50						
14	.73	.68	.74	.67			.76	.65					.76	.65						
15					**.26**	.96					**.24**	.97							**.21**	.98
17	.48	.88	.47	.88					.59	.81					.59	.81				
18					.39	.92					.39	.92							.38	.92
19					.62	.79					.65	.76							.69	.72

DIDF difficulty in identifying and describing feelings, *EOT* externally oriented thinking, *DIF* difficulty in identifying feelings, *DDF* difficulty in describing feelings to others, *PT* pragmatic thinking, *LIE* lack of importance of emotions, *β* standardized factor loading, *ε* measurement error

Factor loadings smaller or equal to .35 are shown in bold type

Table 3 Correlation values between final models' latent factors

Model	Factors			
II		EOT		
	DIDF	**.29**		
III		DDF	EOT	
	DIF	**.72**	**.30**	
	DDF		**.31**	
IV		DDF	PT	LIE
	DIF	**.72**	**.31**	**.29**
	DDF		**.33**	**.27**
	PT			**.83**

DIDF difficulty in identifying and describing feelings, *EOT* externally oriented thinking, *DIF* difficulty in identifying feelings, *DDF* difficulty in describing feelings to others, *PT* pragmatic thinking, *LIE* lack of importance of emotions

Significant *p* values ($p < .05$) are in boldface

Reliability

In original models, the internal consistency was acceptable for Model I ($\alpha = .82$). The subscales related to difficulty identifying and describing feelings also evidenced an acceptable reliability both considering the two factors together (DIDF, $\alpha = .87$) or separated (DIF, $\alpha = .84$; DDF, $\alpha = .77$). Cronbach's alpha was unacceptably low for EOT ($\alpha = .46$) in Models II and III, and even lower in factors related to the thinking style in Model IV ($\alpha = .30$ and $\alpha = .33$, for PT and LIE, respectively).

Items' removal did not result in significant changes in internal consistency. An adequate reliability remained in the unifactorial model ($\alpha = .82$). A poor internal consistency persisted in PT ($\alpha = .19$) and LIE ($\alpha = .43$) factors. Although the EOT alpha has increased to .49, it remains low. MIIC were calculated for each factor of final Model III and only in the EOT a value below of the minimum accepted was observed (EOT, MIIC = .14; DIF, MIIC = .43; DDF, MIIC = .39). It must be taken into account that the reliability of the EOT subscale can decrease with age (Loas et al., 2017). In order to explore this hypothesis in our sample, we calculated the EOT internal consistency separately by age group. A lower value was found in adolescents ($\alpha = .39$) relative to adults ($\alpha = .55$). This trend was not observed in the DIF (adolescents, $\alpha = .87$; adults, $\alpha = .79$) and DDF (adolescents, $\alpha = .82$; adults, $\alpha = .70$) factors.

Discussion

In this study we tested four-factor competing models for the TAS-20 by means of CFA based on a clinical sample of patients with AN. We found that, although the original three-factor solution (Model III) was the model that best fit the data, globally all models tested with the full scale provided a poor fit to the data. Thus, the standard three-factor structure that was replicated in previous Portuguese validation studies (Prazeres et al., 2000; Veríssimo, 2001) was not supported in this sample. When analyzing the standardized factor loadings we verified that some items repeatedly yielded very low values. Then, we tested non-nested models where poorly performing items were removed. Based on the Wald test modification index, two items (16 and 20) were eliminated from Models II, III, and IV and all EOT items from Model I. The results showed that Model III without items 16 and 20 has an adequate fit and is the most suitable model.

Together, our results helped to clarify three questions that have arisen from the synthesis of findings of factor analytic studies of the TAS-20 in clinical samples held in this article, whether: (a) DIF and DDF comprise two independent dimensions, (b) clinical data exhibit the third factor, and (c) difficulties in the EOT structure and consistency are related to some problematic items.

Regarding the first question, our data supported the independence of DIF and DDF in AN samples. In Model III, the best fitting model, they are two separated, strongly correlated ($r = .72$) factors, as expected, but without multicollinearity. This result is in agreement with select previous studies (e.g., all studies described in Table S1), but contradictory findings also exist (e.g., the majority of studies described in Table S2). It should be noted, however, that this factor solution with DIF and DDF as independent factors is substantially different from the one-dimensional structure reported by Guillén et al. (2014) with ED patients. The adequate reliability of DIF and DDF also reinforces the notion that these two concepts can be distinguished from each other, as indicated by other studies using clinical samples (Maggini & Raballo, 2004; Müller et al., 2003; Pinaquy et al., 2002). This leads us to conclude that, even though verbalization and differentiation of feelings are interconnected, they do not necessarily act together in AN. In other words, there may be situations where, despite individuals' ability to acknowledge their emotional states, emotions are difficult to express. This incongruence between emotion expression and inner experience was documented by Gramaglia et al. (2016) and converges with Fox and Power's (2009) position that the non-expression of feelings in ED is not entirely explained by reduced emotional awareness, but is often due to an intense fear of dealing with emotions that are perceived to be overwhelming to the self. On the other hand, it is also plausible to assume that individuals with high DIF scores may be able to express emotions, as previously reported in AN (Torres, Guerra, Lencastre, Roma-Torres et al., 2011). In these cases, the emotional expression relies on the differentiation between positive and negative emotions, without distinguishing the experiential dimension of each emotion. This perspective is of particular relevance to understand how

emotions can be expressed by individuals that typically present poor interoceptive awareness, i.e., an impaired perception of both emotional states and sensations of hunger and satiety (Myers & Crowther, 2008).

With respect to the second question, as to whether clinical data exhibit the EOT dimension, our conclusion is that EOT is a salient component, as the results of the CFA clearly indicate relative superiority of the three-factor structure. In addition, and relating to the third question, our study provides evidence that the well-documented problems in literature relative to the EOT structure and internal consistency might be partly due to the poor performance of some items. In our sample, item 16 ("I prefer to watch 'light' entertainment shows rather than psychological dramas") yielded a very low parameter estimate and is commensurate with findings from other studies (Ling, Zeng, Yuan, & Zhong, 2016; Müller et al., 2003; Parker, Shaughnessy, Wood, Majeski, & Eastabrook, 2005; Simonsson-Sarnecki et al., 2000; Zhu et al., 2007). This also applies to item 20 ("Looking for hidden meanings in movies or plays distracts from their enjoyment"), in which the low contribution to the EOT factor shown is consistent with previous research (Koch et al., 2015; Müller et al., 2003; Zech, Luminet, Rimé, & Wagner, 1999). In the case of item 20, a low factor loading was also found in the Portuguese validation study of the scale with a non-clinical sample (Prazeres et al., 2000).

In practice, low factor loadings suggest that items are poorly related to the underlying construct. That can happen if item content does not represent the construct or if its interpretation is not clear. In the case of items 16 and 20 we believe that they can be problematic due to contextual complexity and polysemy (Kooiman et al., 2002; Moriguchi et al., 2007), in particular with item 16, and the use of unfamiliar terms (Parker, Eastabrook, Keefer, & Wood, 2010), such as "psychological dramas" in item 20. For this reason, we believe that the quality of measurement of the EOT factor could be improved if items that repeatedly presented weak loadings in several studies are reviewed, as suggested by Meganck et al. (2008). We recommend that these items should be rewritten using simpler and clearer language, in order to increase item comprehension. At a syntactic level, we suggest the use of shorter sentences, not requiring the balance between two facets or double entendre. We believe these changes would be beneficial, not only for the use of the TAS-20 in AN, but also in other settings, since our problems concerning this third factor have been repeatedly replicated.

This procedure might not be enough to improve the quality of the factor. It should be noted that, according to our results, even removing problematic items the EOT factor remains below the optimal internal consistency ($\alpha = .49$). This finding is consistent with other Portuguese validation studies, in which only one of five samples achieved an acceptable value (Prazeres et al., 2000; Veríssimo, 2001).

This problem in reliability is so widespread in literature that it begs attention to the characteristics of the instrument. One of the criticisms most mentioned on the TAS-20 is the disproportionate number of negatively keyed items (four out of five) which are designed to load onto the EOT factor. Although important for avoiding response set bias, negatively worded items create increased complexity, thus making the scale vulnerable to response tendencies (Loas et al., 2001). This issue may be more salient with alexithymic respondents, as they may have decreased mental flexibility (Kojima, Frasure-Smith, & Lespérance, 2001; Richards et al., 2005), particularly pertinent for AN samples, in which a high prevalence of alexithymia is expected. Preliminary data support the presence of a method effect induced by these items (Meganck et al., 2008). However, far much more research is required to elucidate the nature of wording effects associated with the measure. As previously done with other psychological measures, such as the Rosenberg Self-Esteem Scale, future studies should determine if this measurement procedure contributes variance to scores beyond what is attributable to variance in the construct (Salerno, Ingoglia, & Lo Coco, 2017). In addition, it would be pertinent to examine the stability of method effects over time, using a longitudinal methodological approach (Marsh, Scalas, & Nagengast, 2010).

There are two additional speculations to explain the low reliability of the EOT subscale. One is relative to the inclusion of young subjects in the sample. Despite the use of the standard three-factor model with adolescents in previous studies, it was also found that the reliability of the EOT subscale, in particular, can decrease with age (for review see Loas et al., 2017). Findings from our study corroborate this data by the fact that a lower value of internal consistency was observed in the adolescent group ($\alpha = .39$). Nevertheless, this argument can only partially explains the lack of item homogeneity in the EOT subscale, as adults still presented a suboptimal level of reliability ($\alpha = .55$). The other argument is related to the nature of the EOT construct. This factor concerns the tendency to leave emotions unanalyzed and this avoidant attitude might be difficult to capture when it is not entirely conscious.

We considered, however, that the low internal reliability of the EOT factor does not preclude the validity of the three-factor model of the TAS-20 in our sample. It should be noted that "the factor structure of a set of indicators and the internal consistency of the factor scales consisting of the summation of a set of indicators within those factors are distinct (albeit related) issues" (Tsaousis et al., 2010, p. 447). In practice, this means that the assessment of alexithymia in AN can reliably be made by means of self-report. Results from factorial validity analyses provide some evidence that participants are able to reflect on their emotions. This perspective converges with previous research in documenting

that meta-emotional abilities are preserved in AN despite high levels of alexithymia (Torres, Guerra, Lencastre, Roma-Torres et al., 2011). However, some caution should be taken when applying the EOT scale. We recommend that investigation of the different facets of the construct should be supported by other assessment tools, such as the Toronto Structured Interview for Alexithymia (TSIA), which is found to be a more sensitive instrument in detecting EOT in adolescent patients with AN (Balottin, Nacinovich, Bomba, & Mannarini, 2014).

Limitations of the present study are primarily related to the study sample. The sample size of 125 can be considered small for CFA in general populations. However, given the prevalence rate of AN, the sample size is relatively large. Although we recognize that a larger sample could lead to more robust results, research to determine sample size for structural equation modeling applications, particularly within select clinical populations, has been inconclusive (e.g., MacCallum, Roznowski, & Necowitz, 1992; Tanaka, 1987). Various sample size guidelines have been proposed (e.g., 50 observations per variable, no < 100 observations total, 5–10 observations per parameter). The study by Hamilton, Gagne, and Hancock (2003) even suggested that sample size may not bias the parameter estimates to a substantive degree, and recommended the use of samples of at least 100. Two other sample-related limitations concern age range, including different levels of cognitive and emotional maturity, and the impossibility to test measurement invariance between age groups or diagnostic subclasses of AN. In terms of diagnosis, it is still undetermined whether there is a transdiagnostic factor structure of TAS-20 in ED. While not the focus of this paper, we suggest further exploration in future studies. Similarly, consideration should be given to the possibility of the presence or absence of binge eating/purging behaviors to be a critical variable to differentiate ED types on alexithymia. Very recent studies in emotion regulation deficits tend in this direction (Mallorquí-Bagué et al., 2017; Weinbach, Sher, & Bohon, 2017). An additional procedure would be recommended in comparing CFA results between AN and a non-clinical sample. Considering that we used a translated version of the TAS-20, this procedure would allow exploring whether the poor results on factorial structures achieved without items removed may be due to specific characteristics of AN patients or issues related to cross-cultural equivalence in construct operationalization. Given the impact of cross-cultural issues of the alexithymia construct, this point could add importance to the clinical utility and implication of the findings of the current study. Lastly, we stress the absence of a more comprehensive study of construct validity, including convergent and discriminant validity assessment, which could have been more informative about the degree to which the EOT subscale, in particular, indeed measures what it purports to be measuring.

Conclusions

Confirmatory factor analysis was used to determine the number and nature of the factors underlying the TAS-20 in a sample of patients diagnosed with AN, according the recent DSM-5 criteria. The three-factor structure of the TAS-20 was confirmed. Its quality of measurement improves if two items designed to load onto the EOT factor, namely item 16 and 20, are removed. The internal consistency of the EOT is demonstrated to be lower than that of the other TAS-20 scales, which is in accordance with a large body of research literature. Difficulties in item comprehension may be at the root of these problematic items, not only in AN but in other clinical and non-clinical samples. Contrary to DID and DDF, the EOT is a dimension of alexithymia that can be problematic when subjects are young.

Compliance with Ethical Standards

Conflict of interest The authors Sandra Torres, Marina Prista Guerra, Kylee Miller, Patrício Costa, Inês Cruz, Filipa M. Vieira, Isabel Brandão, António Roma-Torres, and Magda Rocha declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent All procedures were in accordance with the ethical standards of the institutional research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

Acquash, H. D. (2013). On the comparison of Akaike information criterion and consistent Akaike information criterion in selection of asymmetric price relationships: Bootstrap simulation results. *Agris On-line Papers in Economics and Informatics, 5*, 3–9.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332. https://doi.org/10.1007/bf02294359.

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

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23–32. https://doi.org/10.1016/0022-3999(94)90005-1.

Balottin, L., Nacinovich, R., Bomba, M., & Mannarini, S. (2014). Alexithymia in parents and adolescent anorexic daughters: Comparing the responses to TSIA and TAS-20 scales. *Neuropsychiatric Disease and Treatment, 10*, 1941–1951. https://doi.org/10.2147/NDT.S67642.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*, 815–824. https://doi.org/10.1016/j.paid.2006.09.018.

Beadle, J. N., Paradiso, S., Salerno, A., & McCormick, L. M. (2013). Alexithymia, emotional empathy, and self-regulation in anorexia aervosa. *Annals of Clinical Psychiatry, 25*, 107–120.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Besharat, M. A., Rostami, R., Pourhosein, R., & Mirzamani, M. (2006). Assessing reliability and validity of Farsi version of the Toronto Alexithymia Scale-20 in a sample of opioid substance use disordered patients. *Iranian Journal of Psychiatry, 1*, 133–139.

Bourke, M. P., Taylor, G. J., Parker, J. D. A., & Bagby, R. M. (1992). Alexithymia in women with anorexia nervosa: A preliminary investigation. *British Journal of Psychiatry, 161*, 240–243.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345–370.

Bressi, C., Taylor, G., Parker, J., Bressi, S., Brambilla, V., Aguglia, E., … Invernizzi, G. (1996). Cross validation of the factor structure of the 20-item Toronto Alexithymia Scale: An Italian multicenter study. *Journal of Psychosomatic Research, 41*, 551–559. https://doi.org/10.1016/S0022-3999(96)00228-0.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106–148. https://doi.org/10.1111/j.1467-6494.1986.tb00391.x.

Brockmeyer, T., Holtforth, M. G., Bents, H., Kämmerer, A., Herzog, W., & Friederich, H.-C. (2012). Starvation and emotion regulation in anorexia nervosa. *Comprehensive Psychiatry, 53*, 496–501. https://doi.org/10.1016/j.comppsych.2011.09.003.

Brown, T. A., Holland, L. A., & Keel, P. K. (2014). Comparing operational definitions of DSM-5 anorexia nervosa for research contexts. *International Journal of Eating Disorders, 47*, 76–84. https://doi.org/10.1002/eat.22184.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Byrne, B. M. (2010). Structural equation modelling with AMOS: Basic concepts, applications, and programming (2nd ed.). New York: Routledge.

Caglar-Nazali, H. P., Corfield, F., Cardi, V., Ambwani, S., Leppanen, J., Olabintan, O., … Treasure, J. (2014). A systematic review and meta-analysis of 'Systems for Social Processes' in eating disorders. *Neuroscience and Biobehavioral Reviews, 42*, 55–92. https://doi.org/10.1016/j.neubiorev.2013.12.002.

Cleland, C., Magura, S., Foote, J., Rosenblum, A., & Kosanke, N. (2005). Psychometric properties of the Toronto Alexithymia Scale (TAS-20) for substance users. *Journal of Psychosomatic Research, 58*, 299–306. https://doi.org/10.1016/j.jpsychores.2004.11.002.

Clinton, D. (2006). Affect regulation, object relations and the central symptoms of eating disorders. *European Eating Disorders Review, 14*, 203–211. https://doi.org/10.1002/erv.710.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*, 1–9.

Courty, A., Godart, N., Lalanne, C., & Berthoz, S. (2015). Alexithymia, a compounding factor for eating and social avoidance symptoms in anorexia nervosa. *Comprehensive Psychiatry, 56*, 217–228. https://doi.org/10.1016/j.comppsych.2014.09.011.

Dahl, R. E. (2004). Adolescent brain development: A period of vulnerabilities and opportunities. Keynote address. *Annals of the New York Academy of Sciences, 1021*, 1–22. https://doi.org/10.1196/annals.1308.001.

Dahlgren, C. L., & Wisting, L. (2016). Transitioning from DSM-IV to DSM-5: A systematic review of eating disorder prevalence assessment. *International Journal of Eating Disorders, 49*, 975–997. https://doi.org/10.1002/eat.22596.

Dayton, C. M. (2003). Model comparisons using information measures. *Journal of Modern Applied Statistical Models, 2*, 281–292.

De Gucht, V., & Heiser, W. (2003). Alexithymia and somatisation: A quantitative review of the literature. *Journal of Psychosomatic Research, 54*, 425–434. https://doi.org/10.1016/S0022-3999(02)00467-1.

Erni, T., Lötscher, K., & Modestin, J. (1997). Two-factor solution of the 20-item Toronto Alexithymia Scale confirmed. *Psychopathology, 30*, 335–340. https://doi.org/10.1159/000285079.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509–528. https://doi.org/10.1016/S0005-7967(02)00088-8.

Fernández-Jiménez, E., Pérez-San-Gregorio, M., Taylor, G. J., Bagby, R. M., Ayears, L. E., & Izquierdo, G. (2013). Psychometric properties of a revised Spanish 20-item Toronto Alexithymia Scale adaptation in multiple sclerosis patients. *International Journal of Clinical Health & Psychology, 13*, 226–234. https://doi.org/10.1016/S1697-2600(13)70027-9.

Fox, J. R. E., & Power, M. J. (2009). Eating disorders and multi-level models of emotion: An integrated model. *Clinical Psychology and Psychotherapy, 16*, 240–267. https://doi.org/10.1002/cpp.626.

Gilboa-Schechtman, E., Avnon, L., Zubery, E., & Jeczmien, P. (2006). Emotional processing in eating disorders: Specific impairment or general distress related deficiency? *Depression and Anxiety, 23*, 331–339. https://doi.org/10.1002/da.20163.

Gramaglia, C., Ressico, F., Gambaro, E., Palazzolo, A., Mazzarino, M., Bert, F., … Zeppegno, P. (2016). Alexithymia, empathy, emotion identification and social inference in anorexia nervosa: A case-control study. *Eating Behaviors, 22*, 46–50. https://doi.org/10.1016/j.eatbeh.2016.03.028.

Guillén, V., Santos, B., Muñoz, P., Fernández de Corres, B., Fernández, E., Pérez, I., … González-Pinto, A. (2014). Toronto Alexithymia Scale for patients with eating disorder: Of performance using the non-parametric item response theory. *Comprehensive Psychiatry, 55*, 1285–1291. https://doi.org/10.1016/j.comppsych.2014.03.020.

Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003). *The effect of sample size on latent growth models*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Hatch, A., Madden, S., Kohn, M., Clarke, S., Touyz, S., & Williams, L. M. (2010). Anorexia nervosa: Towards an integrative neuroscience model. *European Eating Disorders Review, 18*, 165–179. https://doi.org/10.1002/erv.974.

Haviland, M. G., & Reise, S. P. (1996). Structure of the twenty-item Toronto Alexithymia Scale. *Journal of Personality Assessment, 66*, 116.

Haynos, A. F., & Fruzzetti, A. E. (2011). Anorexia nervosa as a disorder of emotion dysregulation: Evidence and treatment implications. *Clinical Psychology: Science and Practice, 18*, 183–202. https://doi.org/10.1111/j.1468-2850.2011.01250.x.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. https://doi.org/10.1080/10705519909540118.

Jackson, D. L., Gillaspy, J. A. Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview

and some recommendations. *Psychological Methods, 14*, 6–23. https://doi.org/10.1037/a0014694.

Koch, A. S., Kleiman, A., Wegener, I., Zur, B., Imbierowicz, K., Geiser, F., & Conrad, R. (2015). Factorial structure of the 20-item Toronto Alexithymia Scale in a large sample of somatoform patients. *Psychiatry Research, 225*, 355–363. https://doi.org/10.1016/j.psychres.2014.12.013.

Kojima, M., Frasure-Smith, N., & Lespérance, F. (2001). Alexithymia following myocardial infarction: Psychometric properties and correlates of the Toronto Alexithymia Scale. *Journal of Psychosomatic Research, 51*, 487–495. https://doi.org/10.1016/S0022-3999(01)00253-7.

Kooiman, C. G., Spinhoven, P., & Trijsburg, R. W. (2002). The assessment of alexithymia: A critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research, 53*, 1083–1090. https://doi.org/10.1016/S0022-3999(02)00348-3.

Kutlesic, V., Williamson, D. A., Gleaves, D. H., Barbin, J. M., & Murphy-Eberenz, K. P. (1998). The Interview for the Diagnosis of Eating Disorders—IV: Application to DSM-IV diagnostic criteria. *Psychological Assessment, 10*, 41–48. https://doi.org/10.1037/1040-3590.10.1.41.

Lambert, M. C., Samms-Vaughan, M. E., Lyubansky, M., Rose, D., Grandison, T., Holness, A., … Durst, J. (1999). Emotions and psychopathology. *The West Indian Medical Journal, 48*, 203–207.

Li, S., Zhang, B., Guo, Y., & Zhang, J. (2015). The association between alexithymia as assessed by the 20-item Toronto Alexithymia Scale and depression: A meta-analysis. *Psychiatry Research, 227*, 1–9. https://doi.org/10.1016/j.psychres.2015.02.006.

Ling, Y., Zeng, Y., Yuan, H., & Zhong, M. (2016). Cross-cultural validation of the 20-item Toronto Alexithymia Scale in Chinese adolescents. *Journal of Psychiatric and Mental Health Nursing, 23*, 179–187. https://doi.org/10.1111/jpm.12298.

Loas, G., Braun, S., Delhaye, M., & Linkowski, P. (2017). The measurement of alexithymia in children and adolescents: Psychometric properties of the Alexithymia Questionnaire for Children and the twenty-item Toronto Alexithymia Scale in different non-clinical and clinical samples of children and adolescents. *PLoS ONE, 12*, 1–17. https://doi.org/10.1371/journal.pone.0177982.

Loas, G., Corcos, M., Stephan, P., Pellet, J., Bizouard, P., Venisse, J. L., … Jeammet, P. (2001). Factorial structure of the 20-item Toronto Alexithymia Scale: Confirmatory factorial analyses in nonclinical and clinical samples. *Journal of Psychosomatic Research, 50*, 255–261. https://doi.org/10.1016/S0022-3999(01)00197-0.

Loas, G., Otmani, O., Verrier, A., Fremaux, D., & Marchand, M. P. (1996). Factor analysis of the French version of the 20-Item Toronto Alexithymia Scale (TAS-20). *Psychopathology, 29*, 139–144.

Lulé, D., Schulze, U. M. E., Bauer, K., Schöll, F., Müller, S., Fladung, A.-K., & Uttner, I. (2014). Anorexia nervosa and its relation to depression, anxiety, alexithymia and emotional processing deficits. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 19*, 209–216. https://doi.org/10.1007/s40519-014-0101-z.

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: Implications for understanding and treating health problems. *Journal of Personality Assessment, 89*, 230–246. https://doi.org/10.1080/00223890701629698.

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

Maggini, C., & Raballo, A. (2004). Alexithymia and schizophrenic psychopathology. *Acta Bio-Medica: Atenei Parmensis, 75*, 40–49.

Mallorquí-Bagué, N., Vintró-Alcaraz, C., Sánchez, I., Riesco, N., Agüera, Z., Granero, R., … Fernández-Aranda, F. (2017). Emotion regulation as a transdiagnostic feature among eating disorders: Cross-sectional and longitudinal approach. *European Eating Disorders Review*. https://doi.org/10.1002/erv.2570.

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*, 366–381. https://doi.org/10.1037/a0019225.

Meganck, R., Vanheule, S., & Desmet, M. (2008). Factorial validity and measurement invariance of the 20-item Toronto Alexithymia Scale in clinical and nonclinical samples. *Assessment, 15*, 36–47. https://doi.org/10.1177/1073191107306140.

Miyake, Y., Okamoto, Y., Onoda, K., Shirao, N., Mantani, T., & Yamawaki, S. (2009). Neural correlates of alexithymia in response to emotional stimuli: A study of anorexia nervosa patients. *Hiroshima Journal of Medical Sciences, 58*, 1–8.

Miyake, Y., Okamoto, Y., Onoda, K., Shirao, N., Okamoto, Y., & Yamawaki, S. (2012). Brain activation during the perception of stressful word stimuli concerning interpersonal relationships in anorexia nervosa patients with high degrees of alexithymia in an fMRI paradigm. *Psychiatry Research: Neuroimaging, 201*, 113–119. https://doi.org/10.1016/j.pscychresns.2011.07.014.

Moriguchi, Y., Maeda, M., Igarashi, T., Ishikawa, T., Shoji, M., Kubo, C., & Komaki, G. (2007). Age and gender effect on alexithymia in large, Japanese community and clinical samples: A cross-validation study of the Toronto Alexithymia Scale (TAS-20). *Biopsychosocial Medicine, 1*, 7. https://doi.org/10.1186/1751-0759-1-7.

Müller, J., Bühner, M., & Ellgring, H. (2003). Is there a reliable factorial structure in the 20-item Toronto Alexithymia Scale?: A comparison of factor models in clinical and normal adult samples. *Journal of Psychosomatic Research, 55*, 561–568. https://doi.org/10.1016/S0022-3999(03)00033-3.

Mustelin, L., Silén, Y., Raevuori, A., Hoek, H. W., Kaprio, J., & Keski-Rahkonen, A. (2016). The DSM-5 diagnostic criteria for anorexia nervosa may change its population prevalence and prognostic value. *Journal of Psychiatric Research, 77*, 85–91. https://doi.org/10.1016/j.jpsychires.2016.03.003.

Myers, T. A., & Crowther, J. H. (2008). Is self-objectification related to interoceptive awareness? An examination of potential mediating pathways to disordered eating attitudes. *Psychology of Women Quarterly, 32*, 172–180. https://doi.org/10.1111/j.1471-6402.2008.00421.x.

Nowakowski, M. E., McFarlane, T., & Cassin, S. (2013). Alexithymia and eating disorders: A critical review of the literature. *Journal of Eating Disorders, 1*, 1–14. https://doi.org/10.1186/2050-2974-1-21.

Oldershaw, A., Lavender, T., Sallis, H., Stahl, D., & Schmidt, U. (2015). Emotion generation and regulation in anorexia nervosa: A systematic review and meta-analysis of self-report data. *Clinical Psychology Review, 39*, 83–95. https://doi.org/10.1016/j.cpr.2015.04.005.

Parker, J. D. A., Eastabrook, J. M., Keefer, K. V., & Wood, L. M. (2010). Can alexithymia be assessed in adolescents? Psychometric properties of the 20-item Toronto Alexithymia Scale in younger, middle, and older adolescents. *Psychological Assessment, 22*, 798–808. https://doi.org/10.1037/a0020256.

Parker, J. D. A., Shaughnessy, P. A., Wood, L. M., Majeski, S. A., & Eastabrook, J. M. (2005). Cross-cultural alexithymia: Validity of the 20-item Toronto Alexithymia Scale in North American aboriginal populations. *Journal of Psychosomatic Research, 58*, 83–88.

Parling, T., Mortazavi, M., & Ghaderi, A. (2010). Alexithymia and emotional awareness in anorexia nervosa: Time for a shift in the measurement of the concept? *Eating Behaviors, 11*, 205–210. https://doi.org/10.1016/j.eatbeh.2010.04.001.

Pascual, A., Etxebarria, I., & Cruz, M. S. (2011). Emotional differences between women with different types of eating disorders. *International Journal of Psychology and Psychological Therapy, 11*, 1–11.

Pinaquy, S., Chabrol, H., & Barbe, P. (2002). Alexithymie et obésité: Étude de la structure factorielle et de la consistance interne de la version française de la Toronto Alexithymia Scale (TAS 20) chez les femmes obèses [Factorial analysis and internal consistency of the French version of the Toronto Alexithymia Scale (TAS 20), in obese women]. *L'Encéphale: Revue de psychiatrie clinique biologique et thérapeutique, 28*, 277–282.

Powell, D. A., & Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta-analysis. *Journal of Educational and Behavioral Statistics, 26*, 105–132.

Prazeres, N., Parker, J., & Taylor, G. (2000). Adaptação Portuguesa da Escala de Alexitimia de Toronto de 20 Itens (TAS-20) [Portuguese adaptation of the 20-item Toronto Alexithymia Scale (TAS-20)]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica (RIDEP), 9*, 9–21.

Richards, H. L., Fortune, D. G., Griffiths, C. E. M., & Main, C. J. (2005). Alexithymia in patients with psoriasis: Clinical correlates and psychometric properties of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research, 58*, 89–96. https://doi.org/10.1016/j.jpsychores.2004.03.009.

Salerno, L., Ingoglia, S., & Lo Coco, G. (2017). Competing factor structures of the Rosenberg Self-Esteem Scale (RSES) and its measurement invariance across clinical and non-clinical samples. *Personality and Individual Differences, 113*, 13–19. https://doi.org/10.1016/j.paid.2017.02.063.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th ed.). New York: Taylor & Francis.

Sexton, M. C., Sunday, S. R., Hurt, S., & Halmi, K. A. (1998). The relationship between alexithymia, depression, and axis II psychopathology in eating disorder inpatients. *International Journal of Eating Disorders, 23*, 277–286.

Simonsson-Sarnecki, M., Lundh, L. G., Törestad, B., Bagby, R. M., Taylor, G. J., & Parker, J. D. (2000). A Swedish translation of the 20-item Toronto Alexithymia Scale: Cross-validation of the factor structure. *Scandinavian Journal of Psychology, 41*, 25–30.

Southgate, L., Tchanturia, K., & Treasure, J. (2005). Building a model of the aetiology of eating disorders by translating experimental neuroscience into clinical practice. *Journal of Mental Health, 14*, 553–566. https://doi.org/10.1080/09638230500347541.

Speranza, M., Loas, G., Guilbaud, O., & Corcos, M. (2011). Are treatment options related to alexithymia in eating disorders? Results from a three-year naturalistic longitudinal study. *Biomedicine and Pharmacotherapy, 65*, 585–589. https://doi.org/10.1016/j.biopha.2010.01.009.

Speranza, M., Loas, G., Wallier, J., & Corcos, M. (2007). Predictive value of alexithymia in patients with eating disorders: A 3-year prospective study. *Journal of Psychosomatic Research, 63*, 365–371. https://doi.org/10.1016/j.jpsychores.2007.03.008.

Swift, L., Stephenson, R., & Royce, J. (2006). The 20-item Toronto Alexithymia Scale: Validation of factor solutions using confirmatory factor analysis on physiotherapy out-patients. *Psychology and Psychotherapy: Theory, Research and Practice, 79*, 83–88. https://doi.org/10.1348/147608305x42875.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*, 134–146.

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge: Cambridge University Press.

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research, 55*, 277–283. https://doi.org/10.1016/S0022-3999(02)00601-3.

Taylor, G. J., Ryan, D., & Bagby, M. R. (1985). Toward the development of a new self-report alexithymia scale. *Psychotherapy and Psychosomatics, 44*, 191–199.

Thorberg, F. A., Young, R. M., Sullivan, K. A., Lyvers, M., Hurst, C., Connor, J. P., & Feeney, G. F. X. (2010). A confirmatory factor analysis of the Toronto Alexithymia Scale (TAS-20) in an alcohol-dependent sample. *Psychiatry Research, 178*, 565–567. https://doi.org/10.1016/j.psychres.2009.09.015.

Torres, S., Guerra, M. P., Lencastre, L., Miller, K., Vieira, F. M., Roma-Torres, A., … Costa, P. (2015). Alexithymia in anorexia nervosa: The mediating role of depression. *Psychiatry Research, 225*, 99–107. https://doi.org/10.1016/j.psychres.2014.10.023.

Torres, S., Guerra, M. P., Lencastre, L., Roma-Torres, A., Brandão, I., Queirós, C., & Vieira, F. (2011). Cognitive processing of emotions in anorexia nervosa. *European Eating Disorders Review, 19*, 100–111. https://doi.org/10.1002/erv.1046.

Torres, S., Guerra, M. P., Lencastre, L., Vieira, F., Roma-Torres, A., & Brandão, I. (2011). Prevalência da alexitimia na anorexia nervosa e sua associação com variáveis clínicas e sociodemográficas [Prevalence of alexithymia in anorexia nervosa and its association with clinical and sociodemographic variables]. *Jornal Brasileiro de Psiquiatria, 60*, 182–189. https://doi.org/10.1590/S0047.

Torres, S., Guerra, M. P., Lencastre, L., Williamson, D. A., Roma-Torres, A., & Vieira, F. (2008). Contribuição para o estudo da adaptação portuguesa da Entrevista de Diagnóstico das Perturbações do Comportamento Alimentar - IV (IDED-IV) específica para a Anorexia Nervosa [Contribution to the Portuguese adaptation of the Interview for the Diagnosis of Eating Disorders-IV (IDED-IV) specifically for Anorexia Nervosa]. *Arquivos de Medicina, 22*, 113–119.

Treasure, J., Corfield, F., & Cardi, V. (2012). A three-phase model of the social emotional functioning in eating disorders. *European Eating Disorders Review, 20*, 431–438. https://doi.org/10.1002/erv.2181.

Troop, N. A., Schmidt, U. H., & Treasure, J. L. (1995). Feelings and fantasy in eating disorders: A factor analysis of the Toronto Alexithymia Scale. *International Journal of Eating Disorders, 18*, 151–157.

Tsaousis, I., Taylor, G., Quilty, L., Georgiades, S., Stavrogiannopoulos, M., & Bagby, R. M. (2010). Validation of a Greek adaptation of the 20-item Toronto Alexithymia Scale. *Comprehensive Psychiatry, 51*, 443–448. https://doi.org/10.1016/j.comppsych.2009.09.005.

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment, 87*, 35–50. https://doi.org/10.1207/s15327752jpa8701_03.

Veríssimo, R. (2001). A versão Portuguesa da Escala de Alexitimia de Toronto de 20 itens - I. Adaptação linguística, validação semântica, e estudo de fiabilidade [The Portuguese version of the 20-item Toronto Alexithymia Scale - I. Linguistic adaptation, semantic validation, and reliability study]. *Acta Medica Portuguesa, 14*, 529–536.

Veríssimo, R. (2003). Inteligência emocional: Da alexitimia ao controlo emocional [Emotional intelligence: From alexithymia to emotional control]. *Acta Medica Portuguesa, 16*, 407–411.

Weinbach, N., Sher, H., & Bohon, C. (2017). Differences in emotion regulation difficulties across types of eating disorders during

adolescence. *Journal of Abnormal Child Psychology*. https://doi.org/10.1007/s10802-017-0365-7.

Westwood, H., Kerr-Gaffney, J., Stahl, D., & Tchanturia, K. (2017). Alexithymia in eating disorders: Systematic review and meta-analyses of studies using the Toronto Alexithymia Scale. *Journal of Psychosomatic Research, 99*, 66–81. https://doi.org/10.1016/j.jpsychores.2017.06.007.

Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research, 16*, 118–154. https://doi.org/10.1177/0049124187016001005.

WHO. (1995). *Physical status: The use and interpretation of anthropometry. Report of a WHO Expert Committee (WHO Technical Report Series 854)*. Geneva: World Health Organization.

Wildes, J. E., Ringham, R. M., & Marcus, M. D. (2010). Emotion avoidance in patients with anorexia nervosa: Initial test of a functional model. *International Journal of Eating Disorders, 43*, 398–404. https://doi.org/10.1002/eat.20730.

Zech, E., Luminet, O., Rimé, B., & Wagner, H. (1999). Alexithymia and its measurement: Confirmatory factor analyses of the 20-item Toronto Alexithymia Scale and the Bermond–Vorst Alexithymia Questionnaire. *European Journal of Personality, 13*, 511–532.

Zhu, X., Yi, J., Yao, S., Ryder, A. G., Taylor, G. J., & Bagby, R. M. (2007). Cross-cultural validation of a Chinese translation of the 20-item Toronto Alexithymia Scale. *Comprehensive Psychiatry, 48*, 489–496. https://doi.org/10.1016/j.comppsych.2007.04.007.