

Assessing Alexithymia: Psychometric Properties and Factorial Invariance of the 20-Item Toronto Alexithymia Scale in Nonclinical and Psychiatric Samples

David Preece¹ · Rodrigo Becerra¹ · Ken Robinson¹ · Justine Dandy¹

Published online: 9 November 2017
© Springer Science+Business Media, LLC 2017

Abstract The 20-item Toronto Alexithymia Scale (TAS-20) is a self-report questionnaire designed to measure the three components of alexithymia; difficulty identifying feelings in the self (DIF), difficulty describing feelings (DDF), and externally orientated thinking (EOT). We examined the scale's psychometric properties in Australian nonclinical ($N = 428$) and psychiatric ($N = 156$) samples. In terms of factorial validity, confirmatory factor analyses found the traditional 3-factor correlated model (DIF, DDF, EOT) to be the best and most parsimonious solution, but it did not reach adequate levels of goodness-of-fit in either sample. Several EOT items loaded poorly on their intended factor, and a reverse-scored item method factor was present; the factor structure of the scale was invariant across both samples. A higher-order factor model (with a single higher-order factor) was slightly inferior to the correlated models, but still tenable. The total scale score and DIF and DDF subscales displayed sound internal consistency, but the EOT subscale did not. We conclude that the TAS-20 has, for the most part, adequate psychometric properties, though interpretation should focus only on the total scale score and DIF and DDF subscales; we recommend the EOT subscale score not be used. Implications for clinical use and future revision of the scale are discussed.

Keywords Alexithymia · Toronto alexithymia scale · TAS-20 · Psychometric · Factor structure · Invariance

✉ David Preece
d.preece@ecu.edu.au

¹ Psychology Department, School of Arts and Humanities, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia

Introduction

Alexithymia (coined from the Greek, *a* = lack, *lexis* = words, *thymos* = feeling) is a trait involving difficulties in the cognitive processing of emotions (Nemiah and Sifneos 1970; Nemiah 1984; Sifneos 1973; Sifneos 1996). Contemporary theorists define it as multidimensional construct, comprised of three interrelated (positively correlated) components: difficulty identifying feelings in the self (DIF); difficulty describing feelings (DDF); and an externally orientated thinking style (EOT) whereby one tends to not focus their attention on their emotions.¹ In other words, people with high levels of alexithymia rarely pay *attention* to their emotional states (EOT) and have difficulty accurately *appraising* what those states are (DIF, DDF) (Preece et al. 2017). Such difficulties are understood to result from underdeveloped emotion schemas (Bucci 1997; Lane and Schwartz 1987; Preece et al. 2017) and the habitual use of experiential avoidance as an emotion regulation strategy (Bilotta et al. 2015; Coriale et al. 2012; Panayiotou et al. 2015; Preece et al. 2017).

The trait is normally distributed in the general population (Mattila et al. 2010; Parker et al. 2008) and is of substantial interest to psychiatry. High levels of alexithymia are an important transdiagnostic risk factor for a range of psychopathologies (Taylor et al. 1999) and have been found to reduce the efficacy of some psychotherapy approaches (Leweke et al.

¹ Some authors disagree with respect to how alexithymia should be defined, positing that difficulty fantasising or reduced emotional reactivity might also be components of the construct (e.g., Vorst and Bermond 2001). Most empirical work has, however, suggested that the latent structure of the construct is comprised of only DIF, DDF and EOT (for a review, see Preece et al. 2017). We, hence, follow this empirically informed three-component definition when defining alexithymia in this paper. These are the three components that the TAS-20 is designed to assess (Bagby et al. 1994), and this is how alexithymia is defined within the *attention-appraisal model of alexithymia* (Preece et al. 2017).

2009). The assessment of alexithymia is, hence, of import. Several psychometric tools have been developed for this purpose (for a review, see Bermond et al. 2015), with the most widely used being a self-report questionnaire, the 20-item Toronto Alexithymia Scale (TAS-20; Bagby et al. 1994).

The TAS-20 has 20 items designed to measure the three components of alexithymia (DIF, 7 items; DDF, 5 items; EOT, 8 items). Each item is comprised of a statement which respondents rate on a 5-point Likert scale. Standard scoring involves the calculation of a subscale score for DIF, DDF and EOT, and the summation of all items into a total scale score as a marker of overall alexithymia. Higher scores indicate higher levels of alexithymia. Five items (four EOT items and one DDF item) are reverse-scored by the examiner because their content describes a low level of alexithymia.

Although there is a growing body of literature examining the psychometric properties of the TAS-20, we consider that five key issues remain unresolved. (1) It is unclear whether the factor structure of the scale is best represented by three or four first-order factors. (2) TAS-20 total and subscale scores are often compared across nonclinical and psychiatric groups, for such comparisons to be meaningful it must be demonstrated that the factor structure of the TAS-20 is invariant across these populations (Cheung and Rensvold 1999), but little research has examined the factorial invariance of the scale. (3) The calculation of a total scale score assumes that the first-order factors (subscales) of the TAS-20 load meaningfully together onto a single higher-order factor (Brown 2014), but few studies have examined whether this is the case. We also have some concerns about (4) the content validity of several EOT items and (5) the often low internal consistency of the EOT subscale. To determine the adequacy of the TAS-20 as a measure of alexithymia these issues require further examination.

The purpose of this study is to provide clarity on these five issues by examining the psychometric properties of the TAS-20 in nonclinical and psychiatric populations. In the remainder of this introduction, we firstly critique the *content validity* of the TAS-20 and then summarise the existing psychometric literature with respect to the *factor structure*, *concurrent/criterion validity*, and *internal consistency reliability* of the scale.

Content Validity

Our view of the content validity of the TAS-20 is that whilst all the DIF and DDF items appear satisfactory, only three of the eight EOT items appear to be satisfactory. We consider all the DIF and DDF items to have satisfactory content validity because they all reference one's ability to recognise, differentiate, or communicate internal feelings. Likewise, we consider three of the eight EOT items to have satisfactory content validity because these three items reference one's tendency to focus attention on their emotions, and this what we (Preece et al. 2017) and others (e.g., Vorst and Bermond 2001) consider to

be the core of EOT. The other five EOT items, however, we consider potentially problematic, because they do not share this same emphasis. Namely, EOT item 16 and item 20 concern one's preference for different entertainment genres or an aversion to analysing entertainment shows too closely, whereas EOT item 5 and item 8 refer to a tendency to analyse everyday events, and item 18 is about one's capacity to form close interpersonal relationships. Thus, although some of these five EOT items mention emotion related phenomena, they seem to move away from the theoretical definition of EOT.

Factor Structure

Several items of the EOT subscale have, indeed, been found to load poorly (factor loading < .40) on their intended latent factor in most factor analytic work (e.g., items 8, 10, 15, 16, and 20 in Bagby et al. 1994; items 5, 8, 16 and 20 in Koch et al. 2015; items 8, 18, 19 and 20 in Meganck et al. 2008). Nonetheless, in line with the theoretical structure of the alexithymia construct, most early studies (e.g., Bach et al. 1996; Bagby et al. 1994; Bressi et al. 1996; Loas et al. 2001; Pandey et al. 1996) found that the scale conformed to a correlated 3-factor structure (comprised of positively correlated DIF, DDF, and EOT factors) when assessed via exploratory or confirmatory factor analyses. The overall goodness-of-fit of this 3-factor correlated model has often been adequate across nonclinical and psychiatric samples, but results have been equivocal, with some later studies finding inadequate levels of fit (Cleland et al. 2005; Haviland and Reise 1996; Koch et al. 2015; Kooiman et al. 2002; Mattila et al. 2010; Thorberg et al. 2010; Watters et al. 2016; Zech et al. 1999) or finding an alternate 2-factor correlated model or 4-factor correlated model to be superior.

The alternate 2-factor correlated model (DIF/DDF, EOT) was endorsed by Kooiman et al. (2002), Erni et al. (1997), and Loas et al. (1996) on the basis of their EFA results showing that the DIF and DDF items loaded on the same factor. The use of EFA is, however, considered less appropriate than CFA when a clear hypothesis about factor structure is present (Fabrigar et al. 1999), and when CFA has been used, the 3-factor correlated model (DIF, DDF, EOT) has always been found to be superior to the 2-factor correlated model (see e.g., Meganck et al. 2008; Zhu et al. 2007); hence there is little psychometric support for the 2-factor correlated model. More convincing evidence is emerging, though, for a 4-factor correlated model (DIF, DDF, PR, IM) where the EOT factor is split into separate pragmatic thinking (PR, 3 items) and lack of importance of emotions (IM, 5 items) factors. The distinction between PR and IM roughly corresponds to a separation between those EOT items which directly reference emotions (IM) and those that do not (PR).

Six CFA studies have directly assessed this 4-factor correlated model (Gignac et al. 2007; Meganck et al. 2008; Müller et al. 2003; Tsaousis et al. 2010; Watters et al. 2016; Zhu et al.

2007).² In nonclinical samples, four out of six studies supported the 4-factor correlated model over the 3-factor correlated model (Gignac et al. 2007; Müller et al. 2003; Watters et al. 2016; Zhu et al. 2007), but in clinical samples the 4-factor correlated model has only been supported in one out of three studies (Müller et al. 2003). Thus the psychiatric status of the sample might influence the factor structure obtained; a distinction between PR and IM may be more common in nonclinical samples. Only one study has, however, formally examined the factorial invariance of the TAS-20 across nonclinical and psychiatric samples (Meganck et al. 2008); in this study, it was found that the 3-factor correlated model was *partially* invariant across Belgian nonclinical and psychiatric samples, with only one item varying.

Of note, in all but two of the abovementioned studies (Gignac et al. 2007; Meganck et al. 2008) the higher-order factor structure of the TAS-20 was not examined. This trend is unfortunate, as researchers using the TAS-20 frequently use only the total scale score (e.g., McGillivray et al. 2017) and the summing of the subscales into a total scale score assumes that all subscales (first-order factors) load meaningfully on a single higher-order factor. This assumption must therefore be confirmed statistically before the total scale score can be used confidently (Brown 2014). Promisingly though, in the two studies to examine a 3-factor higher-order model (where the DIF, DDF and EOT first-order factors were specified to load onto a single higher-order factor) the results offered tentative support for the presence of a higher-order factor (Gignac et al. 2007; Meganck et al. 2008). Further studies are needed, however, to establish the suitability of deriving a total scale score across various populations.

Some authors have also recently begun to examine whether a reverse-scored item method factor might be present in the TAS-20. This has been motivated by findings within the general psychometric literature whereby reverse-scored items in self-report scales are often found to have a problematic influence on factor structure (e.g., van Sonderen et al. 2013). Such an influence is, typically, tested via CFA whereby an additional factor (a method factor) is specified in the model. To date, four CFA studies have used this approach with the TAS-20 (Meganck et al. 2008; Mattila et al. 2010; Watters et al. 2016; Gignac et al. 2007), with a majority finding a prominent method effect. Because most of the TAS-20's reverse-scored items are concentrated in the EOT subscale (four reverse-scored items), some authors have subsequently speculated that this method effect may be the cause of the EOT subscale's low internal consistency (e.g., Meganck et al. 2008). To date, however, no studies have examined the internal consistency of the EOT subscale when the reverse-scored items are excluded.

² In one of these studies, Gignac et al. (2007), the 4-factor correlated model was assessed as part of a bifactor model, with a general alexithymia factor also included in the model loading on all the items.

Concurrent and Criterion Validity

Though there are concerns about some aspects of the TAS-20's internal psychometrics, the scale as a whole does still seem to measure a variable relevant to psychopathology (Taylor and Bagby 2004). The TAS-20 total scale can discriminate between psychiatric and nonclinical populations (e.g., McGillivray et al. 2017), and is strongly associated with self-reported psychological distress (e.g., Leising et al., 2009) and emotion regulation difficulties (e.g., Swart et al., 2009). The TAS-20 also correlates highly with other measures of alexithymia, and displays evidence of convergent and discriminant validity at a subscale level (e.g., Bagby et al. 2006; Preece et al. 2017; Vorst and Bermond 2001). Preece et al. (2017), for example, found in a nonclinical sample that the TAS-20 EOT subscale score correlated highly with other self-report measures of EOT (namely, subscales of the Bermond-Vorst Alexithymia Questionnaire [BVAQ; Vorst and Bermond 2001] and the Difficulties in Emotion Regulation Scale [DERS; Gratz and Roemer 2004]) and, in factor analysis, loaded on the same underlying factor as these other EOT measures. Some of the variance in the TAS-20 EOT subscale score does therefore appear to capture its intended construct.

Internal Consistency Reliability

In almost all studies, however, the internal consistency reliability of the EOT subscale has been below .70 and is frequently below .60 (e.g., Cleland et al. 2005; Kooiman et al. 2002; Loas et al., 2001; Taylor et al. 2003; Thorberg et al. 2010; but see Parker et al. 2003), indicating that more than 50% of the variance in this subscale score is usually attributable to error variance. Most psychometricians agree that the reliability coefficient of a scale score must be at least .70 for it to be useful for research purposes, and ideally around .90 if it is to be used in clinical decision making (e.g., Groth-Marnat, 2009). Hence, whilst the TAS-20 total scale score and DIF and DDF subscales have regularly met these standards, the EOT subscale has not (for a review, see Kooiman et al. 2002). We know of one study that has reported on the reliability of the EOT subscale when it is split into PR and IM subscales, and in this instance, the PR and IM subscales also had poor internal consistency ($\alpha \leq .56$; Müller et al. 2003).

Purpose of the Present Study

To clarify the psychometric strengths and limitations of the TAS-20, the purpose of this study is to comprehensively examine the factor structure, factorial invariance and internal consistency of the TAS-20 across nonclinical and psychiatric samples.

Method

Participants and Procedure

All participants were English speaking and current residents of Australia.

Nonclinical Sample The nonclinical sample was comprised of 428 adults (60.5% female) with an average age of 41.62 ($SD = 16.77$, range = 18–83). The distribution of educational attainment within this sample was, roughly, similar to that of the Australian population as a whole (Australian Bureau of Statistics 2016). For 30.4% their highest level of completed education was high school, for 36% it was a technical diploma, and for 33.4% it was a university degree. Under one quarter of the participants (21.5%) were currently studying at university. The nonclinical sample completed the TAS-20 as part of a battery of psychological questionnaires administered via an anonymous online survey. Participants were recruited via three avenues: an online survey recruiting company (Qualtrics panels), an advertisement placed on a social media website, or an advertisement placed on the unit website of an undergraduate psychology course.³ Some additional participants ($n = 47$, recruited in the same manner) also completed the survey, but their data was excluded during quality screening because they failed at least one of three attention check questions and/or completed the survey impossibly quickly (suggesting inattentive responding). Participants in the nonclinical sample were required to complete all items in order to submit the online survey, hence there were no missing items.

Psychiatric Sample The psychiatric sample was comprised of 156 adults. These patients had been diagnosed (by a psychiatrist) with a psychiatric disorder using the ICD-10 and were attending an outpatient group psychotherapy program at Fremantle hospital in Western Australia. In terms of primary diagnosis, the most common ICD-10 diagnostic category was mood (affective) disorders (F30-F39; 49.4%), then neurotic, stress related and somatoform disorders (F40-F49; 26.9%), disorders of adult personality and behaviour (F60-F69; 14.1%), schizophrenia, schizotypal and delusional disorders (F20-F29; 6.4%), and behavioural syndromes associated with physiological disturbances and physical factors (F50-F59; 0.6%). Diagnostic information was unavailable for 2.6% of the patients. Compared to the nonclinical sample, the proportion of females was slightly higher in the psychiatric sample (71.2%) and the average age was slightly lower ($M = 41.10$, $SD = 12.17$, range = 18–66). The proportion of university

graduates was also lower in this sample (22.4%); for 50% high school was their highest level of completed education, and for 18.6% it was a technical diploma. Prior to the completion of their first group psychotherapy session, patients completed the TAS-20 as part of a battery of psychological questionnaires. Completion of the scale was supervised by a clinical psychologist. Some additional patients ($n = 19$, recruited in the same manner) also completed the TAS-20, but did not complete enough items for their data to be used (data were missing for more than one item in a subscale, or more than two items overall; G. Taylor, personal communication, 28 April, 2016). Of the 156 patients remaining in the sample, 18 had an acceptable level of missing data and missing items were replaced using the expectation maximisation method (Gold and Bentler 2000).

Materials

20-Item Toronto Alexithymia Scale The TAS-20 (Bagby et al. 1994) is a 20-item measure of alexithymia. Items are intended to correspond to three subscales; DIF (7 items, e.g., “I am often confused about what emotion I am feeling”), DDF (5 items, e.g., “It is difficult for me to find the right words for my feelings”), and EOT (8 items, e.g., “Being in touch with emotions is essential” [reverse-scored]). All items are also summed into a total scale score as a marker of overall alexithymia. Each item is comprised of a statement that respondents score on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicate higher levels of alexithymia. Five items (four of which are in the EOT subscale) are reverse-scored items.

Analysis

CFAs were conducted using AMOS 24, all other analyses used SPSS 24. Item scores for the TAS-20 were reasonably normally distributed in both samples (maximum skewness = 1.05, maximum kurtosis = -1.33).

Factor Structure Using a series of CFAs (maximum likelihood estimation based on a Pearson covariance matrix), we analysed the factor structure of the TAS-20 in the nonclinical and psychiatric samples separately.

In the first phase of our CFA testing, we examined three basic first-order models to determine which first-order structure best represented the TAS-20 (see Fig. 1). These first-order factor structures were: a 1-factor model (where all items were specified to load onto a single factor), the traditional 3-factor correlated model (where items were specified to load on either a DIF, DDF, or EOT factor), and the 4-factor correlated model (where the EOT factor was split, and items were specified to load on either a DIF, DDF, PR, or IM factor).

³ Some undergraduate students were included in the nonclinical sample in order to (1) increase the size of the sample and (2) make the distribution of age and gender in the nonclinical sample closer to that of the psychiatric sample.

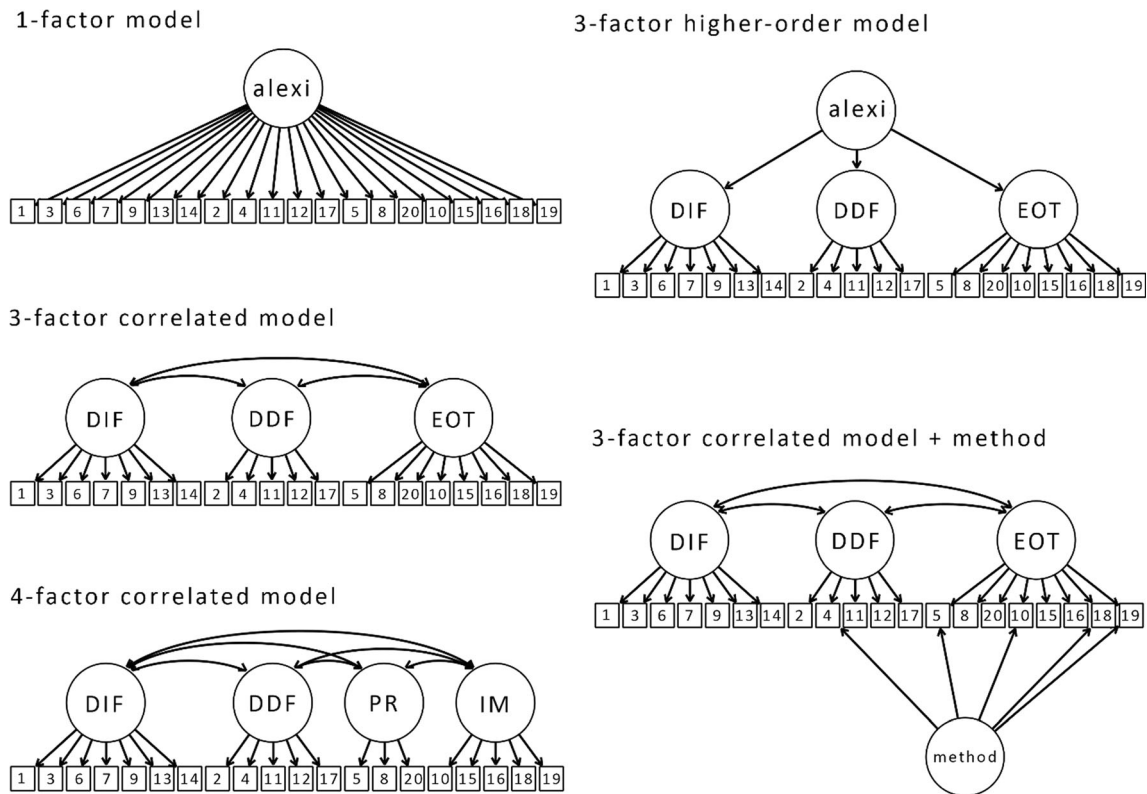


Fig. 1 The assessed confirmatory factor analysis models for the TAS-20. Note. Item error terms are not displayed. Alexi = alexithymia, DIF = difficulty identifying feelings, DDF = difficulty describing

The goodness-of-fit of the CFA models was judged based on the pattern of factor loadings and intercorrelations within each model (Marsh et al. 2004), and three fit indices: the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). These three fit indices were selected as they are considered to be among the best indicators of model fit (Byrne 2016). CFI and TLI values $\geq .90$ were judged to indicate acceptable fit, as were RMSEA values $\leq .08$ (Bentler and Bonett 1980; Browne and Cudeck 1992; Marsh et al. 2004). The models were also directly compared using the Akaike Information Criterion (AIC) and change in CFI. AIC includes a penalty for more complex models, with lower AIC values indicating better levels of fit; a difference between CFI values of $> .01$ also indicates a better fitting model (Byrne 2016; Cheung and Rensvold 2002). Factor loadings $\geq .40$ were considered meaningful loadings (Stevens 1992).

Reverse-Scored Item Method Factor Once the best of these first-order factor models had been determined, we then tested the best model with the addition of a method factor loading on the reverse-scored items. This method factor was specified to be orthogonal to the other first-order factors in the model (see Fig. 1). Models that include the method factor are denoted with the label '+method'.

feelings, EOT = externally orientated thinking, PR = pragmatic thinking, IM = lack of importance of emotions, method = reverse-scored item method factor

Higher-Order Factor We then examined a higher-order version of the best fitting first-order model. In this higher-order model, the first-order factors were specified to load on a single higher-order factor (see Fig. 1).

Factorial Invariance The best fitting model for the nonclinical and psychiatric samples was then examined in terms of whether it was invariant across the samples. Following the procedure outlined by Byrne (2016), a baseline *configural model* was firstly tested with no equality constraints imposed; a *measurement model* was then tested with all factor loadings constrained to be equal across the samples; and a *structural model* was tested with all factor loadings and factor covariances constrained to be equal. A difference in CFI of $< .01$ between the configural model and the measurement and structural models was required for the factor structure to be judged as invariant (Cheung and Rensvold 2002).

Internal Consistency Reliability Cronbach's alpha reliability coefficients were calculated for both samples. We also calculated Cronbach's alpha for a 15-item version of the scale that had all the reverse-scored items removed. Cronbach's alpha $\geq .70$ was used as the criteria for acceptable levels of reliability (Groth-Mamat 2009).

Results

Descriptive Statistics

Descriptive statistics are displayed in Table 1.

Factor Structure

CFAs of the 1-factor model, 3-factor correlated model, and 4-factor correlated model, indicated that the 3-factor correlated model was the best of the first-order factor solutions in both samples (see Table 2).

The 3-factor correlated model was significantly better fitting than the 1-factor model, indicating that the TAS-20 was measuring a multidimensional construct. The 4-factor correlated model offered no advantage over the 3-factor correlated model in either sample, indicating that splitting EOT into IM and PR was unnecessary. The 3-factor correlated model and 4-factor correlated model were roughly equivalent in terms of fit ($\Delta CFI < .01$), but because the 3-factor correlated model was more parsimonious, it was selected as the best. Factor intercorrelations in the 3-factor correlated model were consistent with the theoretical structure of the alexithymia construct, with the DIF, DDF and EOT factors all being significantly positively correlated (see Table 3). However, none of these models reached acceptable levels of fit. According to RMSEA, the 3-factor correlated model demonstrated marginal levels of fit in the nonclinical sample and acceptable levels of fit in the psychiatric sample, but CFI and TLI indicated very poor levels of fit in both samples. Overall fit was, therefore, deemed to be inadequate. Inspection of factor loadings (see Table 4) for the 3-factor correlated model revealed that four items (EOT items 5, 16, 18, 20) had poor loadings in the nonclinical sample, and five items (EOT items 5, 10, 16, 20; DDF item 4) had poor loadings in the psychiatric sample. Most of these items were reverse-scored items.

Reverse-Scored Item Method Factor The addition of the method factor substantially improved the fit of the 3-factor

Table 1 Descriptive statistics and Cronbach’s alpha reliability coefficients for the TAS-20 in the nonclinical and psychiatric samples

	Nonclinical sample			Psychiatric sample		
	<i>M</i>	<i>SD</i>	α	<i>M</i>	<i>SD</i>	α
Total scale	47.08	12.36	.87	59.26	12.42	.86
DIF	14.82	6.05	.87	22.93	6.42	.87
DDF	12.75	4.72	.82	15.73	4.39	.74
EOT	19.51	4.73	.64	20.60	4.75	.61

DIF, difficulty identifying feelings; *DDF*, difficulty describing feelings; *EOT*, externally orientated thinking

correlated model (see Table 2). The 3-factor correlated model + method demonstrated acceptable levels of fit in both samples according to RMSEA, though fit was still unacceptable according to CFI and TLI. Inspection of factor loadings indicated that either four (in the nonclinical sample) or all (in the psychiatric sample) of the reverse-scored items loaded more heavily on the method factor than on their intended alexithymia factor. With the method-factor added, a greater number of items displayed poor loadings on their intended substantive factor (in the nonclinical sample, EOT items 5, 8, 18, 19, and 20; in the psychiatric sample, EOT items 5, 10, 16, 18, 19 and DDF item 4). The size of the positive correlation between the EOT factor and the DIF and DDF factors was also larger in the 3-factor correlated model + method, suggesting that by removing the variance attributable to the method factor, the remaining EOT factor was more closely related to DIF and DDF (see Table 3).

As the overall fit of the 3-factor correlated model + method, nonetheless, remained unacceptable we inspected modification indices for further sources of misspecification. In both samples, there was a large covariance between the error terms of two DIF items, items 3 and 7 (both items refer to being confused by physical sensations in the body). Allowing for this error covariance in the model (models including this covariance are denoted by the label ‘+covariance’) substantially improved the fit of the factor solution. The 3-factor correlated model + method + covariance, however, still did not quite reach *globally* acceptable levels of fit. In the nonclinical sample, RMSEA and CFI indicated acceptable fit whilst TLI indicated marginal fit; in the psychiatric sample RMSEA indicated acceptable fit and CFI and TLI indicated marginal fit (see Table 2). Ultimately, the poor factor loadings of several EOT items seemed to suppress fit index values.

Higher-Order Factor Higher-order versions of the first-order correlated models produced slight decrements in fit in both samples (see Table 2). Nonetheless, in the 3-factor higher-order model, the DIF, DDF and EOT factors all loaded highly and significantly on the higher-order factor in both samples (higher-order factor loadings for the nonclinical sample, DIF = .86, DDF = .85, EOT = .61, $ps < .01$; for the psychiatric sample, DIF = .87, DDF = 1.02, EOT = .56, $ps < .01$), and this higher-order model was not substantially worse fitting than the 3-factor correlated model ($\Delta CFI < .01$). Thus, whilst the relationship between DIF, DDF and EOT could not be perfectly accounted for by a single higher-order factor, there did appear to be enough common variance explained by the higher-order factor to support the calculation of a total scale score.

Factorial Invariance Having established that the 3-factor correlated model + method + covariance was the best fitting model in the nonclinical and psychiatric samples, we then tested the factorial invariance of this model across the two

Table 2 Goodness-of-fit index values for the assessed confirmatory factor analysis models for the nonclinical and psychiatric samples

Model	χ^2 (<i>df</i>)	CFI	TLI	RMSEA (90%CI)	AIC
Nonclinical sample					
First-order models					
1-factor model	1051.313 (170)	.700	.665	.110 (.104–.117)	1131.313
3-factor correlated model	646.768 (167)	.837	.814	.082 (.075–.089)	732.768
4-factor correlated model	641.748 (164)	.837	.812	.083 (.076–.089)	733.748
Models with a method factor					
3-factor correlated model + method	560.157 (162)	.865	.841	.076 (.069–.083)	656.157
3-factor correlated model + method + covariance	456.117 (161)	.900	.881	.066 (.058–.073)	554.117
Higher-order models					
3-factor higher-order model	667.147 (168)	.830	.808	.083 (.077–.090)	751.147
3-factor higher-order model + method	557.780 (163)	.859	.835	.077 (.070–.084)	671.780
3-factor higher-order model + method + covariance	471.329 (162)	.895	.877	.067 (.060–.074)	567.329
Psychiatric sample					
First-order models					
1-factor model	412.591 (170)	.756	.727	.096 (.084–.108)	492.591
3-factor correlated model	337.272 (167)	.829	.805	.081 (.069–.094)	423.272
4-factor correlated model	330.024 (164)	.833	.807	.081 (.068–.093)	422.024
Models with a method factor					
3-factor correlated model + method	286.835 (162)	.874	.853	.071 (.057–.084)	382.835
3-factor correlated model + method + covariance	267.471 (161)	.893	.874	.065 (.051–.079)	365.471
Higher-order models					
3-factor higher-order model	344.529 (168)	.822	.799	.082 (.070–.095)	428.529
3-factor higher-order model + method	290.711 (163)	.872	.850	.071 (.058–.084)	384.711
3-factor higher-order model + method + covariance	271.032 (162)	.890	.871	.066 (.052–.079)	367.032

For all examined models, χ^2 $p < .001$

CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; AIC, Akaike information criterion; CI, confidence interval

samples. Compared to the configural model, the measurement model and structural model were not substantially worse fitting ($\Delta\text{CFI} < .01$); thus, the factor structure of the TAS-20 was invariant across our nonclinical and psychiatric samples (see Table 5).

Internal Consistency Reliability

In both samples, the internal consistency of the TAS-20 total scale score and DIF and DDF subscales was acceptable, however, the EOT subscale had unacceptably low levels of internal consistency (see Table 1). Splitting the EOT subscale into PR and IM subscales did not improve its internal consistency (for the nonclinical sample, PR $\alpha = .30$, IM $\alpha = .57$; for the psychiatric sample, PR $\alpha = .33$, IM $\alpha = .54$) nor did the removal of the reverse-scored items (for the nonclinical sample, EOT [four items] $\alpha = .58$; for the psychiatric sample, EOT [four items] $\alpha = .55$).

Discussion

Our purpose in this study was to examine the psychometric properties of the TAS-20 in nonclinical and psychiatric samples. Our findings suggest that the TAS-20 has, for the most part, adequate psychometric properties and operates similarly

across these populations, though the EOT subscale and the reverse-scored items appear problematic.

The factor structure of the TAS-20 was broadly consistent with the established theoretical structure of the alexithymia construct (Preece et al. 2017). Out of the examined first-order correlated models, the 3-factor correlated model (DIF, DDF, EOT) was the best and most parsimonious solution in both samples. This finding is consistent with the majority of previous literature (e.g., Bagby et al. 1994; Bressi et al. 1996; Loas et al. 2001; Meganck et al. 2008), though it is inconsistent with some recent studies that endorsed the 4-factor correlated model (DIF, DDF, IM, PR), particularly in nonclinical samples (Gignac et al. 2007; Müller et al. 2003; Watters et al. 2016; Zhu et al. 2007). Earlier we speculated, based on these other recent studies, that the 4-factor correlated model might more commonly emerge in nonclinical (as opposed to psychiatric) samples, but our data suggest that psychiatric status is an inadequate explanation for these differences. Language differences are also an inadequate explanation, as some studies using the same (English) version of the scale have previously supported the 4-factor correlated model (Gignac et al. 2007; Watters et al. 2016). Sample demographics could account for some of these differences, but ultimately we consider our results to highlight that it is, practically speaking, somewhat arbitrary to decide between three and four first-order factors, because the EOT items perform poorly (i.e., exhibit low internal consistency) regardless of whether they are grouped into EOT, PR, or IM subscales (see also, Müller et al. 2003). In turn,

Table 3 For the nonclinical and psychiatric samples, estimated factor intercorrelations for the 3-factor correlated model, the 3-factor correlated model + method, and the 4-factor correlated model

Model/Factor	F1	F2	F3	F4
3-factor correlated model				
F1 DIF	–	.876**	.365**	–
F2 DDF	.727**	–	.605**	–
F3 EOT	.441**	.616**	–	–
3-factor correlated model + method				
F1 DIF	–	.880**	.476**	–
F2 DDF	.760**	–	.660**	–
F3 EOT	.466**	.689**	–	–
4-factor correlated model				
F1 DIF	–	.877**	.542**	.246*
F2 DDF	.727**	–	.612**	.528**
F3 PR	.436**	.496**	–	.815**
F4 IM	.429**	.637**	.878**	–

Correlations below the diagonal are for the nonclinical sample, those above the diagonal are for the psychiatric sample

DIF, difficulty identifying feelings; *DDF*, difficulty describing feelings; *EOT*, externally orientated thinking; *PR*, pragmatic thinking; *IM*, lack of importance of emotions

***p* < .01, **p* < .05

whilst the 3-factor correlated model was superior to the one and four factor models in our samples, it failed to reach acceptable levels of goodness-of-fit. These fit issues seemed attributable to two issues; (1) several EOT items had poor factor loadings and (2) a reverse-scored item method factor was present.

Our finding that several EOT items displayed poor factor loadings is not unexpected, as most studies have reported poor factor loadings for multiple EOT items (e.g., Kooiman et al. 2002). Some authors had posited that these issues may be caused by the reverse-scored nature of half the EOT items (e.g., Meganck et al. 2008), and there was indeed a reverse-scored method factor in our data. The addition of this method factor substantially improved the fit of the 3-factor correlated model in both our samples, and at least four of the reverse-scored items loaded more heavily on the method factor than their intended alexithymia factor. The 3-factor correlated model + method + covariance was invariant across the samples, suggesting that our nonclinical and psychiatric populations were similarly affected by the method factor. The method factor could not, however, fully explain the problems of the EOT subscale, because several non-reverse-scored EOT items also had poor loadings and removing the reverse-scored items did not improve the low internal consistency of this subscale. Of note, when the reverse-scored items were removed, three of the four remaining EOT items were items that we considered to have unsatisfactory content validity. These results are, therefore, in line with our critique of the content validity of the EOT subscale.

That said, consistent with the previous findings of Gignac et al. (2007) and Meganck et al. (2008), the EOT factor, alongside the DIF and DDF factors, did still load meaningfully on a single higher-order factor in our samples. This suggests that some of the variance in the EOT factor score does capture a construct relevant to alexithymia. Ultimately though, similar to previous psychometric investigations (e.g., Kooiman et al. 2002; Meganck et al. 2008; Müller et al. 2003), we found that information garnered about EOT from the TAS-20 is not sufficiently robust. Whilst the TAS-20 total scale score and DIF and DDF subscales had acceptable levels of internal consistency reliability, the internal consistency of the EOT subscale was unacceptably low across both samples. These reliability issues, as aforementioned, were not improved by removing the reverse-scored items, nor were they improved by splitting the EOT subscale into PR and IM.

Implications

It is a concern that across our samples, regardless of the psychiatric status of the respondent, none of the theoretically informed factor structures for the TAS-20 displayed globally good levels of fit. Even when the method effect is accounted for, the content validity problems of several EOT items appear to suppress fit index values. We, nonetheless, think that the traditional TAS-20 total scale score and DIF and DDF subscales can be used with reasonable confidence; the caveat being that the EOT subscale score is not robust enough to be used in clinical or research settings. Given the poor factor loadings of several EOT items, some scholars might question the wisdom of still including these items within the TAS-20 total scale score. We agree that this is not ideal, however we also think there is enough evidence to support that the TAS-20 total scale score, in its traditional form, still assesses a variable relevant to psychopathology. Namely, as aforementioned, across the literature this total scale score regularly displays good levels of internal consistency (e.g., Taylor et al. 2003), correlates highly with other measures of alexithymia (e.g., Vorst and Bermond 2001), and discriminates between non-clinical and psychiatric samples (e.g., McGillivray et al. 2017). Additionally, in our samples, whilst the relationship between the DIF, DDF and EOT factors was not perfectly accounted for by the higher-order factor, in our view, the size of the EOT factor’s loading on the higher-order factor was still sufficient to support the calculation of a total scale score as a rough estimate of overall alexithymia (Brown 2014).

The TAS-20, therefore, seems a viable option for examiners wanting to measure overall levels of alexithymia via self-report, particularly since alternative measures also have some weaknesses. Whilst the BVAQ (Vorst and Bermond 2001) and DERS (Gratz and Roemer 2004), for example, have an advantage over the TAS-20 in that they have EOT subscales with acceptable reliability, the 40-item BVAQ is arguably

Table 4 Standardised factor loadings for the 3-factor correlated model, 4-factor correlated model, and 3-factor correlated model + method

Factor/ Item	Nonclinical sample			Psychiatric sample		
	3-factor correlated model	4-factor correlated model	3-factor correlated model + method	3-factor correlated model	4-factor correlated model	3-factor correlated model + method
DIF						
1	.766	.766	.741	.770	.770	.769
3	.527	.526	.585	.638	.636	.637
6	.755	.755	.729	.710	.710	.710
7	.647	.647	.714	.606	.608	.605
9	.776	.776	.813	.808	.810	.809
13	.737	.737	.716	.734	.733	.734
14	.673	.672	.654	.639	.636	.640
DDF						
2	.793	.794	.795	.785	.795	.788
4(r)	.687	.690	.638 (.288)	.332	.337	.313 (.340)
11	.738	.736	.751	.719	.712	.722
12	.572	.571	.580	.522	.513	.529
17	.663	.663	.711	.656	.654	.646
EOT or PR						
5(r)	.231	.232	-.006* (.520)	.295	.280	.158* (.356)
8	.490	.565	.374	.493	.551	.492
20	.333	.372	.294	.293	.322	.412
EOT or IM						
10(r)	.637	.631	.400 (.466)	.288	.355	.104* (.423)
15	.439	.451	.429	.580	.528	.677
16	.345	.348	.485	.257	.225	.354
18(r)	.360	.355	.227 (.293)	.559	.618	.373 (.534)
19(r)	.592	.583	.315 (.562)	.476	.518	.259 (.619)

Loadings < .40 are in boldface. Factor loadings in brackets indicate loadings on the method factor

DIF, difficulty identifying feelings; *DDF*, difficulty describing feelings; *EOT*, externally orientated thinking; *PR*, pragmatic thinking; *IM*, lack of importance of emotions; *r*, reverse-scored item

* $p > .05$

unnecessarily long and includes 16 items which do not assess DIF, DDF or EOT, and the 36-item DERS is an incomplete measure of alexithymia in that it has no DDF items.

For clinical or research questions requiring the *isolation* of EOT though (e.g., Bankier et al. 2001; Leweke et al. 2011; Son et al. 2012; Subic-Wrana et al. 2005), the TAS-20 is not appropriate and would need to be administered as part of a larger battery of tests. Our favoured approach for addressing such questions in clinical practice, presently, is to administer the TAS-20 and DERS. The TAS-20 functions as the primary indicator of overall alexithymia and DIF and DDF, whilst the

DERS compliments it by providing reliable information about EOT (via the awareness subscale) as well as information about emotion regulation (Gratz and Roemer 2004; Gross 2015). This seems an adequate solution in the short term, but in the long term, such assessments will ideally be streamlined by revising the TAS-20 so as to improve the content validity and reliability of the EOT subscale. These revisions should be done, in our view, by rewriting the EOT items so that they more directly reference one's tendency to focus attention on emotions (Preece et al. 2017). To remove the problematic influence of the method-factor, we further recommend that

Table 5 Factorial invariance of the 3-factor correlated model + method + covariance across the nonclinical and psychiatric samples

Model	χ^2 (df)	CFI	TLI	RMSEA (90% CI)	AIC
Configural model	723.588 (322)	.898	.879	.046 (.042–.051)	919.588
Measurement model	774.690 (343)	.890	.878	.047 (.042–.051)	928.690
Structural model	775.925 (346)	.891	.880	.046 (.042–.051)	923.925

For all examined models, $\chi^2 p < .001$. In the configural model, no equality constraints were imposed. In the measurement model, all factor loadings were constrained to be equal across the samples. In the structural model, all factor loadings and factor covariances were constrained to be equal across the samples

CFI, comparative fit index; *TLI*, Tucker-Lewis index; *RMSEA*, root mean square error of approximation; *AIC*, Akaike information criterion; *CI*, confidence interval

all TAS-20 items be written so as to not require reverse-scoring (see also, van Sonderen et al. 2013).

Limitations

We think our study makes a strong contribution, but some limitations should be noted. Chiefly, the size of our psychiatric sample was modest, though it was still large enough for the factor analysis to be robust according to widely accepted criteria (Kline 1979) and was similar to that of other studies (e.g., Meganck et al. 2008). Our results also apply only to adults; further research is needed to examine the psychometrics of the TAS-20 in adolescent populations (e.g., Parker et al. 2010).

Conclusions

Regardless of an adult examinee's psychiatric status, it appears the TAS-20 can be used in its current form as an adequate measure of overall alexithymia and the DIF and DDF components of the construct. The EOT subscale score, however, is not reliable enough to be used in isolation as a marker of EOT. Researchers and clinicians should be aware of this limitation when interpreting scores from the TAS-20. Future work should focus on revising the EOT items so as to improve the utility of the scale.

Acknowledgements We would like to thank Fremantle Hospital in Western Australia for the use of their site for data collection.

Compliance with Ethical Standards

Conflict of Interest David Preece, Rodrigo Becerra, Ken Robinson and Justine Dandy declares that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Experiment Participants This study was approved by ethics committees from the South Metropolitan Health Service and Edith Cowan University. Informed consent was obtained from all individual participants.

References

- Australian Bureau of Statistics. (2016). *Education and work, Australia* (no. 6227.0). Retrieved from <http://www.abs.gov.au/ausstats/abs@.nsf/mf/6227.0>.
- Bach, M., Bach, D., De Zwaan, M., Serim, M., & Böhmer, F. (1996). Validation of the German version of the 20-item Toronto alexithymia scale in normal persons and psychiatric patients. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *46*, 23–28.
- Bagby, R. M., Parker, J. D., & 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.
- Bagby, R. M., Taylor, G. J., Parker, J. D., & Dickens, S. E. (2006). The development of the Toronto structured interview for alexithymia: Item selection, factor structure, reliability and concurrent validity. *Psychotherapy and Psychosomatics*, *75*, 25–39.
- Bankier, B., Aigner, M., & Bach, M. (2001). Alexithymia in DSM-IV disorder: Comparative evaluation of somatoform disorder, panic disorder, obsessive-compulsive disorder, and depression. *Psychosomatics*, *42*, 235–240.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.
- Bermond, B., Oosterveld, P., & Vorst, H. C. M. (2015). Measures of alexithymia. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 227–256). San Diego: Academic.
- Bilotta, E., Giacomantonio, M., Leone, L., Mancini, F., & Coriale, G. (2015). Being alexithymic: Necessity or convenience. Negative emotionality × avoidant coping interactions and alexithymia. *Psychology and Psychotherapy: Theory, Research and Practice*, *89*, 261–275.
- Bressi, C., Taylor, G., Parker, J., Bressi, S., Brambilla, V., Aguglia, E., et al. (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.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258.
- Bucci, W. (1997). Symptoms and symbols: A multiple code theory of somatization. *Psychoanalytic Inquiry*, *17*, 151–172.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*, 1–27.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
- 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.
- Coriale, G., Bilotta, E., Leone, L., Cosimi, F., Porrari, R., De Rosa, F., & Ceccanti, M. (2012). Avoidance coping strategies, alexithymia and alcohol abuse: A mediation analysis. *Addictive Behaviors*, *37*, 1224–1229.
- Erni, T., Lötscher, K., & Modestin, J. (1997). Two-factor solution of the 20-item Toronto alexithymia scale confirmed. *Psychopathology*, *30*, 335–340.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.
- Gignac, G. E., Palmer, B. R., & Stough, C. (2007). A confirmatory factor analytic investigation of the TAS-20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment*, *89*, 247–257.
- Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, *7*, 319–355.

- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, 26, 41–54.
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26, 1–26.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. Hoboken: Wiley.
- Haviland, M. G., & Reise, S. P. (1996). Structure of the twenty-item Toronto alexithymia scale. *Journal of Personality Assessment*, 66, 116–125.
- Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.
- 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.
- 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.
- Lane, R. D., & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *The American Journal of Psychiatry*, 144, 133–143.
- Leising, D., Grande, T., & Faber, R. (2009). The Toronto Alexithymia Scale (TAS-20): A measure of general psychological distress. *Journal of Research in Personality*, 43, 707–710.
- Leweke, F., Bausch, S., Leichsenring, F., Walter, B., & Stingl, M. (2009). Alexithymia as a predictor of outcome of psychodynamically oriented inpatient treatment. *Psychotherapy Research*, 19, 323–331.
- Leweke, F., Leichsenring, F., Kruse, J., & Hermes, S. (2011). Is alexithymia associated with specific mental disorders. *Psychopathology*, 45, 22–28.
- 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.
- Loas, G., Corcos, M., Stephan, P., Pellet, J., Bizouard, P., Venisse, J. L., et al. (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.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320–341.
- Mattila, A. K., Keefer, K. V., Taylor, G. J., Joukamaa, M., Jula, A., Parker, J. D., & Bagby, R. M. (2010). Taxometric analysis of alexithymia in a general population sample from Finland. *Personality and Individual Differences*, 49, 216–221.
- McGillivray, L., Becerra, R., & Harms, C. (2017). Prevalence and demographic correlates of alexithymia: A comparison between Australian psychiatric and community samples. *Journal of Clinical Psychology*, 73, 76–87.
- 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.
- 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.
- Nemiah, J. C. (1984). The psychodynamic view of anxiety. In R. O. Pasnau (Ed.), *Diagnosis and treatment of anxiety disorders* (pp. 117–137). Washington, DC: American Psychiatric Press.
- Nemiah, J. C., & Sifneos, P. E. (1970). Psychosomatic illness: A problem in communication. *Psychotherapy and Psychosomatics*, 18, 154–160.
- Panayiotou, G., Leonidou, C., Constantinou, E., Hart, J., Rinehart, K. L., Sy, J. T., & Björgvinsson, T. (2015). Do alexithymic individuals avoid their feelings? Experiential avoidance mediates the association between alexithymia, psychosomatic, and depressive symptoms in a community and a clinical sample. *Comprehensive Psychiatry*, 56, 206–216.
- Pandey, R., Mandal, M. K., Taylor, G. J., & Parker, J. D. (1996). Cross-cultural alexithymia: Development and validation of a Hindi translation of the 20-item Toronto alexithymia scale. *Journal of Clinical Psychology*, 52, 173–176.
- Parker, J. D., Taylor, G. J., & Bagby, R. M. (2003). The 20-item Toronto alexithymia scale: III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research*, 55, 269–275.
- Parker, J. D., Keefer, K. V., Taylor, G. J., & Bagby, R. M. (2008). Latent structure of the alexithymia construct: A taxometric investigation. *Psychological Assessment*, 20, 385–396.
- Parker, J. D., Eastbrook, 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.
- Preece, D., Becerra, R., Allan, A., Robinson, K., & Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences*, 119, 341–352.
- Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, 22, 255–262.
- Sifneos, P. E. (1996). Alexithymia: Past and present. *The American Journal of Psychiatry*, 153, 137–142.
- Son, S., Jo, H., Rim, H. D., Kim, J. H., Kim, H. W., Bae, G. Y., & Lee, S. J. (2012). A comparative study on alexithymia in depressive, somatoform, anxiety, and psychotic disorders among Koreans. *Psychiatry Investigation*, 9, 325–331.
- van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS One*, 8, e68967.
- Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale: Erlbaum.
- Subic-Wrana, C., Bruder, S., Thomas, W., Lane, R. D., & Köhle, K. (2005). Emotional awareness deficits in inpatients of a psychosomatic ward: A comparison of two different measures of alexithymia. *Psychosomatic Medicine*, 67, 483–489.
- Swart, M., Kortekaas, R., & Aleman, A. (2009). Dealing with feelings: Characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PLoS One*, 4, e5751.
- Taylor, G. J., & Bagby, R. M. (2004). New trends in alexithymia research. *Psychotherapy and Psychosomatics*, 73, 68–77.
- Taylor, G. J., Bagby, R. M., & Parker, J. D. (1999). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge: Cambridge University Press.
- Taylor, G. J., Bagby, R. M., & Parker, J. D. (2003). The 20-item Toronto alexithymia scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, 55, 277–283.
- Thorberg, F. A., Young, R. M., Sullivan, K. A., Lyvers, M., Hurst, C., Connor, J. P., & Feeney, G. F. (2010). A confirmatory factor analysis of the Toronto alexithymia scale (TAS-20) in an alcohol-dependent sample. *Psychiatry Research*, 178, 565–567.
- 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.
- Vorst, H. C., & Bermond, B. (2001). Validity and reliability of the Bermond–Vorst alexithymia questionnaire. *Personality and Individual Differences*, 30, 413–434.

- Watters, C., Taylor, G. J., Ayearst, L., & Bagby, R. M. (2016). Measurement invariance of the English and French language versions of the 20-item Toronto alexithymia scale. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000365>.
- 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.