JHP

Full-length report

Journal of Health Psychology
1–15

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13591053261449068
journals.sagepub.com/home/hpq
Sage

Reliability and validity of the Dutch interoceptive accuracy scale and interoceptive attention scale

Jesper Mulder[1,2], Marije T. Elferink-Gemser[2], Juriena D. de Vries[3] and Jessica C. Kiefte-de Jong[1]

Abstract

Interoception, the perception of internal bodily signals, plays a key role in health behavior. This study evaluated the reliability and validity of the Dutch Interoceptive Accuracy Scale (IAS-D) and Interoceptive Attention Scale (IATS-D) in 779 Dutch participants (mean age 52.98 ± 16.47, 49.6% female). Both questionnaires showed good to excellent internal consistency (IAS-D $\alpha = 0.895$, $\omega = 0.989$; IATS-D $\alpha = 0.940$, $\omega = 0.980$). The IAS-D was negatively associated with the Interoceptive Confusion Questionnaire (ICQ; $\beta = -0.638$), depression ($\beta = -0.298$), and alexithymia ($\beta = 0.528$), all $p < 0.001$. The IATS-D was positively associated with the ICQ ($\beta = 0.481$), Body Perception Questionnaire (BPQ; $\beta = 0.207$), depression ($\beta = 0.377$), and alexithymia ($\beta = 0.528$), all $p < 0.001$. The IAS-D and IATS-D are reliable and valid instruments for assessing self-reported interoceptive accuracy and attention in Dutch-speaking samples, supporting interoceptive and health behavior research.

Keywords

interoception, self-report, alexithymia, depression, health behavior

Introduction

Interoception is defined as the perception of the internal state of the body, including a wide range of physical states like heart rate, respiration, temperature, fatigue, hunger, satiety, muscle ache, and pain (Craig, 2002, 2003, 2009; Murphy et al., 2019). Research in the field of interoception has increased over the years, which has provided a better understanding of the important role of interoception in physical and mental well-being. Interoception helps in maintaining homeostasis (i.e. the process of maintaining a stable internal environment) and allostasis (i.e. the process of actively adjusting and adapting to changing conditions; Quadt et al., 2018). Changes in internal bodily states can also affect behaviors (Critchley and Harrison, 2013; Farb et al., 2015; Quadt et al.,

[1]Leiden University Medical Center, The Hague, The Netherlands
[2]University Medical Center Groningen, University of Groningen, The Netherlands
[3]Leiden University, The Netherlands

Corresponding author:
Jesper Mulder, Health Campus The Hague, Department of Public Health and Primary Care, Leiden University Medical Center, Spui 5, The Hague BL 2511, The Netherlands.
Email: j.mulder6@lumc.nl

Data Availability Statement included at the end of the article

2018). Several mental disorders have been associated with impaired interoception, such as depressive disorders and alexithymia (i.e. an inability to recognize or describe one's own emotions). Lower scores on interoceptive measures have been related to more depressive symptoms (Pollatos et al., 2009; Quadt et al., 2018), and higher alexithymia scores (Brewer et al., 2016; Murphy et al., 2020). In order to combat these interoceptive deficits, it has been shown that clinical populations may potentially benefit from interoception-based interventions (Heim et al., 2023), allowing patients to more effectively use internal signals to decrease symptoms (Quadt et al., 2018). However, interoceptive dimensions (e.g. accuracy or attention) and bodily domains (e.g. heartrate or respiration) are associated with these clinical symptoms in different ways. For instance, in anxiety respiratory interoceptive accuracy is decreased, while thermoceptive interoceptive accuracy is increased (Schoeller et al., 2025). This shows the complexity of interoception and the need for specific measures for different interoceptive dimensions and bodily domains.

Definitions and measures of interoceptive dimensions and bodily domains vary greatly across studies (Forkmann et al., 2016). Additionally, research has shown that there is low convergence between different bodily domains and measurement types (e.g. behavioral tasks vs questionnaires), possibly leading to heterogeneous conclusions (Desmedt et al., 2025). Therefore, it is necessary to clearly define these dimensions and measurement types to accurately describe interoception and decreasing the risk of overgeneralization, while also acknowledging its multimodal and multifaceted nature. A promising model in this regard is the 2×2 factorial structure for interoception (Figure 1) developed by Murphy and colleagues (Murphy et al., 2019). This model suggests two factors. The first factor makes the distinction between two dimensions: interoceptive accuracy (i.e. an individual's ability to perceive interoceptive signals accurately) and attention (i.e. the degree to which interoceptive signals are objects of attention). The second factor distinguishes between the way these dimensions (i.e. interoceptive accuracy or attention) is measured: an objective performance measure or a self-report measure. Measuring different facets of interoception may provide more accurate assessments (Schoeller et al., 2025). In the current study we will focus on self-reported interoception. Validated self-report measures of interoceptive dimensions are essential for large-scale intervention and population studies, where objective interoceptive tasks are often impractical. Based on the 2×2 factorial model of interoception, two self-report measures of interoception have been developed: the Interoceptive Accuracy Scale (IAS; Murphy et al., 2020) and the Interoceptive Attention Scale (IATS; Gabriele et al., 2022). These two questionnaires were developed as unidimensional measures in general populations, while existing questionnaires such as the Interoceptive Confusion Questionnaire (ICQ) and Body Perception Questionnaire (BPQ) are often multidimensional and more tailored to clinical dysfunction or broad autonomic nervous system awareness. Since the accuracy and attention dimensions of interoception seem to be distinct but complementary (Brand et al., 2023; Gabriele et al., 2022; Murphy et al., 2019, 2020; Tünte et al., 2024), and questionnaires focusing on these dimensions are not yet available in Dutch, there is interest to investigate the psychometric properties of Dutch versions of the IAS and IATS.

The objective of the current study is to investigate the psychometric properties of Dutch translations of the IAS and IATS (IAS-D and IATS-D) in a sample of the general Dutch population. We will investigate the internal consistency, factor structures, and convergent, discriminant, and criterion validity of the questionnaires. Consequently, we compare the IAS-D and IATS-D to validated questionnaires: the ICQ, BPQ Short Form (BPQ-SF), and questionnaires on depression and alexithymia. Our hypotheses are based on the findings of previous studies (Brand et al., 2023; Gabriele et al., 2022; Murphy et al., 2019, 2020; Tünte et al., 2024): we expect to find a one-factor model for both the IAS and IATS, as both questionnaires were originally created as single-factor scales.

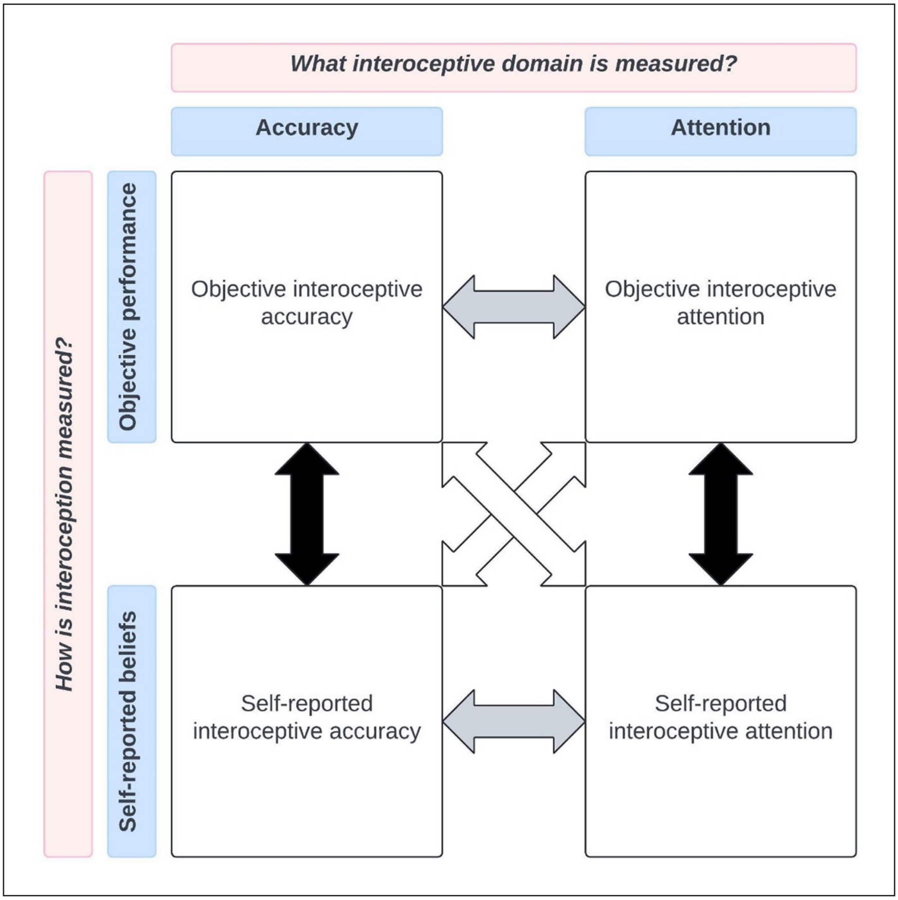

Figure 1. Simplified version of the 2 × 2 factorial structure for interoception developed by Murphy et al. (2019) The quadrants represent interoceptive accuracy assessed with objective performance tasks (top left), interoceptive attention measured with objective performance tasks (top right), interoceptive accuracy assessed with self-report measures (bottom left), and interoceptive attention assessed with self-report measures (bottom right). The arrows represent the interaction between objective and self-report measures (i.e. interoceptive awareness; black arrows), interoceptive accuracy and attention (gray arrows), and across different interoceptive dimensions and measures (white arrows).

The IAS is expected to have a significant negative relation with the ICQ (i.e. when accuracy increases, confusion decreases), and the IATS is expected to have a significant positive relation with both the ICQ and BPQ-SF (i.e. paying more attention to internal signals may increase difficulties in perceiving signals and increase interoceptive confusion, and individuals may interpret awareness, as assessed by the BPQ-SF, as attention; Gabriele et al., 2022). We expect no significant relation between the IAS and BPQ-SF. To investigate criterion validity the correlation between the IAS, IATS, depression, and alexithymia will be examined. We hypothesize that interoceptive accuracy relates negatively, and interoceptive attention positively, to depressive symptoms (Brand et al., 2023; Tünte et al., 2024). This is expected because emotional blunting (i.e. a reduction in range or intensity of emotional experiences) and autonomic dysregulation reduce accuracy in perceiving bodily signals, while simultaneously heightened inward focus and rumination increase interoceptive attention (Quadt et al., 2018). It is also expected that interoceptive accuracy negatively relates to alexithymia, and interoceptive attention positively to alexithymia (Murphy et al., 2020; Tünte et al., 2024), as those with alexithymia seem to show amplified bodily sensation (i.e. increased attention; Longarzo et al., 2015) while not being able to accurately relate these sensations to interoceptive states (Brewer et al., 2016).

Methods

Participants

Participants in this study were recruited through a Dutch internet panel, administered by research agency Flycatcher (www.flycatcher.eu). This panel consists of 20,000 members from the Dutch public, and is designed to be broadly representative with respect to age, sex, region, and socioeconomic status. Although the panel is designed to be broadly representative, participation was limited to individuals with internet access and willingness to complete online surveys. To be eligible to participate in the study, a subject had to meet the following criteria: the subject was (1) at least 18 years old, and (2) had a clear understanding of the Dutch language. The participants in this study were contacted via email to take part in this study. Participants were given points that could be spend on (charity) gift vouchers for completing the surveys. Participants voluntarily and actively indicated their willingness to participate in the online surveys, through Flycatcher's "double-active-opt-in." Flycatcher meets high quality requirements and is ISO-certified. Sample size calculation for multiple linear regression analyses using G*Power 3.1 (Faul et al., 2009) indicated a minimum sample size of 812 participants, based on an effect size of 0.02, alpha of 0.005 (0.05 divided by 10 predictors in the original, overarching study protocol), and a power of 0.8. The surveys remained open until the minimum sample size was reached. A total of 815 participants completed the surveys. Data was collected between July 2024 and August 2024.

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The local research committee (reference number WSC-2024-26/ SP) and local ethics committee (METC number N24.072) approved this study as non-WMO. Informed consent was obtained from all individual participants included in the study.

Measures

Self-reported interoceptive accuracy. The Interoceptive Accuracy Scale (IAS; Murphy et al., 2020) is a questionnaire constructed to assess individuals' accuracy in evaluating sensations that have either been described as interoceptive or are associated with activation in the insula, an area commonly associated with the processing of interoceptive signals. An example of a question in this questionnaire is "I can always accurately perceive when my heart is beating fast." The scale comprises of 21 items rated on a scale from strongly disagree (1) to strongly agree (5), with total scores ranging from 21 to 105. Higher scores indicate greater self-reported interoceptive accuracy. Internal consistency of the IAS in the current sample ranged from good to excellent ($\alpha = 0.895$; $\omega = 0.989$).

The Interoceptive Confusion Questionnaire (ICQ; Brewer et al., 2016) assesses the degree to which individuals feel that they struggle to interpret their own non-affective interoceptive states, such as hunger, temperature, and arousal. Examples of questions in this questionnaire are "I am very sensitive to changes in my heartrate" and "When I adjust the heat of a room or car, others find it uncomfortable." Items are scored on a scale from 1 ("Does not describe me") to 5 ("Describes me very well"). Total scores range between 20 and 100. Higher scores indicate greater self-reported interoceptive confusion. Internal consistency of the ICQ in the current sample ranged from sufficient to excellent ($\alpha = 0.694$; $\omega = 0.954$).

Self-reported interoceptive attention. The Interoceptive Attention Scale (IATS; Gabriele et al., 2022) is a questionnaire constructed to quantify the extent to which internal signals are the object of one's attention. The signals exactly match the signals included in the IAS. An example of a question in this questionnaire is "Most of the time my attention is focused on whether my heart is beating fast." The IATS comprises 21 items rated from strongly disagree (1) to strongly agree (5), with total scores ranging from 21 to 105. Higher scores indicate greater self-reported attention to internal signals. Internal consistency of the IATS in the current sample were excellent ($\alpha = 0.940$; $\omega = 0.980$).

The Body Perception Questionnaire Short Form (BPQ-SF; Cabrera et al., 2018) was developed to assess the subjective experiences of the function and reactivity of target organs and structures that are innervated by the autonomic nervous system. The original BPQ has 122 items and assesses body awareness, autonomic nervous system reactivity, cognitive-emotional-somatic stress response, body and cognitive stress response styles, and health history. The short form focuses on the awareness and autonomic reactivity subscales of the BPQ, resulting in 46 items scored from never (1) to always (5). Total scores range between 46 and 230. Example items of the BPQ-SF are "During most situations I am aware of swallowing frequently" and "During most situations I am aware of how hard my heart is beating." Higher scores indicate greater self-reported interoceptive attention. Internal consistency of the BPQ-SF in the current sample were excellent ($\alpha = 0.952$; $\omega = 0.981$). In German samples the BPQ-SF was found to be significantly related to the IAS (Brand et al., 2023; Tünte et al., 2024), but in English samples this was not the case (Gabriele et al., 2022; Murphy et al., 2020). The BPQ-SF was found to be significantly related to the IATS (Gabriele et al., 2022; Tünte et al., 2024).

Depression. The Beck Depression Inventory (BDI-II; Beck et al., 1996) measures the severity of depressive symptoms. The questionnaire consists of 21 groups of statements assessing the presence of psychological and physiological symptoms of major depression. Statements are assigned point values (ranging from 0 to 3) reflecting the severity of depressive symptoms. Total scores range from 0 to 63, with a higher score indicating more depressive symptoms. Example answers from an item are: "I do not feel sad," "I feel sad," "I am sad all the time and I cannot snap out of it," or "I am so sad and unhappy that I cannot stand it." The internal consistency of the Dutch BDI-II has been reported as Cronbach's $\alpha = 0.95$ (Roelofs et al., 2013). Internal consistency of the BDI-II in the current sample ranged from good to excellent ($\alpha = 0.871$; $\omega = 0.989$).

Alexithymia. The Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994) is a 20-item scale to assess alexithymia traits. The scale includes items like "I am often confused about what emotion I am feeling" and "I have physical sensations that even doctors do not understand." The scale is rated on a 5-point forced-choice answer format ranging from strongly disagree (1) to strongly agree (5), and grouped in three subscales: difficulty identifying feelings, difficulty describing feelings, and externally oriented thinking. Higher scores indicate stronger alexithymia traits. Total scores range from 20 to 100. Internal consistency of the TAS-20 in the current sample ranged from acceptable to good ($\alpha = 0.712$; $\omega = 0.885$).

Procedures

The IAS and IATS questionnaires were translated to Dutch, using cross-cultural guidelines (Sousa and Rojjanasrirat, 2011), resulting in the IAS-D and IATS-D. The translation was performed by an expert from an organization specializing in translating medical questionnaires. After translation, the Dutch questionnaires were evaluated in individual think-aloud interviews with a small group of Dutch speaking individuals. Interviewees were asked to read out loud the questions and answer options from the questionnaires, and verbally express how they would answer the questions, any difficulties in interpretation, and possible alternatives wordings whenever such issues arose. Where necessary, adjustments were made to the Dutch versions of the questionnaires based on feedback from the think-aloud interviews (Supplemental Appendix A). No back-translation of the questionnaires was performed. Adjustments were discussed with the other researchers until consensus was reached, resulting in a final version of both questionnaires (Supplemental Appendix B and C). The ICQ was not available in Dutch and was translated using the same method as described above, however this questionnaire was not validated in the current study (Supplemental Appendix D).

Statistical analysis and software

All analyses were performed in Python version 3.12, using the factor-analyzer (version 0.5), scikit-learn (version 1.5), scipy (version 1.14), semopy (version 2.3), and statsmodels (version 0.14) packages. Normality of data for all measures was checked beforehand, by visually examining the distribution of data using histograms and Q-Q plots, and performing the Shapiro-Wilk test. Variables with missing data showed signs of data missing at random (MAR > 0.5), and were imputed using the Multiple Imputation by Chained Equations with random forests (MICEforest version 6.0) package in Python, generating 10 imputed datasets and running 20 iterations. Outlier detection was performed on total scores of all questionnaires, and outliers outside three standard deviations were removed from the dataset. Descriptive statistics (mean, standard deviation, median, interquartile range, minimum, and maximum) were calculated to summarize the characteristics and questionnaire scores of the participants.

Cronbach's alpha and McDonald's omega were calculated, with an alpha and omega of 0.7 or higher considered acceptable (Gliem and Gliem, 2003). For factor structure and reliability analyses, the dataset was split into two independent samples using stratified random sampling. This resulted in separate samples for the exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), stratified on gender to maintain comparable gender distributions across the two samples. The EFA and CFA were performed separately for the IAS and IATS. For the EFA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were calculated. For the KMO, values above 0.5 are recommended (Williams et al., 2010). Parallel analyses were conducted to determine the optimal number of factors. Kaiser's criterion (eigenvalue > 1) was calculated for comparison to the parallel analyses. A principal axis factoring extraction method and oblimin rotation were used. To confirm the structure of the Dutch IAS and IATS as identified in the EFA, a CFA with weighted least squares model estimation was performed. The fit indices comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were calculated. For the CFI and TLI values above 0.90 are usually considered to be indicators of acceptable fit (Hooper et al., 2008). For the RMSEA and SRMR values below 0.08 indicate acceptable fit (Hooper et al., 2008; Hu and Bentler, 1999). Measurement invariance testing across gender was conducted using a hierarchical approach (Vandenberg and Lance, 2000). Three models were evaluated: (1) configural invariance (same factor structure, no constraints), (2) metric invariance (factor loadings constrained equal), and (3) scalar invariance (loadings and intercepts constrained equal). Model comparison was based on changes in fit indices, with $\Delta CFI \leq 0.010$ and $\Delta RMSEA \leq 0.015$ indicating acceptable invariance (Cheung and Rensvold, 2002).

Convergent, discriminant, and criterion validity were determined using linear regression analyses. The dependent variables were the total scores of the IAS and IATS. The independent variables were the total scores of the ICQ, BPQ-SF, TAS-20, and BDI-II. For convergent validity the IAS was compared to the ICQ, and the IATS to the BPQ-SF. For discriminant validity the IAS was compared to the BPQ-SF, and the IATS to the ICQ. To investigate criterion validity the directions of the relations

Table 1. Mean age and average total scores per questionnaire from 779 Dutch adults.

Variable	Mean	SD	Median	Q1	Q3	Min	Max
Age	52.98	16.47	54	41	67	19	92
Interoceptive accuracy							
IAS-D	80.01	9.39	80	75	84	50	105
ICQ	47.36	7.18	47	42	53	27	66
Interoceptive attention							
IATS-D	50.58	14.72	49	41	62	21	86
BPQ-SF	101.02	25.02	100	82	120	46	172
Depression and alexithymia							
BDI-II	7.42	6.37	6	3	11	0	31
TAS-20	55.08	7.83	55	50	60	33	79

IAS-D: Interoceptive Accuracy Scale Dutch version; ICQ: Interoceptive Confusion Questionnaire; IATS-D: Interoceptive Attention Scale Dutch version; BPQ-SF: Body Perception Questionnaire Short Form; BDI-II: Beck Depression Inventory; TAS-20: Toronto Alexithymia Scale.

between the IAS, IATS, TAS-20, and BDI-II were investigated. In all analyses, alpha was set at 0.05 a priori.

Results

In total, 815 participants responded to the questionnaires. After removal of outliers, 779 participants were included in the analyses. The remaining participants consisted of 393 (50.4%) males and 386 (49.6%) females, with an average age of 52.98 (SD=16.47) years. Table 1 shows the average score per questionnaire for the complete sample.

Interoceptive accuracy scale factor structure

Cronbach's alpha and McDonald's omega indicated good to excellent internal consistency for the Dutch IAS ($\alpha=0.895$; $\omega=0.989$). The KMO measure of sampling adequacy was 0.883, which was above the recommend value of 0.5 (Williams et al., 2010). Bartlett's test of sphericity was significant ($\chi^2=2986.81$, $p<0.001$), indicating significant correlations among variables.

An EFA with principal axis factoring and oblimin rotation was used to examine the factor structure of the IAS-D. Parallel analyses and Kaiser's criterion both indicated a four-factor structure, which explained 46.9% of the total variance. A CFA was performed to evaluate the fit of the four-factor model (Table 2). The CFA showed poor fit: $\chi^2(183)=648.55$, CFI=0.511, TLI=0.438, RMSEA=0.081, SRMR=0.306 (Table 4). Despite the poor fit, measurement invariance testing across gender was conducted to evaluate whether the measurement structure differs between males and females. Both males (CFI=0.779, RMSEA=0.093) and females (CFI=0.697, RMSEA=0.107) displayed poor fit. Despite this poor fit, results indicated that the relationship between individual items and what they measure was equivalent across genders (ΔCFI=−0.006, ΔRMSEA=0.002). However, men and women appear to use the response scale differently (ΔCFI=−0.019, ΔRMSEA=0.005).

Interoceptive attention scale factor structure

Cronbach's alpha and McDonald's omega indicated excellent internal consistency for the Dutch IATS ($\alpha=0.940$; $\omega=0.980$). The KMO measure of sampling adequacy was 0.946, which was above the recommend value of 0.5 (Williams et al., 2010). Bartlett's test of sphericity was significant ($\chi^2=4969.17$, $p<0.001$), indicating significant relations among variables.

Table 2. Confirmatory Factor Analysis (CFA) standardized factor loadings for the IAS-D four-factor model (CFA sample $n = 390$).

Item	Factor	β	SE	*z*	*p*
Item 8: Vomiting	1	1.000			
Item 11: Temperature	1	0.935	0.158	5.91	<0.001
Item 15: Muscle soreness	1	0.769	0.137	5.60	<0.001
Item 17: Pain	1	0.466	0.107	4.38	<0.001
Item 19: Affective touch	1	−0.645	0.171	−3.78	<0.001
Item 20: Tickling	1	0.115	0.095	1.21	0.226
Item 21: Itching	1	0.071	0.064	1.10	0.270
Item 1: Heart	2	1.000			
Item 16: Bruising	2	−0.094	0.062	−1.53	0.127
Item 18: Blood sugar	2	0.479	0.184	2.60	0.009
Item 7: Tasting	3	1.000			
Item 9: Sneezing	3	−5.113	1.756	−2.91	0.004
Item 10: Coughing	3	−1.644	0.498	−3.30	0.001
Item 13: Wind	3	−0.778	0.302	−2.58	0.010
Item 14: Burping	3	−0.617	0.230	−2.68	0.007
Item 2: Hunger	4	1.000			
Item 3: Breathing	4	0.303	0.109	2.79	0.005
Item 4: Thirst	4	0.377	0.097	3.87	<0.001
Item 5: Urination	4	1.238	0.194	6.39	<0.001
Item 6: Defecation	4	1.231	0.190	6.50	<0.001
Item 12: Sexual arousal	4	0.147	0.109	1.35	0.178

An EFA with principal axis factoring and oblimin rotation was performed to examine the factor structure of the IATS-D. Parallel analyses indicated a single-factor structure, which explained 48.1% explained variance. A CFA testing the single-factor model (Table 3) indicated poor fit to the data: $\chi^2(189) = 942.00$, CFI = 0.589, TLI = 0.543, RMSEA = 0.101, SRMR = 0.432 (Table 4). Measurement invariance testing across gender showed poor fit for both males (CFI = 0.759, RMSEA = 0.129) and females (CFI = 0.803, RMSEA = 0.109). IATS-D items relate to interoceptive attention in the same way for both genders (ΔCFI = −0.006, ΔRMSEA = 0.001), but men and women use the rating scale differently (ΔCFI = −0.019, ΔRMSEA = 0.005).

Regression analyses

Univariate linear regression analyses were performed with the IAS-D as dependent variable, and the ICQ, BPQ-SF, BDI-II, and TAS-20 as independent variables. Our results indicated significant negative associations between the IAS-D scores and the ICQ (β = −0.638, 95% CI [−0.719, −0.558]), BDI-II (β = −0.298, 95% CI [−0.399, −0.196]), and TAS-20 (β = −0.298, 95% CI [−0.380, −0.216]). Lower levels of interoceptive confusion, depressive symptoms, and alexithymia traits were associated with greater interoceptive accuracy. We observed no significant association between the IAS-D and the BPQ-SF (Table 5).

The same analyses were performed with the IATS-D as dependent variable. The ICQ (β = 0.481, 95% CI [0.341, 0.621]), BPQ-SF (β = 0.207, 95% CI [0.168, 0.246]), BDI-II (β = 0.377, 95% CI [0.216, 0.537]), and TAS-20 (β = 0.528, 95% CI [0.400, 0.655]) were significant determinants of IATS-D scores (Table 5). Higher levels of interoceptive confusion, autonomic nervous system reactivity, depressive symptoms, and alexithymia traits were associated with greater interoceptive attention.

Table 3. Confirmatory Factor Analysis (CFA) standardized factor loadings for the IATS-D single-factor model (CFA sample $n = 390$).

Item	β	SE	*z*	*p*
Item 1: Heart	1.000			
Item 2: Hunger	0.735	0.045	16.20	<0.001
Item 3: Breathing	1.052	0.038	27.76	<0.001
Item 4: Thirst	0.867	0.039	22.17	<0.001
Item 5: Urination	0.658	0.041	16.11	<0.001
Item 6: Defecation	0.979	0.036	27.59	<0.001
Item 7: Tasting	0.132	0.043	3.08	0.002
Item 8: Vomiting	1.053	0.042	25.29	<0.001
Item 9: Sneezing	1.122	0.040	28.37	<0.001
Item 10: Coughing	1.123	0.038	29.63	<0.001
Item 11: Temperature	0.895	0.045	19.70	<0.001
Item 12: Sexual arousal	0.856	0.034	24.88	<0.001
Item 13: Wind	0.921	0.031	29.44	<0.001
Item 14: Burping	1.129	0.041	27.60	<0.001
Item 15: Muscle soreness	0.842	0.038	21.91	<0.001
Item 16: Bruising	0.858	0.047	18.11	<0.001
Item 17: Pain	0.872	0.039	22.43	<0.001
Item 18: Blood sugar	0.993	0.040	24.68	<0.001
Item 19: Affective touch	0.336	0.043	7.81	<0.001
Item 20: Tickling	0.850	0.039	21.81	<0.001
Item 21: Itching	0.987	0.033	29.47	<0.001

Table 4. Confirmatory Factor Analysis (CFA) goodness of fit indices ($n = 390$) and measurement invariance testing (configural, metric, and scalar) across gender (male $n = 197$, female $n = 193$) for the IAS-D and IATS-D models.

Confirmatory Factor Analyses goodness of fit								
Scale	Model	χ^2	df	*p*	CFI	TLI	RMSEA	SRMR
IAS-D	Four-factor	648.55	183	<0.001	0.511	0.438	0.081	0.306
IATS-D	Single-factor	942.00	189	<0.001	0.589	0.543	0.101	0.432
Measurement invariance testing models								
Scale	Model	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA	Result	
IAS-D	Configural	0.692	0.665	0.112			Baseline	
	Metric	0.686	0.660	0.113	−0.006	0.001	Supported	
	Scalar	0.667	0.641	0.118	−0.019	0.005	Not supported	
IATS-D	Configural	0.781	0.756	0.119			Baseline	
	Metric	0.775	0.751	0.120	−0.006	0.001	Supported	
	Scalar	0.756	0.732	0.125	−0.019	0.005	Not supported	

Table 5. Univariate regression analyses (performed on the complete dataset, $n=779$) with the Interoceptive Accuracy Scale and Interoceptive Attention Scale as dependent variables, and the ICQ, BPQ-SF, BDI-II, and TAS-20 as independent variables.

Variables	β	SE	*t*	*p*	95% CI	R^2	Adj. R^2	*F*	AIC	BIC
Dependent variable: IAS-D										
ICQ	−0.638	0.041	−15.568	<0.001	−0.719, −0.558	0.238	0.237	242.35	5492.02	5501.34
BPQ-SF	0.0002	0.013	0.017	0.986	−0.026, 0.027	0.000	−0.001	0.00	5703.50	5712.82
BDI-II	−0.298	0.052	−5.749	<0.001	−0.399, −0.196	0.041	0.040	33.05	5671.06	5680.37
TAS-20	−0.298	0.042	−7.144	<0.001	−0.380, −0.216	0.062	0.060	51.04	5653.95	5663.26
Dependent variable: IATS-D										
ICQ	0.481	0.072	6.725	<0.001	0.341, 0.621	0.055	0.054	45.23	6359.18	6368.50
BPQ-SF	0.207	0.020	10.471	<0.001	0.168, 0.246	0.124	0.123	109.65	6300.43	6309.74
BDI-II	0.377	0.082	4.606	<0.001	0.216, 0.537	0.027	0.025	21.21	6382.28	6391.59
TAS-20	0.527	0.065	8.147	<0.001	0.400, 0.655	0.079	0.078	66.38	6339.40	6348.72

IAS-D: Interoceptive Accuracy Scale Dutch version; ICQ: Interoceptive Confusion Questionnaire; IATS-D: Interoceptive Attention Scale Dutch version; BPQ-SF: Body Perception Questionnaire Short Form; BDI-II: Beck Depression Inventory; TAS-20: Toronto Alexithymia Scale.

Discussion

This study aimed to develop Dutch versions of the IAS (i.e. interoceptive accuracy; Murphy et al., 2020) and IATS (i.e. interoceptive attention; Gabriele et al., 2022), and to examine their internal consistency, factor structure, and convergent, discriminant, and criterion validity. The IAS-D showed good to excellent internal consistency. A four-factor structure was identified in the EFA, but showed poor fit in the CFA. The IATS-D showed excellent internal consistency. For the IATS-D, a single-factor structure was found in the EFA, with poor CFA fit indices. The regression analyses confirmed our hypotheses for both the IAS-D and IATS-D.

Four underlying factors were identified for the IAS-D in the EFA. However, CFA fit indices indicated poor fit for this four-factor structure, possibly indicating this factor structure is not stable across our split samples or may be sample-specific. This finding may be in line with the rationale behind the construction of the original IAS, which was developed with a single-factor structure in mind (Murphy et al., 2020). Additionally, interpretation of the underlying themes provides mixed results: the first two factors of the IAS-D show no clear coherent themes, while factors three and four may respectively capture concepts like respiration and basic physiological needs. Furthermore, the moderate correlations between factors suggest these dimensions may not be truly distinct, instead referring to a unidimensional structure. The finding that the four-factor structure found in the EFA could not be replicated in the CFA, the lack of coherent themes, and good to excellent reliability may suggest that, like the original questionnaire, the Dutch version of the IAS should be used as a unidimensional scale. Moreover, the use of split-sampling prevented overfitting and showed the four-factor structure was not replicable, further supporting the one-dimensionality of the IAS-D.

Unlike the IAS-D, the EFA for the IATS-D indicated a single-factor structure. However, the current study found poor model fit for this single-factor structure in the CFA. These findings contrast those of a previous study, where good CFA model fit indices were found (Tünte et al., 2024), as well as the original validation (Gabriele et al., 2022). The discrepancy between our findings and prior validations suggests the poor fit may be related to the current sample or the Dutch adaptation of the scale, rather than fundamental problems with the dimensional structure of the scale. Even though the CFA model fit indices showed poor fit, the clear unidimensional structure found in the EFA, along with excellent internal consistency, suggests that the use of IATS-D total scores maintains comparability to previous literature, and the intended development of the original scale.

Both the IAS-D and IATS-D showed good internal consistency, comparable to those found in other samples (Brand et al., 2023; Murphy et al., 2020; Tünte et al., 2024). However, in the factor structures some discrepancies were found. In the current study a four-factor structure was identified for the IAS-D, which was different from previous findings where single-factor structures were proposed (Brand et al., 2023; Murphy et al., 2020). Conversely, the single-factor structure found in the IATS-D was in line with the development of the original scale and subsequent validation studies, even though CFA fit indices were poor in the current sample (Gabriele et al., 2022; Tünte et al., 2024). One explanation for these differing findings in factor structures and model fit might be cultural differences. Even though people from the Netherlands, Germany, and the United Kingdom share similarities based on their Western European heritage, there are also several cultural differences that may influence study results. There might be differences in response styles (Harzing, 2006; van de Vijver and Leung, 2021), interpretation and translation of the questionnaires (Behr, 2017; Brislin, 1970), and cultural dimensions that impact underlying structures (Hofstede, 2001; Schwartz, 1999). For example, in terms of item interpretation, during the back-translation process, interviewees expressed difficulty in interpreting item 7 (tasting) as either being able to distinguish between different tastes (e.g. sweet, sour, bitter, etc.) or the ability to recognize when one food item tastes different from another food item. These subtle differences in interpretation of questions could have resulted in discrepancies in findings between the current and previous studies. Additionally, sample characteristics may differ from prior validations. Our sample included substantial age variability ($M = 53.0$, SD = 16.5, range 19–92) and 39% reported chronic health conditions. Older adults or those with chronic conditions may demonstrate different patterns of interoception than younger, healthier validation samples. Summarizing, while CFA fit indices were suboptimal for both scales, current evidence supports their validity: (1) both showed excellent reliability comparable to prior research, (2) the IATS-D demonstrated clear structural consistency with theory and prior validation, (3) unidimensional use of the IAS-D aligns with original development and is indicated by excellent reliability, and (4) current findings are interpretable within the context of potential cultural and sample differences. In addition to these fit indices, measurement invariance testing identified that for both questionnaires men and women seem to interpret the questions in fundamentally the same way, but appear to use the response scales in different ways. This means the questionnaires can be used to study relationships between interoceptive accuracy and attention, and other factors in both men and women, but direct comparisons of average scores between genders should be interpreted with caution.

For the convergent, discriminant, and criterion validity, all our hypotheses were confirmed. Interoceptive accuracy showed significant negative associations with other measures of interoceptive accuracy, depressive symptoms, and alexithymia traits. This means that individuals who think they are accurate in perceiving their internal signals, are less likely to display depressive symptoms and alexithymia traits, and vice versa. A measure of interoceptive attention (i.e. BPQ-SF) did not determine interoceptive accuracy scores, indicating that self-reported accuracy of perceiving internal signals is not the same as paying attention to them. These findings mirror results from English and German versions of the IAS (Brand et al., 2023; Gabriele et al., 2022; Murphy et al., 2020; Tünte et al., 2024). Interoceptive attention was significantly and positively associated with other measures of interoceptive attention, interoceptive accuracy, depression, and alexithymia. The implication of these findings is that those who pay more attention to their bodily signals display more depressive symptoms and alexithymia traits, which is suggestive of maladaptive (i.e. insufficient or excessive attention to bodily signals, unfavorable for health-promoting behavior) forms of interoception (Trevisan et al., 2023). Again, these findings are in line with previous studies on the IATS in different cultural samples (Gabriele et al., 2022; Tünte et al., 2024). These differences between the IAS-D and the IATS-D signify that interoceptive accuracy and attention are distinctive interoceptive dimensions that interact with each other, and with depressive symptoms and alexithymia, in varying ways (Schoeller et al., 2025). Additionally, this shows that these questionnaires can distinguish between different interoceptive dimensions, and are valid instruments for measuring perceived interoception.

One of the strengths of this validation study was the large sample size and the use of split sample methods, enhancing statistical power and robustness of the validation. The sample of 779 participants, with a diverse background in terms of age, region of the Netherlands where they reside, and socio-economic status, contributing to the high external validity. Additionally, this is the first time the IAS and IATS were translated to Dutch and validated in a sample from the Netherlands. This allows for the new insights into the potential underlying structures and use of the IAS and IATS, as well as adding valuable knowledge on cultural differences in interoceptive processes. This study also has its limitations. First, this study had a cross-sectional design, and the questionnaires were not validated on test-retest reliability. This absence of a longitudinal aspect limits the conclusions that can be drawn about the stability of the IAS-D and IATS-D over longer periods of time. However, test-retest reliability is challenging because of learning effects: individuals may become more aware of their internal bodily signals every time they fill out a questionnaire, causing them to change their answers over time. Additionally, the self-report measures of interoception used in this study were not compared to objective interoceptive tasks, thus we are not able explain results in relation to these kinds of tests. It is suggested that researchers should consider the complex relationship between objective and subjective interoceptive measures, in order to provide more accurate assessments (Schoeller et al., 2025). Second, we made use of a translated but not validated version of the ICQ. While this may have influenced the current findings, to be able to directly compare to previous studies using the same questionnaires the decision was made to use a back-translated version of the ICQ. The ICQ was translated using the same procedure as with the IAS and IATS, ensuring quality of translations and increasing confidence in the use of these measures. Finally, data collection in this study was performed using online surveys. Online testing may be limited because of selection bias (Bethlehem, 2010; Sue and Ritter, 2007) or participants rushing through surveys by providing random answers (Huang et al., 2012). Future studies might consider adding attention control items (e.g. “Please select option #3 when you are reading this”) to improve the quality of responses and prevent participants rushing through the survey.

The model fit statistics of the CFA showed an imperfect fit for both the IAS-D and the IATS-D. Discrepancies in CFA model fit between the current and previous studies indicate the need for future research to further investigate the underlying structures, the use of the questionnaires, and how the questionnaire items can be improved. Additionally, since test-retest reliability was not included in the current study, and no comparisons were made with objective interoceptive tests, the questionnaires could benefit from more longitudinal research in several interoceptive dimensions. Nonetheless, the current study has shown that the IAS-D and IATS-D are valid measures of interoceptive accuracy and attention. The questionnaires can distinguish between these dimensions of interoception. This is important, since it has been suggested that different interoceptive dimensions and bodily domains associate in varying ways with clinical symptoms (Schoeller et al., 2025). In addition to traditional psychometric and clinical applications, validated self-report measures of interoception may play an important role in future data-driven approaches to mental health research. Our findings supported the theoretical distinction between adaptive versus maladaptive interoceptive patterns. Dimension-specific measures that are able to identify these interoceptive profiles can support structured, theory-grounded inputs for data-driven approaches that identify interoceptive risk profiles in clinician-centered ways, supporting early detection and personalized interventions while maintaining interpretability (Triberti et al., 2022). Finally, while some items may need refinement to improve interpretation, the total scores of the questionnaires are easy to use, and are therefore relevant for use in practical settings.

In conclusion, the current study found that the IAS-D and IATS-D are reliable and valid instruments for the assessment of self-report interoceptive accuracy and attention in Dutch speaking populations. The results show that the questionnaires can distinguish between interoceptive dimensions, and are related to depression and alexithymia as expected from previous research. Future longitudinal research with these questionnaires in Dutch samples should focus on the association with objective interoception and test-retest reliability. These valid and reliable Dutch questionnaires offer accessible tools for assessing interoceptive accuracy and attention, supporting public health research in understanding the role of interoception in health behavior.

ORCID iDs

Jesper Mulder https://orcid.org/0009-0003-1501-7331
Marije T. Elferink-Gemser https://orcid.org/0000-0003-2555-4782
Juriena D. de Vries https://orcid.org/0000-0002-9385-2290
Jessica C. Kiefte-de Jong https://orcid.org/0000-0002-8136-0918

Ethical considerations

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The local research committee (reference number WSC-2024-26/ SP) and local ethics committee (METC number N24.072) approved this study as non-WMO.

Consent to participate

Consent to participate was obtained from all individual participants included in the study.

Consent to publication

Consent to publication was obtained from all individual participants included in the study.

Author contributions

JM, MEG, and JKJ conceptualized and designed the study. JM performed data collection. JM and JKJ conducted statistical analyses. JM, MEG, JdV, and JKJ contributed to the interpretation of the results. JM drafted the initial manuscript, with MEG, JdV, and JKJ providing critical revisions and refining the final manuscript. All authors reviewed and approved of the final version of the manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is part of a research project that is funded by the Velux Stiftung (project number 1815). The sponsor had no influence on the content of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

Data will be made available by the corresponding author upon reasonable request. Analysis code is available at: https://github.com/JesperMulder/dutch-interoceptive-scales-validation.

Supplemental material

Supplemental material for this article is available online.

References

Bagby RM, Parker JD and Taylor GJ (1994) The twenty-item Toronto Alexithymia scale–I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research* 38(1): 23–32.

Beck AT, Steer RA and Brown GK (1996) *BDI-II: Beck Depression Inventory Manual*. Psychological Corporation.

Behr D (2017) Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology* 20(6): 573–584.

Bethlehem J (2010) Selection bias in web surveys. *International Statistical Review* 78(2): 161–188.

Brand S, Meis AC, Tünte MR, et al. (2023) A multisite German validation of the interoceptive accuracy scale and its relation to psychopathological symptom burden. *Communications psychology* 1(1): 14.

Brewer R, Cook R and Bird G (2016) Alexithymia: A general deficit of interoception. *Royal Society Open Science* 3(10): 150664.

Brislin RW (1970) Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology* 1(3): 185–216.

Cabrera A, Kolacz J, Pailhez G, et al. (2018) Assessing body awareness and autonomic reactivity: Factor structure and psychometric properties of the Body Perception Questionnaire-Short Form (BPQ-SF). *International Journal of Methods in Psychiatric Research* 27(2): e1596.

Cheung GW and Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9(2): 233–255.

Craig AD (2002) How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience* 3(8): 655–666.

Craig AD (2003) Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology* 13(4): 500–505.

Craig AD (2009) How do you feel — Now? The anterior insula and human awareness. *Nature Reviews Neuroscience* 10(1): 59–70.

Critchley HD and Harrison NA (2013) Visceral influences on brain and behavior. *Neuron* 77(4): 624–638.

Desmedt O, Luminet O, Maurage P, et al. (2025) Discrepancies in the definition and measurement of human interoception: A comprehensive discussion and suggested ways forward. *Perspectives on Psychological Science* 20(1): 76–98.

Farb N, Daubenmier J, Price CJ, et al. (2015) Interoception, contemplative practice, and health. *Frontiers in Psychology* 6: 763.

Faul F, Erdfelder E, Buchner A, et al. (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41(4): 1149–1160.

Forkmann T, Scherer A, Meessen J, et al. (2016) Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology* 109: 71–80.

Gabriele E, Spooner R, Brewer R, et al. (2022) Dissociations between self-reported interoceptive accuracy and attention: Evidence from the interoceptive attention scale. *Biological Psychology* 168: 108243.

Gliem J and Gliem R (2003) Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales. In: *2003 midwest research to practice conference in adult, continuing, and community education*.

Harzing AW (2006) Response styles in cross-national survey research. *International Journal of Cross Cultural Management* 6(2): 243–266.

Heim N, Bobou M, Tanzer M, et al. (2023) Psychological interventions for interoception in mental health disorders: A systematic review of randomized-controlled trials. *Psychiatry and Clinical Neurosciences* 77(10): 530–540.

Hofstede G (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Orgaizations Across Nations*. Sage Publications.

Hooper D, Coughlan J and Mullen M (2008) Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods* 6(1): 53–60.

Huang JL, Curran PG, Keeney J, et al. (2012) Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology* 27(1): 99–114.

Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6(1): 1–55.

Longarzo M, D'Olimpio F, Chiavazzo A, et al. (2015) The relationships between interoception and alexithymic trait. The self-Awareness Questionnaire in healthy subjects. *Frontiers in Psychology* 6: 1149.

Murphy J, Brewer R, Plans D, et al. (2020) Testing the independence of self-reported interoceptive accuracy and attention. *Quarterly Journal of Experimental Psychology* 73(1): 115–133.

Murphy J, Catmur C and Bird G (2019) Classifying individual differences in interoception: Implications for the measurement of interoceptive awareness. *Psychonomic Bulletin & Review* 26(5): 1467–1471.

Pollatos O, Traut-Mattausch E and Schandry R (2009) Differential effects of anxiety and depression on interoceptive accuracy. *Depression and Anxiety* 26(2): 167–173.

Quadt L, Critchley HD and Garfinkel SN (2018) The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences* 1428(1): 112–128.

Roelofs J, van Breukelen G, de Graaf LE, et al. (2013) Norms for the Beck Depression Inventory (BDI-II) in a large Dutch community sample. *Journal of Psychopathology and Behavioral Assessment* 35(1): 93–98.

Schoeller FA, Zhang B, Garcia T, et al. (2025) There is no such thing as interoception. *Frontiers in Psychology* 16: 1488415.

Schwartz SH (1999) A theory of cultural values and some implications for work. *Applied Psychology* 48(1): 23–47.

Sousa VD and Rojjanasrirat W (2011) Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice* 17(2): 268–274.

Sue VM and Ritter LA (2007) *Conducting Online Surveys*. Sage.

Trevisan DA, Tsheringla S and McPartland JC (2023) On the relation between interoceptive attention and health anxiety: Distinguishing adaptive and maladaptive bodily awareness. *Cogent Psychology* 10(1): 2262855.

Triberti S, Durosini I, La Torre D, et al. (2022) Artificial intelligence in healthcare practice: How to tackle the “human” challenge. In: Lim CP, Chen YW, Vaidya A, et al. (eds) *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*. Springer International Publishing, pp.43–60.

Tünte MR, Petzke TM, Brand S, et al. (2024) He who seeks finds (Bodily signals): German validation of the Interoceptive Attention Scale (IATS) and its relationship with subclinical psychopathology. *Journal of Personality Assessment* 106(6): 787–797.

Vandenberg RJ and Lance CE (2000) A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for Organizational Research. *Organizational Research Methods* 3(1): 4–70.

van de Vijver FJR and Leung K (2021) *Methods and Data Analysis for Cross-Cultural Research*. Cambridge University Press.

Williams B, Onsman A and Brown T (2010) Exploratory factor analysis: A five-step guide for Novices. *Australasian Journal of Paramedicine* 8: 1–13.