

**Untangling Self-Reported Interoceptive Attention and Accuracy: Evidence from the
European Portuguese Validation of the Body Perception Questionnaire and the Interoceptive
Accuracy Scale**

Carlos Campos^{1,2,3}, Nuno Barbosa Rocha³ and Fernando Barbosa¹

¹Laboratory of Neuropsychophysiology, Faculty of Psychology and Education Sciences,
University of Porto, Portugal

²Faculty of Medicine, University of Porto, Porto, Portugal

³Neurocognition Group | LabRP, Center for Rehabilitation Research, School of Health,
Polytechnic Institute of Porto, Porto, Portugal

Author Notes

This project was supported by a Ph. D. grant funded by the Portuguese Foundation for Science and Technology and the POCH/FSE Program (SFRH/BD/136296/2018). We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Carlos Campos, Laboratory of Neuropsychophysiology, Faculty of Psychology and Education Sciences, University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal. Email: carloscampos@fpceup.pt.

Acknowledgments

We would like to acknowledge: Maria J. Trigueiro, Bruno Melo, Filipa Carvalho, and Mariana R. Pereira for their contributions in the translation of the IAS and the BPQ; Jennifer Murphy, Stephen W. Porges, and Jacek Kolacz for providing their inputs on the cross-cultural adaptation of the questionnaires; Catarina Fernandes and Ana Correia for their support in data collection.

Abstract

This study included an online community-sample ($n = 515$) and aimed to: (1) explore the factor structure and psychometric properties of the European Portuguese versions of the Interoceptive Accuracy Scale (IAS) and the Body Perception Questionnaire (BPQ) - Body Awareness and Autonomic Reactivity scales; (2) examine the association between self-reported interoceptive attention and accuracy (indexed by the BPQ Body Awareness and IAS, respectively). Parallel-analysis indicated a one-factor solution for the BPQ Body Awareness. For the IAS and BPQ Autonomic Reactivity, bifactor ESEM models were retained. Ancillary measures revealed that both scales were mainly unidimensional and presented reliable total scores. All scales displayed excellent internal consistency, although test-retest reliability was modest. There was a quadratic U-shaped association between the IAS and BPQ Body Awareness. Alexithymia was negatively correlated with the IAS and unassociated with the BPQ Body Awareness. These findings suggest that interoceptive attention and accuracy may be dissociated using self-report questionnaires.

Keywords: interoception; interoceptive attention; interoceptive accuracy; quadratic association; factor analysis; bifactor models; dimensionality

Introduction

Interoception can be broadly defined as the perception and subjective experience of one's internal bodily states (Ceunen et al., 2013; Craig, 2002). Contemporary definitions argue that the central nervous system is capable of processing and integrating a wide array of signals from within the body, allowing a continuously updated model of our inner landscape (Khalsa et al., 2018). These interoceptive signals encompass physiological processes that typically take place on unconscious levels (e.g., cardiovascular, respiratory, gastrointestinal), conscious nonpainful stimuli (e.g., temperature, muscle tension, itch, tickle), as well as painful inputs stemming from visceral, somatic, and skeletal pathways.

In recent years, there has been a growing interest in interoception in psychology and neurosciences, as this construct has provided valuable insights to understand cognition and fine-tune theories of neural functioning (e.g., Critchley & Harrison, 2013; Seth & Friston, 2016). Interoception has also been linked to processes such as visual perception and action states (Marshall et al., 2017, 2018), as well as with higher-order cognitive domains, such as emotion processing (Barrett, 2016; Pereira et al., 2019) and social cognition (Fotopoulou & Tsakiris, 2017; Ondobaka et al., 2017). Furthermore, interoception has been implicated in a wide range of psychiatric conditions (e.g., mood and anxiety disorders) and transdiagnostic traits such as alexithymia (Khalsa et al., 2018; Murphy et al., 2017; Trevisan et al., 2019).

Theoretical Frameworks for Interoception Measurement

The ongoing increase in interoception-related research has led scholars to propose various frameworks that aim not only to define interoceptive domains but also to address the challenge of measuring individual differences in interoception. Although interoception has been consistently described as a multifaceted construct, there is still no clear consensual taxonomy. Interoceptive awareness has been frequently used as a broad term for distinct interoceptive components that can be consciously available, which can evidently influence findings regarding self-report measures of interoception (Khalsa et al., 2018).

Ceunen et al. (2013) first introduced the discussion addressing the dissociation between interoceptive awareness and interoceptive accuracy, that is, being cognizant of interoceptive inputs vs. accurately perceiving changes in bodily signals. Subsequently, a more complex three-dimensional

framework to measure interoception was suggested and included the following domains (Garfinkel et al., 2015; Garfinkel & Critchley, 2013): (a) interoceptive accuracy, described as the ability to accurately detect and track internal bodily sensations (measured using objective tasks); (b) interoceptive sensitivity, comprising the self-perceived dispositional tendency to be interoceptively focused as well as individual's beliefs about interoceptive abilities (measured through self-reported questionnaires or confidence ratings during objective interoception tasks); (c) interoceptive awareness, defined as the metacognitive ability to be aware of our own interoceptive accuracy (indexed by the relationship between objective interoceptive accuracy and subjective interoceptive sensibility). Although this is likely the most widespread framework for interoception measurement, differential associations between interoceptive awareness and clinical outcomes could not still be properly understood within this model (Garfinkel et al., 2016; Rae et al., 2019).

More recently, Murphy et al. (2019) proposed a 2 x 2 factorial framework to account for individual differences in interoception (Figure 1). The first factor addresses what is being measured, namely interoceptive attention or interoceptive accuracy. Interoceptive attention is defined as the extent to which interoceptive inputs are the target of our attention, while interoceptive accuracy refers to the degree we can accurately perceive our true bodily states. Conversely, the second factor distinguishes how interoception is measured, that is, whether interoception is evaluated via self-report or objective measures. This model further refines the previous three-dimensional model due to several reasons. First, as previously described, interoceptive sensibility encompassed one's beliefs regarding both interoceptive accuracy (e.g., confidence ratings on accuracy tasks) and interoceptive attention (e.g., self-report measures). This new model clearly distinguishes the ability to accurately perceive interoceptive information from the predisposition to allocate attentional resources to bodily signals. Secondly, the previous three-dimensional model framed interoceptive awareness as the correspondence between interoceptive accuracy and sensibility, regardless of whether the measure of interoceptive sensibility was measuring beliefs about attention or accuracy. This was a major issue, as interoceptive awareness should be a metacognitive measure assessing the correspondence between objective and subjective measures of the same underlying domain. The 2 x 2 factorial model proposes two distinct metacognitive measures, that is, awareness of interoceptive attention

and awareness of interoceptive accuracy. Hence, this newly developed framework can be a valuable tool to further understand the role of interoception in behavior and psychopathology.

[INSERT FIGURE 1 ABOUT HERE]

Psychometric Proprieties of Self-Report Measures for Interoceptive Attention and Accuracy

Self-report questionnaires to measure interoceptive attention and accuracy are still scarce, as are studies exploring their psychometric properties. Two of the instruments that can be used to measure these constructs are the Body Perception Questionnaire (BPQ) and the Interoceptive Accuracy Scale (IAS).

The BPQ was developed over 25 years ago aiming to evaluate subjective experiences directly and functionally linked to the autonomic nervous system, a set of neural pathways connecting the brain and the body (Porges et al., 1993, 2015). The original version encompassed several domains (e.g., cognitive emotional-somatic stress response, body and cognitive stress response styles) but has recently been reduced to include only body awareness and autonomic reactivity. The BPQ Body Awareness scale is a sensitivity measure for internal bodily functions, reflecting underlying afferent projections about the status of several organs and tissues. On the other hand, the BPQ Autonomic Reactivity evaluates responses of autonomically-innervated organs and tissues, which are driven by efferent control systems. The BPQ has been widely used as a self-report measure of interoception (e.g., Herman et al., 2019; Petzschner et al., 2019; Salvato et al., 2019; von Mohr et al., 2020), even if its application across fields has been far from being consistent and theoretically sound.¹ Moreover, although the BPQ has been translated to several languages, until now there are only two published studies exploring the psychometric properties of the full questionnaire (Cabrera et al., 2018; Cerritelli et al., 2021). In the original validation study, factor analysis across three samples suggested that a one-factor solution was suitable for BPQ Body Awareness, while BPQ Autonomic Reactivity reflected two distinct factors, namely supradiaphragmatic reactivity and subdiaphragmatic reactivity. Cabrera et al. (2018) also reported good internal consistency and excellent test-retest reliability across these subscales. However, this first psychometric endeavor should be interpreted with caution for two main reasons: firstly, the authors conducted an exploratory factor analysis (EFA) for each scale - Body Awareness and Autonomic Reactivity - separately, but subsequently conducted

a CFA (in another sample) encompassing all BPQ items within a single three-factor solution; secondly, the authors conducted factor analysis using a binary scoring system instead of the full 5-point Likert scale, as the latter resulted in extremely complex structures with a very large number of factors. The more recent psychometric study of the Italian version of the BPQ also used the binary scoring system (Cerritelli et al., 2021). Thus, these uncommon options for psychometric analysis should be contrasted with alternative strategies exploring the dimensionality and factor structure of the BPQ.

It is also important to highlight that Murphy et al. (2019) labeled the BPQ Body Awareness as a self-report measure of interoceptive attention within the 2 x 2 factorial model. These authors conducted a robust CFA for the BPQ Body Awareness using the 5-point scoring system, which indicated that the one-factor solution provided an acceptable but imperfect model (Murphy et al., 2020). There is also evidence suggesting excellent internal consistency and acceptable test-retest reliability for the BPQ Body Awareness using the full scoring system (Gabriele et al., 2020; Murphy et al., 2020).

Finally, the IAS was recently developed and validated by Murphy et al. (2020) as a self-report trait-based measure assessing perceived interoceptive accuracy as advocated by their 2 x 2 factorial model. In this validation study, principal component analysis suggested a two-factor solution for this scale, although there was no clear theoretical rationale for these factors. Subsequent confirmatory factor analysis (CFA) with another sample indicated that the two-factor solution provided an acceptable although imperfect explanation of the data. Thus, despite evidence for good internal consistency and test-retest reliability, there is still a need for further psychometric testing on the IAS (Gabriele et al., 2020; Murphy et al., 2020).

Untangling Self-Reported Interoceptive Attention and Accuracy

Despite the recent development of the 2 x 2 factorial model of interoception, there is already evidence indicating that self-report measures of interoceptive attention are uncorrelated with self-reported interoceptive accuracy (Murphy et al., 2020). The latest findings even suggest that greater attention to interoceptive signals may be associated with lower interoceptive accuracy, further reinforcing these constructs as dissociable (Gabriele et al., 2020). Furthermore, it is important to highlight that these studies

measured interoceptive attention and accuracy using the BPQ Body Awareness and IAS (respectively), although alternative questionnaires revealed similar findings (Gabriele et al., 2020).

It is also important to highlight that our team recently conducted an exploratory analysis using data from Murphy et al. (2020) where we found a quadratic association between interoception attention and accuracy (Campos et al., 2020). More precisely, both high and low levels of self-reported interoceptive attention (quantified by the BPQ Body Awareness) were associated with enhanced self-reported interoceptive accuracy (indexed by the IAS and other interoceptive accuracy measures). Whilst this quadratic association supports the assertion that these measures may assess different constructs, it also suggests that they may be differentially associated at different degrees of interoceptive attention.

Further clarification about the dissociation of self-reported interoceptive attention and accuracy can also be provided by exploring the link between both constructs and alexithymia. Alexithymia can be defined as difficulties to identify, describe, and interpret one's own emotional experiences (Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994) and has been widely related to emotional embodiment and interoception (Scarpazza & di Pellegrino, 2018). Recent meta-analytical evidence suggests a negative correlation between interoception and alexithymia, although the direction of this association was moderated by the specific component of interoception being measured (Trevisan et al., 2019). Self-reported interoceptive accuracy scores were negatively associated with alexithymia, while the BPQ Body Awareness (presumably indexing self-reported interoceptive attention) displayed a positive association, although there were only a few studies for the latter analysis. Murphy et al. (2020) also reported a large negative correlation between the IAS and alexithymia, even after controlling for other factors such as self-esteem, depression, and anxiety. Conversely, these authors reported no association between BPQ Body Awareness and alexithymia.

Finally, sociodemographic variables may also provide important insights to dissociate these interoception-related domains. While age has been reported to be negatively correlated with both interoceptive attention and accuracy (Khalsa et al., 2009; Murphy et al., 2020; Murphy, Geary, et al., 2018), sex appears to be differentially associated with both constructs. More specifically, it has been argued that

interoceptive attention may be higher in females, while interoceptive accuracy is enhanced in male subjects (Murphy, Viding, et al., 2019).

The Present Study

As previously mentioned, both studies exploring the psychometric proprieties of the BPQ had several limitations, while the IAS is a fairly recent scale that still requires further psychometric testing. Thus, the first major goal of this study was to translate, validate, and examine the psychometric properties of the European Portuguese versions of the BPQ and the IAS. This is intrinsically valuable as validating these measures will provide additional tools to assess interoception for researchers and clinicians within the field in Portugal. Moreover, a data-driven approach was implemented to identify how many specific factors are associated with each scale, as well as modern factor analytic techniques aiming to test several models and evaluate scale dimensionality. To further address the limitations of previous psychometric testing for the BPQ, factor structure analysis was conducted separately for the Body Awareness and Autonomic Reactivity subscales, as we consider them clearly distinct for several reasons.² Moreover, the analysis was conducted using full item scoring as it represents the scale originally presented to participants during data collection and provides more sensitivity to individual differences. Regardless, supplemental analyses using the binary scoring system were also completed for comparability purposes.

Finally, the second goal of this study was to examine the association between self-reported interoceptive attention and accuracy, as measured by the BPQ Body Awareness and the IAS, respectively. Additional evidence disentangling self-reported interoceptive attention and accuracy stemming from other countries is critical to further establish these constructs as dissociable and strengthen the theoretical rationale of the 2 x 2 factorial model. To confirm this dissociation, a lack of a significant correlation between the BPQ Body Awareness and the IAS was hypothesized, although a quadratic U-shaped association was also expected between these scores. Furthermore, alexithymia should be differentially associated with self-reported interoceptive attention and accuracy, that is, a positive association with BPQ Body Awareness and a negative relation with the IAS. Consequently, the correlation between the BPQ Body Awareness and alexithymia should be significantly different from the correlation between the IAS and alexithymia.

Altogether, exploring the association between the BPQ Body Awareness and the IAS, as well as the relationship of both scales with alexithymia is critical to establish their convergent and divergent validity. Finally, examining the relation between sociodemographic variables and self-reported interoceptive attention and accuracy can also be useful to dissociate these constructs. It was hypothesized that BPQ Body Awareness scores would be greater in female participants, IAS scores would be higher in male subjects, and both scale scores would be negatively correlated with age.

Methods

This observational study encompassed two stages. First, subjects completed a full version of the online survey, encompassing sociodemographic data and the questionnaires described in the instruments subheading.³ Participants who provided their contact information were subsequently emailed to complete stage 2, where they were only required to complete the BPQ and the IAS to assess test-retest reliability and agreement for both questionnaires. This study was preregistered at <https://osf.io/5jhcw> and <https://osf.io/ktpzc>.⁴ Datasets and supplemental materials are also available at <https://osf.io/hmtz9/>.

Sampling & Participants

Subjects were recruited using a non-list based, non-probability sample. As suggested by Callegaro et al. (2015) participant recruitment was spread as broadly as possible, using multiple recruitment channels such as social media, as well as personal contacts (snowball sampling). Summary statistics were accessed periodically to check whether recruited participants were sociodemographically heterogeneous, allowing to develop advertising strategies to target specific communities of participants when required (e.g., using recruitment channels that typically attract more male or female subjects as needed).

In the first stage of data collection, 515 participants⁵ aged between 18 and 72 were recruited ($M_{\text{age}} = 30.74$; $SD = 10.52$; 59.61% female). The rationale for sample size calculation is available at <https://osf.io/5jhcw>. Education level was reasonably high ($M_{\text{years}} = 15.03$; $SD = 2.99$) and a fair number of participants were university students (30.68%). A total of 89 subjects (17.28%) reported current or past psychiatric conditions, while 12 described neurological disorders. Moreover, 88 subjects (17.09%) reported other chronic health conditions (cardiovascular, respiratory, gastrointestinal, among others). Finally, only

6 subjects considered that Portuguese was not their first language. A subsample of 134 subjects completed stage 2 of data collection for test-retest analysis ($M_{\text{age}} = 32.28$; $SD = 11.19$; 54.48% female).

Instruments

Body Perception Questionnaire (BPQ; Porges et al., 1993, 2015). This questionnaire includes two main domains: BPQ Body Awareness (26 items), which can be used as a self-report measure of interoceptive attention, with higher scores indicating that interoceptive inputs are more frequently the target of attention; BPQ Autonomic Reactivity (20 items), which evaluates the responses of autonomically-innervated organs, with higher scores implying greater reactivity. The original validation (Cabrera et al., 2018) suggested a 2-factor structure for the BPQ Autonomic Reactivity, dissociating responses of organs above the diaphragm (supradiaphragmatic reactivity, 15 items) and responses of gastrointestinal organs below the diaphragm (subdiaphragmatic reactivity, 6 items), although item 41 was loaded and scored on both factors. BPQ items were rated on a 5-point Likert scale ranging from never (1) to always (5).

Interoceptive Accuracy Scale (IAS; Murphy et al., 2020). Self-report scale of perceived interoceptive accuracy. This scale comprises 21 items rated on a 5-point Likert scale, ranging from strongly agree (5) to strongly disagree (1). Higher scores imply greater interoceptive accuracy.

20-Item Toronto Alexithymia Scale (TAS-20; Bagby, Parker, et al., 1994; Bagby, Taylor, et al., 1994). Self-report scale for alexithymia comprising 20 items, rated on a 5-point Likert scale, and ranging from strongly disagree (1) to strongly agree (5). Higher scores imply greater alexithymia traits. The factor structure of the European Portuguese version (Prazeres et al., 2000, 2008) encompasses three subscales: Difficulty Identifying Feelings (7 items), Difficulty Describing Feelings (5 items), and Externally-Oriented Thinking (8 items), but only total scores were used for hypothesis testing.

Procedures

Questionnaire Translation and Cross-Cultural Adaptation

The translation and adaption of the BPQ and the IAS were conducted according to established procedures aiming to maximize semantic and normative equivalence between the original English and the European Portuguese versions (Behling & Law, 2000). This process encompassed the following stages:

(1) translation - two external researchers proficient in both languages completed independent translations of the original English questionnaires, which were then reviewed by the first author to develop a consensus version; (2) piloting and thinking aloud - the provisional versions were administered to a sample of 6 university students followed by a semi-structured interview (debriefing with a thinking aloud format) to ensure that all items were properly understood; results from this pilot and required changes were subsequently discussed by the research team; (3) back-translation - the adapted Portuguese versions were back-translated to English by an external researcher who is a native speaker of both languages and had no previous contact with any of the questionnaires; (4) comparison with the original version - the back-translated version of each questionnaire was sent to the original authors for comparison and an online meeting was conducted to debate any required changes. An external researcher proficient in both languages further provided additional input to fine-tune some of the items. The final version of each questionnaire was approved by the original authors and our research team. A further detailed description of the translation and adaptation procedures can be found in the preregistration (<https://osf.io/5jhcw>).

Data Collection

This study was approved by the institutional Ethics Committee of ESS|P.PORTO and the Data Protection Officer. Data was collected using an online survey developed in a web-based tool (LimeSurvey, version 3.22.18+200603, Hamburg, Germany), which was piloted with 7 subjects (3 female, different age groups and educational levels) prior to proceeding with data collection. Before starting the online survey, participants received a brief explanation about the purpose of the study as well as an electronic consent form. After completing stage 1 of data collection, participants were asked to provide their contact information for the subsequent stage, if they were interested. Participants were invited to enroll in stage 2 by email, 30 days after completing the first stage. Participants received reminders for completing the second stage and were excluded if their responses were submitted more than 40 days after stage 1. To avoid order effects, questionnaires were presented in a randomized order, after the consent form and sociodemographic data collection. Participants did not receive any sort of compensation for their participation.

Data Analysis

Full description of statistical models, assumption testing, and other analytical details can be found at the preregistration. The analysis was initially conducted with the full sample (only excluding subjects that reported less than 18 years of age) to examine the preregistered hypothesis. Subsequently, the analysis was replicated excluding careless respondents, outliers, and/or other possible confounders (non-native speakers, self-report history of psychiatric or neurological conditions). Replicating the analysis under these conditions allowed to understand whether any of these factors played a role in the attained results.

Factor Structure Analysis. The strategy was to first identify how many and which factors were driven by the current dataset for each questionnaire using parallel analysis and EFA. This data-driven approach was selected because, although the translation and cross-cultural adaptation procedures were extremely thorough, items can still cluster differently when compared to the original versions. Parallel analysis was conducted to determine the number of factors for each scale using the procedures recommended by Garrido et al. (2013) for ordinal variables. Parallel analysis compares the eigenvalues of factors retrieved from the observed data with those from a random data matrix with similar size. Parallel analysis was implemented using MATLAB Release 2020 (The MathWorks, Inc., Natick, Massachusetts), employing 100 permutations of the real dataset and smoothing of non-positive definite polychoric matrices (Garrido, 2017)

Subsequently, an EFA using a robust weighted least squares with mean and variance adjustment (WLSMV) estimation was implemented in MPlus 8.1.5 (Muthén & Muthén, 1998-2017) for each scale. This estimator uses the polychoric correlation matrix and is more suitable for ordinal variables, including binary responses (Barendse et al., 2015; Gana & Broc, 2019; Muthén, 1983). An oblique geomin rotation was selected as there is evidence suggesting that it outperforms other oblique rotations when little is known about the factor structure (Asparouhov & Muthén, 2009). The number of extracted factors for each questionnaire was set accordingly to the results from the parallel analysis. For subsequent confirmatory models, each item was attributed to the factor where it loaded the most.

Finally, based on the analytical strategy described by Morin et al. (2016), CFA and exploratory structural equation modeling (ESEM) were used to compare different structures encompassing the factors retrieved from parallel analysis and EFA. This approach was critical to assess scale dimensionality, that is,

whether each scale was mainly unidimensional or if there were putative sources of construct-relevant psychometric multidimensionality. Thus, the following models were implemented:

1. One-factor CFA model - good fit would suggest a unidimensional model and even if the one-factor model did not provide a better fit than other models, it could still be the most adequate and parsimonious alternative as it implies a simpler solution.

2. First-order CFA and ESEM models, allowing factors to correlate - CFA models usually rely on restrictive criteria in which cross-loadings between items and nontarget factors are assumed to be zero, which is not the case for ESEM models.

3. Bifactor models, allowing specific factors to correlate (CFA and/or ESEM, accordingly to the selected first-order model) - in these models, the common variance among all items is accounted by a general factor and variance that is only shared among certain items is attributed to specific lower-level factors.

4. Hierarchical models (CFA and/or ESEM, accordingly to the selected first-order model) - Each item is specified as loading on its specific first-order factor, and each first-order factor loads on a higher-order factor.

5. Previously reported models were also implemented, namely the two-factor CFA (allowing factors to correlate) for BPQ Autonomic Reactivity (Cabrera et al., 2018) and the IAS (Murphy et al., 2020).

Models were implemented using WLSMV estimation in MPlus 8.1.5. For ESEM models, oblique rotations (geomin or bi-geomin) were used. Factor loadings were extracted using a completely standardized solution. Models were compared using several goodness of fit statistics, information criteria, examination of parameter estimates, and theoretical considerations (Gana & Broc, 2019; Hu & Bentler, 1999; Kline, 2016). Chi-square difference tests were also used to compare nested models, while non-nested models were compared using information criteria. If the bifactor model adequately fit the data, ancillary bifactor measures were computed to assess scale dimensionality and model-based reliability of scale scores (Reise et al., 2013; Rodriguez et al., 2016a, 2016b) using the spreadsheet of Dueber (2017). Table 1 provides a detailed description of the goodness of fit statistics, information criteria, and ancillary bifactor measures.

[INSERT TABLE 1 ABOUT HERE]

Internal Consistency. Internal consistency was computed using the MBESS R package version 4.8.0 (Kelley, 2007; Kelley, 2020) ran in R version 4.0.3 (R Core Team, 2020). The categorical omega coefficient ($\omega_{\text{categorical}}$) was used to assess internal consistency as it is more suitable for ordered-categorical items (Kelley & Pornprasertmanit, 2016). Cronbach's alpha was also computed for comparability with previous studies (Cronbach, 1951). Confidence intervals for these measures were computed using bias-corrected and accelerated bootstrapping with 1000 replications. Internal consistency was classified using the following cut-off values: < .50 unacceptable; $\geq .50$ and < .60 poor; $\geq .60$ and < .70 questionable; $\geq .70$ and < .80 adequate/acceptable; $\geq .80$ and < .90 good; $\geq .90$ excellent (Kline, 2016).

Test-Retest Analysis. Test-retest analysis was implemented using IBM SPSS version 27 (IBM Corporation, Armonk, New York, USA) and aimed to assess both test-retest reliability (the ratio of true variance over true plus error variance; Koo & Li, 2016) and test-retest agreement (capacity of the scale to provide exactly identical results when applied twice under equivalent conditions; Berchtold, 2016).

Test-retest reliability was evaluated using the intraclass correlation coefficient (ICC) based on an absolute-agreement, two-way mixed-effect model (Koo & Li, 2016). Values less than .50 are indicative of poor reliability, values $\geq .50$ and < .75 indicate moderate reliability, values $\geq .75$ and < .90 indicate good reliability, and values $\geq .90$ indicate excellent reliability. ICCs were reported together with 95% confidence intervals and formally compared between scales using Fisher's Z test (Donner & Zou, 2002), implemented using the quantpsy web implementation (Lee & Preacher, 2013a). Pearson's correlation coefficients were also computed for comparability with previous studies.

Test-retest agreement was quantitatively examined using the concordance correlation coefficient (Lin, 1989, 2000), computed using an SPSS macro (Garcia-Granero, 2005). The CCC was classified according to the classification proposed by Landis & Koch (1977): < .21 poor/slight; $\geq .21$ and < .40 fair; $\geq .41$ and < .60 moderate; $\geq .61$ and < .80 substantial; $\geq .81$ good/excellent. Spearman rank correlation coefficients were also computed to assess whether there are systematic differences between the mean score and the test-retest difference scores (Bonett & Wright, 2000).

Correlational Analysis and Independent Group Comparisons. These analyses were implemented using IBM SPSS version 27. Pearson's correlation coefficients were used to test the associations between self-reported interoceptive attention, interoceptive accuracy, alexithymia, and age. According to Gignac and Szodorai (2016), Pearson's correlation coefficients were classified as small ($\geq .10$), medium ($\geq .20$), and large ($\geq .30$). These correlations were also formally compared using Steiger's Z-test for dependent correlations (Steiger, 1980), implemented via the quantpsy web implementation (Lee & Preacher, 2013b). Independent-sample t-tests were used to compare self-reported interoceptive attention and accuracy between male and female subjects. Effect sizes were computed using Hedges' g (Borenstein, 2009) and classified as suggested by Cohen (1988): small = $|.20|$, medium = $|.50|$ and large = $|.80|$.

Quadratic Association: Polynomial Regression and Two-Lines Testing Approach. A polynomial regression analysis was used to examine the quadratic association between self-reported interoceptive attention and accuracy. Polynomial regression adds to the model extra predictors that are powers of the original independent variable (Garson, 2012; Gignac, 2019; Wuensch, 2006). Thus, the linear model was evaluated first and, subsequently, the quadratic term of the predictor (squared BPQ Body Awareness scores) was added to the model to test whether it increased model fit (significant R^2 changes). A cubic term (cubing BPQ Body Awareness scores) was further added to examine additional model fit changes. As suggested by Garson (2012), besides testing for significant R^2 changes, the paired t-tests of residual differences method were also used to compare linear, quadratic, and cubic models.

Finally, Simonsohn (2018) argued that quadratic regressions are not suitable to test for U-shaped associations, leading to a high rate of false positives. The authors developed an alternative approach designated two-lines testing to examine U-shaped relationships. In this method, an algorithm is used to estimate the breakpoint for the U-shape, allowing to identify "low" and "high" values for the predictor variable. Then, two regression slopes are estimated (one for each set of values). This allows testing whether each regression slope is individually significant and if there is a sign change between them (positive and negative or vice-versa), which would suggest a U-shaped association. This approach was tested using the

web-based app developed by the authors (<http://webstimate.org/twolines/>). The two-lines testing approach was used to test for a U-shaped relationship between self-reported interoceptive attention and accuracy.

Results

The results presented here refer to the full sample analysis (only excluding subjects that reported less than 18 years of age), although additional control analyses were conducted excluding careless respondents, outliers, and other confounders.⁶ Detailed statistical reports are available at <https://osf.io/hmtz9/>.

Factor Structure Analysis

BPQ Body Awareness

Parallel analysis and EFA results indicated only one factor for the BPQ Body Awareness scale. The one-factor model resulted in a marginally acceptable, although imperfect solution, as suggested by several fit statistics, $\chi^2(299) = 1657.70, p < .001$, WRMR = 1.61, RMSEA = .094, [.090, .098], CFI = .947, TLI = .943. Factor loadings (see Table 2) were all statistically significant and higher than .400, $\lambda = [.461, .883]$. Removing the item with the lowest loading, item 1 (“swallowing frequently”), did not produce any significant changes in model fit, $\chi^2(275) = 1502.02, p < .001$, WRMR = 1.57, RMSEA = .093, RMSEA $CI_{90\%} = [.089, .098]$, CFI = .952, TLI = .948. Thus, despite this imperfect solution, the one-factor model for the BPQ Body Awareness scale was retained as there was no previous theoretical rationale nor data-driven information from the current dataset arguing for a multiple factor model.

[INSERT TABLE 2 ABOUT HERE]

BPQ Autonomic Reactivity

Parallel analysis and EFA results indicated two specific factors for the BPQ Autonomic Reactivity, with items loading quite similarly to the previous findings (Cabrera et al., 2018). Therefore, subsequent CFA and ESEM models were tested using 2-factor solutions.⁷ Fit and information criteria for the BPQ Autonomic Reactivity models are presented in Table 3. Model fit was marginally adequate for the correlated 2-factor CFA model, $\chi^2(169) = 775.44, p < .001$, WRMR = 1.53; RMSEA = .083, CFI = .927, TLI = .917, as well as the correlated 2-factor ESEM model, $\chi^2(151) = 713.47, p < .001$, WRMR = 1.32, RMSEA = .085, CFI = .932, TLI = .914. Although χ^2 difference testing indicated that the correlated 2-factor ESEM

model had significantly better fit than correlated 2-factor CFA model, $\chi^2 (18) = 110.83, p < .001$, confidence intervals for the RMSEA showed a significant overlap between the competing models, CFA RMSEA $CI_{90\%} = [.078, .089]$, ESEM RMSEA $CI_{90\%} = [.079, .091]$, suggesting that there was not a clear differentiation between these models. Thus, it was decided not only to test the bifactor ESEM model for the BPQ Autonomic Reactivity but also the bifactor CFA model⁸, which would be particularly important to examine whether the specific factors could be reliably used as subscales for supradiaphragmatic and subdiaphragmatic reactivity.

The bifactor ESEM model displayed good fit according to all fit statistics, $\chi^2 (133) = 357.63, p < .001$, WRMR = 0.85, RMSEA = .057, CFI = .973, TLI = .961, and χ^2 difference testing indicated that this model had significantly better fit than the correlated 2-factor ESEM, $\chi^2 (18) = 271.71, p < .001$. Ancillary measures for the ESEM model also suggested that BPQ Autonomic Reactivity can be primarily conceptualized as a unidimensional scale, $ECV_{Gen} = .747$, ARPB = 11.4%, and that its total score predominantly reflects a single general factor, $\omega H = .898$. Similar results were also found for the bifactor CFA model, which exhibited good fit as measured by all fit statistics, $\chi^2 (149) = 382.25, p < .001$, WRMR = 0.95, RMSEA = .055, CFI = .972, TLI = .964, and better fit than the correlated 2-factor CFA model, $\chi^2 (20) = 285.66, p < .001$. Within this model, ancillary dimensionality measures suggested that the BPQ Autonomic Reactivity is mainly unidimensional, $ECV_{Gen} = .759$, ARPB = 10.7%. Importantly, although the total score for this scale predominantly reflected a single general factor, $\omega H = .895$, subscale scores were only marginally associated with unique variance from the specific supradiaphragmatic and subdiaphragmatic reactivity factors, $\omega HS = .001$ and $\omega HS = .408$, respectively. These results suggest that BPQ Autonomic Reactivity subscales scores for supradiaphragmatic and subdiaphragmatic reactivity factors are not reliable, as they are mostly re-measuring the general factor.

Comparing the bifactor models, chi-square difference testing indicated that the bifactor ESEM model had significantly better fit than the bifactor CFA model, $\chi^2 (16) = 46.26, p < .001$, even though confidence intervals for the RMSEA showed a significant overlap between the competing models, CFA RMSEA $CI_{90\%} = [.048, .062]$, ESEM RMSEA $CI_{90\%} = [.050, .064]$. The bifactor ESEM model displayed significantly better

fit in comparison to the one-factor solution, $\chi^2(37) = 812.70, p < .001$, as well as the 2-factor model reported by Cabrera et al. (2018), $\chi^2(35) = 300.50, p < .001$. Thus, the bifactor ESEM model was considered the most appropriate model to represent the BPQ Autonomic Reactivity data in the current sample.

[INSERT TABLE 3 ABOUT HERE]

Factor loadings and item-level ancillary measures for the BPQ Autonomic Reactivity models are displayed in Table 4. In the bifactor ESEM model, factor loadings on the general factor were all statistically significant and higher than .400, except for item 43 (“I am constipated”), $\lambda = [.384, .806]$. Most items contributed to the general factor more than the specific factors, $ECV_{Ind} = [.408, .999]$, although items 31 (“I feel shortness of breath”), 44 (“I have indigestion”) and 45 (“After eating I have digestive problems”) displayed ECV_{Ind} values lower than .500. Also, a total of 12 items were very strong measures of the general dimension (i.e., $ECV_{Ind} > .800$).

[INSERT TABLE 4 ABOUT HERE]

Factor Structure for BPQ Scales using the Binary Scoring System

Regarding the BPQ Body Awareness scale, it was not possible to run models when using the binary scoring system because the bivariate table produced several empty cells for items 21 (“General jitteriness”) and 26 (“Feeling constipated”), which would require to exclude these items to proceed with further analysis. Moreover, the correlation matrix was not positive definite which indicated that this scoring system originated too many highly correlated items.

Conversely, results from factor structure analysis for the BPQ Autonomic Reactivity when using the binary scoring system were quite similar to the abovementioned findings for the full scoring system (detailed information at <https://osf.io/hmtz9/>). Parallel analysis and EFA results also suggested the same two specific factors for this scale, with each item loading in the same factor as described for the full scoring system. Likewise, the solutions which provided better fit for binary scoring BPQ Autonomic Reactivity were also the bifactor CFA, $\chi^2(149) = 258.14, p < .001$, WRMR = 0.91, RMSEA = .038, RMSEA CI_{90%} = [.030, .045], CFI = .984, TLI = .979, and the bifactor ESEM models, $\chi^2(133) = 208.49, p < .001$, WRMR = 0.74, RMSEA = .033, RMSEA CI_{90%} = [.024, .042], CFI = .989, TLI = .984, although χ^2 difference

testing indicated that the bifactor ESEM model had significantly better fit than the bifactor CFA model, $\chi^2(16) = 43.59, p < .001$. Ancillary measures for the ESEM model also suggested that binary scoring BPQ Autonomic Reactivity can be conceptualized as a unidimensional scale, $ECV_{Gen} = .767$, $ARPB = 6.3\%$, with its total score predominantly representing a single general factor, $\omega H = .925$. Unidimensionality, $ECV_{Gen} = .719$, $ARPB = 14.6\%$, and total score reliability, $\omega H = .859$, was also acceptable in the bifactor CFA model, although subscale scores for supradiaphragmatic and subdiaphragmatic reactivity factors were clearly unreliable ($\omega H = .070$ and $\omega H = .484$, respectively), as previously observed for the full scoring system. Finally, in the bifactor ESEM model, factor loadings on the general factor were all statistically significant and higher than .500, $\lambda = [.550, .822]$ and all items (except item 31 - “I feel shortness of breath”) contributed to the general factor more than the specific factors, $ECV_{Ind} = [.492, .996]$.

Interoceptive Accuracy Scale

Parallel analysis and EFA results indicated three specific factors for the IAS. Hence, subsequent CFA and ESEM models were tested for 3-factor solutions. Fit statistics and information criteria for the IAS models are presented at Table 5. Model fit was marginally acceptable for both the correlated 3-factor CFA, $\chi^2(186) = 941.29$, $WRMR = 1.61$, $RMSEA = .089$, $RMSEA\ CI_{90\%} = [.083, .094]$, $CFI = .938$, $TLI = .929$, and the correlated 3-factor ESEM model, $\chi^2(150) = 665.51$, $WRMR = 1.12$, $RMSEA = .082$, $RMSEA\ CI_{90\%} = [.075, .088]$, $CFI = .957$, $TLI = .940$. However, all goodness of fit statistics indicated that the correlated 3-factor ESEM solution provided a better representation than the 3-factor CFA model, $\Delta WRMR = -0.49$, $\Delta RMSEA = -.007$, $\Delta CFI = -.019$, $\Delta TLI = -.011$. Also, χ^2 difference testing indicated that the correlated 3-factor ESEM had significantly better fit than correlated 3-factor CFA, $\chi^2(36) = 296.50, p < .001$. Thus, only the bifactor and hierarchical ESEM models were subsequently tested for the IAS.

The hierarchical ESEM model was not able to provide an adequate representation of the data, with fit statistics similar to the correlated 3-factor ESEM solution, $\chi^2(150) = 665.51$, $WRMR = 1.12$, $RMSEA = .082$, $RMSEA\ CI_{90\%} = [.075, .088]$, $CFI = .957$, $TLI = .940$. This result was expected as the variance explained by the correlations between the high-order factors and the three specific factors basically stems

from the variance allocated to the correlations between specific factors in the correlated 3-factor ESEM model (i.e., these models display the exact same number of free parameters).

The bifactor ESEM model displayed good fit according to all fit statistics, $\chi^2 (132) = 459.03$, WRMR = 0.86, RMSEA = .069, RMSEA CI_{90%} = [.063, .076], CFI = .973, TLI = .957, and χ^2 difference testing indicated that this model had significantly better fit than the correlated 3-factor ESEM, $\chi^2 (18) = 182.49$, $p < .001$. Ancillary measures also suggested that the IAS is mainly a unidimensional scale, ECV_{Gen} = .731, ARPB = 5.7%, and that its total score predominantly reflects a single general factor, $\omega H = .939$. The bifactor ESEM model displayed significantly better fit in comparison to the one-factor solution, $\chi^2 (57) = 1070.00$, $p < .001$, and to the 2-factor Murphy et al. (2020) model, $\chi^2 (56) = 842.50$, $p < .001$. Thus, the bifactor ESEM model was the most appropriate solution to represent the IAS data in the current sample.

[INSERT TABLE 5 ABOUT HERE]

Factor loadings and item-level ancillary measures for the IAS models are displayed in Table 6. In the retained bifactor ESEM model, factor loadings on the general factor were all statistically significant and higher than .400, $\lambda = [.460, .724]$. Furthermore, all items except item 20 (“tickle”) reflected the general factor more than the specific factors, ECV_{Ind} = [.465, .973], and a total of 9 items were very strong measures of the general dimension (i.e., ECV_{Ind} > .800).

[INSERT TABLE 6 ABOUT HERE]

Internal Consistency

Internal consistency was analyzed using the $\omega_{categorical}$, even though Cronbach’s Alphas were also computed for comparability purposes (see Table 7). Excellent internal consistency was observed for the BPQ Body Awareness, $\omega_{categorical} = .978$, CI_{99%} = [.972, .981], the BPQ Autonomic Reactivity, $\omega_{categorical} = .944$; CI_{99%} = [.922, .960], and the IAS, $\omega_{categorical} = .970$, CI_{99%} = [.951, .982],

[INSERT TABLE 7 ABOUT HERE]

Test-Retest Reliability and Agreement

Test-retest reliability was primarily assessed using the ICC, although Pearson’s correlations were also computed for comparability purposes (see Table 7). Good test-retest reliability was found for the BPQ

Autonomic Reactivity, $ICC = .832$, $CI_{99\%} = [.672, .908]$, and the IAS, $ICC = .756$, $CI_{99\%} = [.642, .837]$. Conversely, BPQ Body Awareness only displayed moderate test-retest reliability, $ICC = .684$, $CI_{99\%} = [.537, .789]$. Although there were no significant differences in test-retest reliability between the BPQ Body Awareness and the IAS, $Z = -1.68$, $p = .093$, or between the BPQ Autonomic Reactivity and the IAS, $Z = 1.22$, $p = .224$. BPQ Body Awareness actually displayed significantly decreased test-retest reliability in comparison to BPQ Autonomic Reactivity, $Z = -2.90$, $p = .004$.

Test-retest agreement was formally tested using the CCC, although Spearman correlation coefficients between test-retest mean scores and test-retest difference scores were also analyzed (see Table 7). Test-retest agreement was good for the BPQ Autonomic Reactivity, $CCC = .831$, $CI_{99\%} = [.750, .887]$, and substantial for the BPQ Body Awareness, $CCC = .682$, $CI_{99\%} = [.546, .783]$, and the IAS, $CCC = .754$, $CI_{99\%} = [.641, .835]$. Also, Spearman rank correlation coefficients between test-retest mean scores and test-retest difference scores were not significant, which suggests no proportional bias in any of the scales.

Correlational Analysis, Quadratic Association, and Group Comparisons

The correlation matrix for variables included in this study is presented in Table 8. Contrary to our hypothesis, the BPQ Body Awareness was significantly and positively correlated with the IAS, $r = .204$, $CI_{99\%} = [.084, .316]$, $p < .001$. As previously described, it was further explored whether the BPQ Body Awareness and the IAS were quadratically associated. Firstly, the linear regression model indicated that the BPQ Body Awareness was a significant predictor of the IAS, $R^2 = .042$, $F(1, 513) = 22.30$, $p < .001$. The addition of the quadratic term explained an additional 12.7% of variance in comparison to the linear model, $\Delta R^2 = .127$, $F(1, 512) = 78.38$, $p < .001$, while entering the cubic term did not produce significant changes, $\Delta R^2 = .001$, $F(1, 511) = 0.60$, $p = .440$. Paired t-tests of residual differences also indicated significant differences between the linear and quadratic models, $t(514) = 4.72$, $p < .001$, but no differences between the quadratic and cubic models, $t(514) = 1.14$, $p = .255$. Lastly, two-lines testing indicated a U-shaped association between the BPQ Body Awareness and the IAS (Figure 2), with a negative slope for lower BPQ scores and a positive slope for high BPQ scores, Line 1: $b = -0.196$, $z = -3.91$, $p < .001$, Line 2: $b = 0.416$,

$z = 8.79$, $p < .001$, breakpoint = 84. These results suggest a quadratic U-shaped association between the BPQ Body Awareness and the IAS.

[INSERT FIGURE 2 ABOUT HERE]

There were differential associations of the TAS-20 with the BPQ Body Awareness and the IAS. As predicted, the IAS was negatively correlated with TAS-20, $r = -.291$, $CI_{99\%} = [-.395, -.180]$, $p < .001$. Conversely, there was no significant association between the BPQ Body Awareness and TAS-20, $r = -.030$, $CI_{99\%} = [-.135, .098]$, $p = .500$. Importantly, the correlation between the BPQ Body Awareness and TAS-20 was significantly different from the correlation between the IAS and TAS-20, $Z = -4.30$, $p < .001$.

Further analyses were also conducted to examine the relation of sociodemographic variables (age and sex) with both self-reported interoceptive attention and accuracy. BPQ Body Awareness was significantly and negatively correlated with age, $r = -.157$, $CI_{99\%} = [-.267; -.028]$, $p < .001$. In contrast, there was no significant association between the IAS and age, $r = .081$, $CI_{99\%} = [-.030; .173]$, $p = .067$. Moreover, women displayed higher BPQ Body Awareness scores than men, $t(513) = 3.17$, $p = .002$, $g = 0.29$, and there were no significant sex differences regarding IAS, $t(513) = -1.31$, $p = .190$, $g = 0.12$.

Finally, although there were no formal hypotheses regarding BPQ Autonomic Reactivity, additional analyses were also computed for this scale. BPQ Autonomic Reactivity was positively associated with BPQ Body Awareness, $r = .216$, $CI_{99\%} = [.101, .325]$, $p < .001$, and negatively associated with IAS, $r = -.267$, $CI_{99\%} = [-.381, -.143]$, $p < .001$. Also, women displayed higher BPQ Autonomic Reactivity than men, $t(513) = 4.28$, $p < .001$, $g = 0.38$.

[INSERT TABLE 8 ABOUT HERE]

Discussion

In the last decade, there has been a growing interest about interoception across a wide array of disciplines. This led to an increasing need for adequate theoretical frameworks and measurement tools, that can aid researchers and clinicians to further understand interoception as well as its role on human behavior and psychopathology. Furthermore, additional translations for several languages and validation endeavors are required, as most interoception-related scales are only available in English. Thus, the current study had

two major goals: (a) translating and assessing the psychometric properties of the European Portuguese versions of the BPQ and the IAS; (b) dissociating self-reported interoceptive attention and accuracy, as measured by the BPQ Body Awareness and the IAS (respectively), as distinct interoception-related constructs.

Factor Structure Analysis of the BPQ and the IAS

Factor structure analysis of interoception-related instruments can be quite challenging as there is a thin line between perceptive experiences that fully rely on interoceptive inputs from those that can also be easily influenced by exteroceptive information. Consequently, in the current study, an *a priori* decision-making analytical framework that integrated data-driven methods for factor extraction, modern factor analysis techniques, and previous theoretical insights on interoception measurement was implemented. These methods allowed examining not only which model provided the best solution for the internal structure of each scale, but also to explore their unidimensional vs. multidimensional nature.

Within this analytical framework, parallel-analysis inherently suggested the BPQ Body Awareness is unidimensional, although model fit for the one-factor solution was not the most adequate. Similar findings have been previously reported for the BPQ Body Awareness using the full scoring system (Murphy et al., 2020). These results should be interpreted within the context of recent evidence by Gabriele et al. (2020) suggesting a clear lack of consistency among subjects when interpreting the instructions of the BPQ Body Awareness scale, as only 36.4% of participants considered that the scale was assessing interoceptive attention, while 30.4% interpreted the questionnaire as pertaining to interoceptive accuracy. Additionally, 30.8% of participants believed that the scale was assessing how often interoceptive signals objectively occur. This heterogeneous interpretation of BPQ Body Awareness instructions across participants may explain the imperfect model fit for the one-factor unidimensional solution.

Interestingly, previous psychometric studies of the BPQ Body Awareness reported good model fit for the one-factor solution when using binary scoring (Cabrera et al., 2018; Cerritelli et al., 2021). Cabrera et al. (2018) actually stated that using full item scoring resulted in extremely complex structures, which included a very large number of factors. The current results are completely opposite, as the full scoring

system allowed for a one-factor solution and it was not even possible to run models for binary scoring of the BPQ Body Awareness scale, because the bivariate table produced several empty cells and the correlation matrix for binary scoring was not positive definite, suggesting too many highly correlated items. Thus, we advise against the binary scoring system for the BPQ Body Awareness scale for the following reasons: (a) the binary scoring system does not allow for sufficient variability in the data; (b) factor structure analysis with the full scoring system suggested that this scale is unidimensional and can be used as a more adequate measure of interindividual differences in interoceptive attention; (c) the full scoring system represents the scale as it was originally presented to participants during data collection, which seems more adequate.

Regarding the BPQ Autonomic Reactivity, both the bifactor CFA and bifactor ESEM models provided good fits, although the latter was formally retained after statistical testing. Regardless, ancillary measures from bifactor models indicated that BPQ Autonomic Reactivity may be conceptualized primarily as a unidimensional instrument. Previous evidence from Cabrera et al. (2018) suggested that BPQ Autonomic Reactivity consisted of two distinct factors pertaining to supradiaphragmatic and subdiaphragmatic reactivity, although one item that did not conceptually fit its factor had to be excluded and item 41 (“I feel like vomiting”) was scored on both factors. This model was more recently tested on the Italian version of the BPQ using CFA and a good model fit was also achieved (Cerritelli et al., 2021). However, this factor structure was tested using the binary scoring system and the authors did not conduct any sort of dimensionality analysis to confirm the multidimensional nature of the scale. Also, Cabrera et al. (2018) actually stated that *“sympathetic efferent pathways innervate organs both above and below the diaphragm and thus the role of this system is difficult to interpret in light of the observed factor structure”*. The current results allow us to postulate whether the BPQ Autonomic Reactivity scale may display some degree of multidimensionality related to the gastrointestinal system, as item 43 (“I am constipated”) had a small loading on the general factor and items 44 (“I have indigestion”) and 45 (“After eating I have digestive problems”) displayed ECV_{Ind} values lower than .500. Notwithstanding, ancillary measures in the current analysis clearly suggest that although the total score of BPQ Autonomic Reactivity may be used as a general measure of autonomic reactivity, subscales scores should not be implemented as reliable measures of

supradiaphragmatic and subdiaphragmatic reactivity, as they are mostly re-measuring the general factor. Finally, it is important to highlight that the factor structure analysis of the BPQ Autonomic Reactivity provided quite similar results when using both the binary and full scoring system. Regardless, we still argue that researchers and clinicians should prefer the full scoring system due to the same reasons described for BPQ Body Awareness.

Finally, data pertaining to the IAS scale could be more adequately explained by the bifactor ESEM model. Moreover, ancillary measures confirmed the hypothesis that the IAS scale is mainly unidimensional, as almost all items (except one) were more strongly associated with the general factor, rather than any of the specific factors. The original validation of the IAS suggested a two-factor solution (stemming from principal component analysis), but the authors advised caution when using this structure, as it provided an imperfect model for the data (Murphy et al., 2020). The authors also argued that there was no clear rationale for these factors, although the first factor could reflect solely interoceptive signals, while the second either represented signals hard to perceive using only interoceptive information or socially unacceptable bodily functions (sneeze, cough, wind, bruise, blood sugar, tickle, itch). Interestingly, the lowest ECV_{Ind} values found in the current study were all associated with items previously included in the second factor described by Murphy et al. (2020). Thus, it is reasonable to assume that some responders interpret these items relying not only on their beliefs regarding interoceptive inputs but also considering exteroceptive sources of information (e.g., tactile) or even high-order cognitive mechanisms that can allow them to accurately perceive these bodily signals. Regardless, the current results suggest that all IAS items should be retained, as they are largely (or often) associated with beliefs regarding interoceptive information and clearly contribute to the broad general construct that is interoceptive accuracy.

Evaluating Other Psychometric Proprieties of the IAS and the BPQ

Effectively achieving a semantic equivalency when translating interoception-related questionnaires is challenging, especially when trying to dissociate between different constructs such as interoceptive attention vs. interoceptive accuracy. Thus, besides factor structure analysis, further psychometric testing was needed to examine whether these new translated versions can be reliably used. Regarding other

psychometric properties, internal consistency was excellent for every scale, ranging from .944 to .978. Previous studies also reported good to excellent internal consistency for the IAS, ranging from .88 to .90, and the BPQ Body Awareness, ranging from .92 to .98 (Cabrera et al., 2018; Cerritelli et al., 2021; Gabriele et al., 2020; Murphy et al., 2020). Regarding BPQ Autonomic Reactivity, previous results suggested larger internal consistency for supradiaphragmatic reactivity, ranging from .88 to .94, in comparison to subdiaphragmatic reactivity, ranging from .77 to .87 (Cabrera et al., 2018; Cerritelli et al., 2021). As abovementioned, our analysis argued against scoring these subscales separately, which may ultimately explain why internal consistency for the full BPQ Autonomic Reactivity scale in the current study was larger.

Test-retest reliability was good for the IAS and the BPQ Autonomic Reactivity scales, although it was significantly lower for the BPQ Body Awareness scale. Moreover, test-retest agreement was not very high on any of the scales. Cabrera et al. (2018) reported excellent ICC values (ranging from .96 to .99) across BPQ scales, but their analysis relied on the binary scoring system, which likely inflated test-retest reliability. Conversely, results from Murphy et al. (2020) suggest a more modest reliability for the BPQ Body Awareness using the full scoring system, $r = .684$. These authors also reported better test-retest reliability for the IAS, $r = .754$, with results quite similar to the current findings. These modest results regarding test-retest reliability and agreement may be explained by the previously described heterogeneity in participants' comprehension of the underlying constructs being measured by each scale, particularly regarding the BPQ Body Awareness scale. If there is a clear between-subjects lack of consistency when interpreting the instructions of the BPQ Body Awareness scale (Gabriele et al., 2020), it is also reasonable to postulate that there is a within-subject variability in the interpretation of this scale across timepoints, which ultimately reduces test-retest reliability and agreement (e.g., a subset of participants may initially consider items as measuring how often the signal objectively occurs, but then shift to an interoceptive attention interpretation in the retest phase). This challenge may also interfere with test-retest indicators for the IAS, but to a much lesser extent, as most participants (84.6%) truly consider that this scale is measuring interoceptive accuracy (Gabriele et al., 2020).

A Quadratic Association Between Self-Reported Interoceptive Attention and Accuracy?

Dissociating self-reported interoceptive attention and accuracy, as indexed by the BPQ Body Awareness and the IAS respectively, would allow not only to establish convergent and divergent validity for these scales but also to provide further evidence for the 2 x 2 factorial model of interoception proposed by Murphy et al. (2019). Here, the attained results were somewhat inconsistent, although still providing evidence for considering self-reported interoceptive attention and accuracy as separate constructs.

Contrary to our hypothesis, the BPQ Body Awareness and the IAS were positively and significantly correlated, $r = .204$, which could suggest that these constructs overlap, at least partially. Results from several samples mainly suggest that the BPQ Body Awareness is not significantly correlated with the IAS, $r = .022$ to $.091$, although there has also been evidence for a significant positive correlation between these scales, $r = .153$ (Gabriele et al., 2020; Murphy et al., 2020). More importantly, the association between these measures appears to be widely contingent on the interpretation of the BPQ Body Awareness instructions, which may be quite inconsistent as previously described. Gabriele et al. (2020) found that when participants consider the BPQ Body Awareness as assessing attention, there is no relationship between this scale and IAS. Conversely, when the BPQ Body Awareness was interpreted as a measure of interoceptive accuracy, there was a significant positive association with the IAS score. Therefore, the positive correlation between the BPQ Body Awareness and the IAS found in the current Portuguese sample may have been driven by participants who interpreted both measures as pertaining to interoceptive accuracy and not necessarily by a true association between self-reported interoceptive attention and accuracy. It is also possible that some participants considered that both scales assessed interoceptive attention, although this is unlikely according to previous results from Gabriele et al. (2020).

An additional critical finding from this current study was the quadratic U-shaped association between the BPQ Body Awareness and the IAS, confirming the hypothesis stemming from previous exploratory analyses (Campos et al., 2020). This result was fairly robust due to several reasons: (1) adding the quadratic term in the polynomial regression model explained an additional 12.7% of the variance, much more than the initial linear term (4.2%); (2) two different procedures were used to compare linear, quadratic,

and cubic models, both providing similar results: the quadratic model was more adequate than the linear model, but the cubic model was not superior to the quadratic model; (3) two-lines testing was also implemented, as quadratic models have been linked to a high rate of false positives, providing specific evidence for a quadratic U-shaped association; (4) all these results were consistently found after excluding careless responders, outliers, and other confounders. Hence, in the current sample, it was quite clear that both high and low scores in the BPQ Body Awareness were associated with enhanced IAS scores. This might explain the significant linear correlation between both measures, as the positive slope of the U-shaped association (higher than the negative slope - see Figure 2) may have led to a positive correlation coefficient between the scores.

Importantly, this quadratic U-shaped association is extremely intriguing to the field and triggers a wide range of important topics. First, this non-linear association suggests that self-reported interoceptive attention and accuracy are indeed separate constructs, although likely influenced by each other. This is in line with recent evidence that found that interoceptive attention and accuracy are associated with dissociable insular subdivisions (bilateral dorsal mid-anterior insula vs. dorsal anterior portion of the right insula, respectively) which are vastly interconnected (Haruki & Ogawa, 2021). Secondly, it is important to further refine what is driving this association, that is, whether it may be a true non-linear association between self-reported interoceptive attention and accuracy or if these results are due to differences across both questionnaires (e.g., each measure assesses somewhat different interoceptive signals). For instance, Gabriele et al. (2020) assessed the association between several measures of interoceptive attention and accuracy, suggesting that the quadratic model was statistically significant across most analyses, although in some models (including models using the BPQ to index interoceptive attention) adding the cubic term also improved model fit, making their findings hard to interpret. Interestingly, in the models examining the association between the IAS and the newly developed Interoceptive Attention Scale, where the interoceptive signals are exactly matched between both scales, only the quadratic term produced significant changes to model fit, while the cubic term did not. Thus, although it seems apparent that each specific scale may produce different patterns of association, it is still important to further examine whether there is a true

quadratic U-shaped association between interoceptive attention and accuracy in future studies. It would also be extremely important to examine whether there is a similar quadratic association when using performance-based tasks for interoceptive attention and accuracy instead of self-report measures. Finally, another enthralling issue is why both high and low levels of self-reported interoceptive attention may be associated with higher self-reported accuracy. It is reasonable to argue why high levels of interoceptive attention are associated with enhanced interoceptive accuracy, as subjects who consistently allocate their attentional resources to inner signals may develop (over time) an increased ability to accurately describe interoceptive inputs. In contrast, it is somewhat counterintuitive to suggest that individuals who seldom pay attention to interoceptive signals are highly accurate to describe this information when required to do so.

Other Factors Dissociating Self-Reported Interoceptive Attention and Accuracy

Additionally, convincing evidence of the dissociation between interoceptive attention and accuracy stemmed from the correlational analysis of these constructs with alexithymia. As predicted, alexithymia was differentially associated with interoceptive attention and accuracy. More specifically, the IAS was negatively correlated with alexithymia, $r = -.291$, while there was no significant association between BPQ Body Awareness and alexithymia. A positive association between BPQ Body Awareness and alexithymia was expected according to previous preliminary evidence from a meta-analysis (Trevisan et al., 2019), but recent studies also indicated that these variables may be uncorrelated (Murphy et al., 2020). Regardless, the specific association of alexithymia with the IAS and not with the BPQ Body Awareness provides additional evidence for the dissociation between self-reported interoceptive attention and accuracy, further supporting the 2 x 2 factorial model proposed by Murphy et al. (2019).

Finally, analyses exploring the association of demographic variables with self-reported interoceptive attention and accuracy also offer additional insights regarding the dissociation of these constructs. Regarding age, there was a significant negative relationship with the BPQ Body Awareness in the current sample, but there was no association with self-reported interoceptive accuracy. This suggests an age-related decline in self-reported interoceptive attention, whilst interoceptive accuracy remains stable across the lifespan. According to previous findings, it was hypothesized that both interoceptive attention and accuracy

would be somewhat reduced in aging participants. Both the BPQ Body Awareness and the IAS have been negatively associated with age, while objective performance on cardiac interoceptive accuracy tasks also seems to be diminished in older subjects (Cabrera et al., 2018; Khalsa et al., 2009; Murphy et al., 2020; Murphy, Geary, et al., 2018). However, the results presented here suggest that beliefs regarding self-reported interoceptive attention and accuracy are differentially influenced by aging, which further reinforces the argument for the usefulness of the 2 x 2 factorial model. For instance, it is possible that interoceptive accuracy objectively deteriorates with aging, but older participants overestimate the beliefs in their abilities within this domain. This would be in line with previous evidence suggesting that elderly subjects underestimate age-related decline in some exteroceptive domains (Cavazzana et al., 2018), at least assuming that such effect is also likely to occur regarding interoceptive signals.

Regarding sex, it was postulated that self-reported interoceptive attention should be higher in female participants, while interoceptive accuracy would be larger in male subjects, according to previous evidence regarding sex differences in interoception (Murphy, Viding, et al., 2019). Effectively, there were sex differences regarding the BPQ Body Awareness, as women displayed larger self-reported interoceptive attention in comparison to men, although the effect size was small. However, there were no significant sex differences regarding the IAS. It is important to highlight that sex differences in interoception are fairly inconsistent across studies and may ultimately require a clearer discussion within the 2 x 2 framework. While there have been reports of enhanced objective interoceptive accuracy abilities in men (Grabauskaitė et al., 2017; Murphy, Brewer, et al., 2018), others have also reported no differences between men and women in these tasks (Khalsa et al., 2009). There is also evidence suggesting that sex is not significantly associated with neither BPQ Body Awareness and IAS scores (Murphy 2018, 2020), although Cabrera et al. (2018) and (Cerritelli et al., 2021) described higher BPQ Body Awareness scores in female participants when using the binary scoring system. Current findings further suggest differential sex-related differences in the self-reported interoceptive attention and accuracy, which provides further evidence for these constructs as separate measurable constructs. Importantly, sex differences in interoception should be more

clearly explored within the 2 x 2 factorial framework, as they may ultimately be related to sex differences in psychopathology (Murphy, Viding, et al., 2019).

Limitations and Recommendations for Future Research

This study comprised a comprehensive effort to validate and assess the psychometric proprieties of the BPQ and the IAS. However, a few limitations should be highlighted in the presented analysis, especially regarding convergent and divergent validity. First, there was no formal test for convergent validity using questionnaires that presumably assess either interoceptive attention or accuracy, such as the recently developed Interoceptive Attention Scale (Gabriele et al., 2020) or the Interoceptive Confusion Questionnaire (Brewer et al., 2016), as neither of these is available in Portuguese. There is actually an interoception-related questionnaire translated for Portuguese (Multidimensional Assessment of Interoceptive Awareness; (Machorrinho et al., 2019) but it does not provide a clear measure of either interoceptive attention and accuracy. Thus, it was decided to establish convergent and divergent validity by examining the association between both scales and by exploring the theoretically sound differential association between both scales and the well-established alexithymia scale, which has been properly validated for European Portuguese.

Secondly, it is important to question to what extent the BPQ Body Awareness scale is effectively measuring interoceptive attention, at least in most subjects. Gabriele et al. (2020) highlights that *“the BPQ terminology requires participants to report their level of ‘awareness’ of internal bodily signals rather than explicitly assessing the extent to which internal signals are the object of their attention”*. Hence, individual differences in BPQ Body Awareness may be widely contingent on the interpretation of the term “awareness”, which does not even have a direct semantic equivalent in Portuguese. Thus, future studies (particularly validations for new languages) using BPQ Body Awareness to assess self-reported interoceptive attention (as well as other interoception-related scales) should use the same strategy as Gabriele et al. (2020) and provide an additional control question where participants may describe whether they interpreted each questionnaire as pertaining to attention, accuracy, or even other interpretations (e.g., frequency of sensations). Also, the instructions for the BPQ Body Awareness are quite broad, as this scale

was not originally created to measure interoceptive attention as conceptualized nowadays. Thus, it seems reasonable that authors provide additional information to participants (e.g., define interoceptive attention) when using the BPQ Body Awareness as a self-report measure of interoceptive attention.

The findings presented here also allow for some suggestions regarding future studies using the BPQ. Researchers should clearly understand that BPQ Body Awareness and Autonomic Reactivity are measuring different constructs, as each scale was respectively designed to assess either information regarding afferent projections or efferent control systems of the autonomic nervous system. Secondly, we argue against the use of the binary scoring system of the BPQ. Although a similar factor structure was possible for the BPQ Autonomic Reactivity scale using binary scoring, that was not the case for BPQ Body Awareness. Importantly, this work also provided clear evidence of unidimensionality for both the BPQ Body Awareness and Autonomic Reactivity using the full scoring. Thus, there is no reason not to use the full scoring as it provides a much more valid alternative to account for interindividual differences.

Conclusion

This study allowed for the translation and validation of the European Portuguese versions of the BPQ and the IAS. Data-driven strategies, modern factor analysis techniques, dimensionality analysis, and other psychometric properties indicated that these newly proposed versions are reliable and unidimensional. Correlational analysis confirmed that the BPQ Body Awareness and the IAS are differentially associated with alexithymia, indicating that these scales were effectively measuring different constructs. Importantly, the current study found a quadratic U-shaped association between the BPQ Body Awareness and the IAS, which can suggest a non-linear association between self-reported interoceptive attention and accuracy. These results, together with the differential associations with sex and age, further provide evidence for self-reported interoceptive attention and accuracy as proposed by the 2 x 2 factor model of interoception. Future studies are still required to further examine the non-linear nature of the association between these constructs and how self-reported questionnaires are related with objective measurements of interoceptive attention and accuracy.

Figures

Figure 1. The 2 x 2 factorial model of interoceptive abilities proposed by Murphy et al. (2019)

Figure 2. Two-lines testing: quadratic U-shaped association between BPQ Body Awareness and IAS.

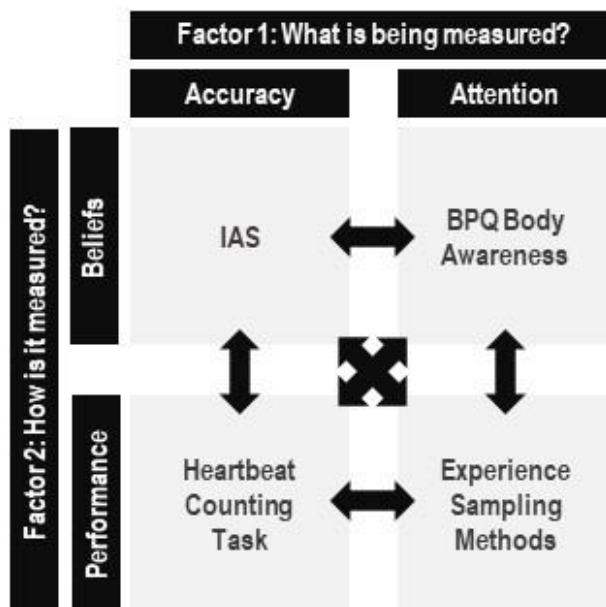
Figure 1

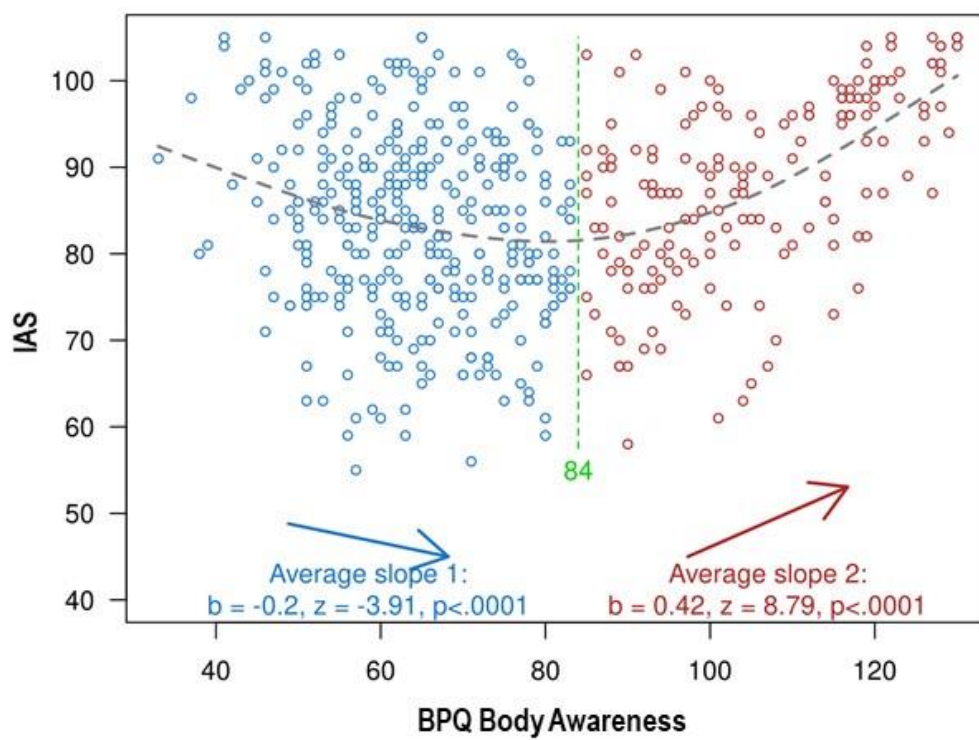
Figure 2

Table 1

Criteria for interpreting goodness of fit statistics, information criteria and ancillary bifactor measures.

<i>Fit Statistics</i>
<p>Weighted Root Mean Square Residual (WRMR): Absolute fit index suitable for categorical data estimated using WLSMV. A value ≤ 1 indicates that the model fits the data well (Gana & Broc, 2019).</p> <p>Root Mean Square Error of Approximation (RMSEA): Values $\leq .06$ and confidence intervals between 0 and .10 indicate a good fit (Hu & Bentler, 1999), while values between .06 and .08 still indicate an acceptable fit (Gana & Broc, 2019).</p> <p>Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI): Values $\geq .95$ indicate good fit (Hu & Bentler, 1999), while values between .90 and .95 indicate an acceptable fit (Gana & Broc, 2019).</p>
<i>Information Criteria</i>
<p>Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC): Parsimony fit indices to compare non-nested models (lower values indicate better fit). AIC and BIC were computed for least squares procedures as suggested by Yamaoka et al. (1978)</p>
<i>Ancillary bifactor measures to assess dimensionality</i>
<p>Percent Uncontaminated Correlations (PUC): Indexes the number of unique correlations in a correlation matrix that are influenced by a single factor divided by the total number of unique correlations.</p> <p>Explained Common Variance of the General Factor (ECV_{Gen}): Indexes the percent of common variance explained by the general factor. When ECV is ≥ 0.70 and PUC ≥ 0.70, common variance is considered mainly unidimensional (Rodriguez et al., 2016a).</p> <p>Individual Explained Common Variance (ECV_{Ind}): Indexes the amount of common variance attributable to the general factor. Values $\geq .50$ indicate that an item reflects the general factor more than the specific factors. Values $\geq .80$ indicate a strongly unidimensional item (Stucky & Edelen, Maria Orlando, 2015).</p>

Average Relative Parameter Bias (ARPB): Compares items loadings on the general factor to loadings on the one-factor solution. A 10-15% upper limit for ARPB was recommended by Muthén et al. (1987).

Ancillary bifactor measures for model-based reliability

Coefficient Omega Hierarchical (ωH): indexes how much variance of the total score is attributable to the general factor. Values $\geq .75$ indicate good reliability, although the minimum acceptable threshold is $\geq .50$ (Reise et al., 2013).

Coefficient Omega Hierarchical Subscales (ωHS): indexes how much variance of the subscale score is attributable to a specific factor, after controlling for its variance due to the general factor. Same cutoffs values as ωH .

Table 2*Standardized factor loadings for BPQ Body Awareness models.*

		One-Factor CFA	One-Factor CFA Without BPQ1
BPQ1	Swallowing frequently	0.461*	-
BPQ2	An urge to cough to clear my throat	0.697*	0.691*
BPQ3	My mouth being dry	0.789*	0.786*
BPQ4	How fast I am breathing	0.795*	0.795*
BPQ5	Watering or tearing of my eyes	0.813*	0.812*
BPQ6	Noises associated with my digestion	0.622*	0.621*
BPQ7	A swelling of my body or parts of my body	0.733*	0.733*
BPQ8	An urge to defecate	0.693*	0.693*
BPQ9	Muscle tension in my arms and legs	0.707*	0.708*
BPQ10	A bloated feeling because of water retention	0.646*	0.648*
BPQ11	Muscle tension in my face	0.678*	0.679*
BPQ12	Goose bumps	0.826*	0.827*
BPQ13	Stomach and gut pains	0.841*	0.844*
BPQ14	Stomach distension or bloatedness	0.722*	0.723*
BPQ15	Palms sweating	0.766*	0.765*
BPQ16	Sweat on my forehead	0.788*	0.789*
BPQ17	Tremor in my lips	0.883*	0.884*
BPQ18	Sweat in my armpits	0.699*	0.701*
BPQ19	The temperature of my face (especially my ears)	0.741*	0.741*
BPQ20	Grinding my teeth	0.626*	0.625*
BPQ21	General jitteriness	0.741*	0.743*
BPQ22	The hair on the back of my neck "standing up"	0.828*	0.828*
BPQ23	Difficulty in focusing	0.661*	0.662*
BPQ24	An urge to swallow	0.744*	0.737*
BPQ25	How hard my heart is beating	0.788*	0.789*
BPQ26	Feeling constipated	0.714*	0.715*

* $p < .001$. Factor loadings $>.400$ are bolded.

Table 3*Goodness of fit statistics and information criteria for BPQ Autonomic Reactivity.*

	χ^2	<i>df</i>	WRMR	RMSEA CI _{90%}	CFI	TLI	AIC	BIC
One-Factor CFA	1501.53	170	2.18	.123 [.118, .129]	.839	.820	308.87	729.04
Correlated 3-Factor CFA	775.44	169	1.53	.083 [.078, .089]	.927	.917	-51.88	372.53
Correlated 3-Factor ESEM	713.47	151	1.32	.085 [.079, .091]	.932	.914	-172.40	328.41
Bi-Factor CFA	382.25	149	0.95	.055 [.048, .062]	.972	.964	-500.15	9.15
Bi-Factor ESEM	357.63	133	0.85	.057 [.050, .064]	.973	.961	-590.90	-13.69
Cabrera <i>et al.</i> (2018) 2-Factor CFA	720.80	168	1.46	.080 [.074, .086]	.933	.924	-100.02	328.64

WRMR: Weighted Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; AIC: Akaike Information Criterion; Bayesian Information Criterion.

Table 4*Standardized factor loadings and individual item ancillary measures for BPQ Autonomic Reactivity models.*

			CFA 2-Factor Model		ESEM 2-Factor Model		Cabrera et al. (2018) 2-Factor CFA		CFA Bi-Factor Model			ESEM Bi-Factor Model				
		One-Factor CFA	F1	F2	F1	F2	F1	F2	GF	F1	F2	ECV _{Ind}	GF	F1	F2	ECV _{Ind}
BPQ27	I have difficulty coordinating breathing and eating.	.706*	.729*		.824*	-.126	.729*		.725*	.289		.863	.762*	-.042	-.225*	.917
BPQ28	When I am eating, I have difficulty talking.	.606*	.630*		.678*	-.057	.630*		.628*	.225		.886	.657*	-.029	-.151	.948
BPQ29	My heart often beats irregularly.	.612*	.634*		.506*	.182*	.634*		.633*	-.145		.950	.581*	.250*	.073	.833
BPQ30	When I eat, food feels dry and sticks to my mouth and throat.	.634*	.661*		.591*	.107	.662*		.665*	.133		.962	.672*	.004	-.015	.999
BPQ31	I feel shortness of breath.	.724*	.753*		.765*	-.002	.754*		.708*	-.438*		.723	.602*	.606*	-.044	.495
BPQ32	I have difficulty coordinating breathing with talking.	.670*	.702*		.728*	-.023	.702*		.706*	-.016		.999	.685*	.220*	-.127	.879
BPQ33	When I eat, I have difficulty coordinating swallowing, chewing, and/or sucking with breathing.	.753*	.777*		.862*	-.113	.776*		.775*	.272		.890	.806*	-.015	-.223	.929
BPQ34	I have a persistent cough that interferes with my talking and eating.	.577*	.600*		.500*	.144	.600*		.599*	.191		.908	.620*	-.098	.025	.974
BPQ35	I gag from the saliva in my mouth.	.500*	.520*		.393*	.178	.520*		.524*	.140		.933	.536*	-.088	.063	.961
BPQ36	I have chest pains.	.614*	.643*		.552*	.132	.643*		.637*	-.305*		.814	.556*	.412*	.051	.642
BPQ37	I gag when I eat.	.660*	.697*		.626*	.107	.697*		.707*	.300*		.847	.750*	-.172*	-.040	.947
BPQ38	When I talk, I often feel I should cough or swallow the saliva in my mouth.	.605*	.633*		.576*	.086	.633*		.635*	.066		.989	.633*	.056	-.028	.990
BPQ39	When I breathe, I feel like I cannot get enough oxygen.	.768*	.797*		.755*	.071	.797*		.769*	-.463*		.734	.652*	.602*	.016	.540
BPQ40	I have difficulty controlling my eyes.	.630*	.654*		.592*	.094	.655*		.657*	.003		1.000	.638*	.137	-.020	.955
BPQ41	I feel like vomiting.	.631*	.723*	.340*	.411*	.326*	.398*	.594*		.285*	.813	.603*	.071	.258*	.835	
BPQ42	I have 'sour' stomach.	.626*	.711*	.144	.609*		.727*	.516*	.466*	.551	.546*	.001	.436*	.611		
BPQ43	I am constipated.	.425*	.492*	.117	.404*		.505*	.361*		.312*	.572	.384*	.003	.284*	.646	
BPQ44	I have indigestion.	.794*	.854*	-.028	.901*		.863*	.512*		.721*	.335	.568*	-.033	.683*	.408	
BPQ45	After eating I have digestive problems.	.841*	.919*	.006	.940*		.931*	.567*	.764*	.355	.619*	.012	.712*	.430		
BPQ46	I have diarrhea.	.480*	.547*	.187*	.389*		.559*	.424*	.295*	.674	.442*	.025	.263*	.737		

* $p < .001$. Factor loadings $> .400$ are bolded. ECV_{Ind} $> .500$ are bolded. GF: General Factor.

Table 5*Goodness of fit statistics and information criteria for IAS.*

	χ^2	<i>df</i>	WRM R	RMSEA CI _{90%}	CFI	TLI	AIC	BIC
One-Factor CFA	1837.75	189	2.32	.130 [.125, .136]	.864	.848	424.95	862.10
Correlated 3- Factor CFA	941.29	186	1.61	.089 [.083, .094]	.938	.929	54.42	504.30
Correlated 3- Factor ESEM	665.51	150	1.12	.082 [.075, .088]	.957	.940	-249.16	353.52
Bi-Factor ESEM	459.03	132	0.86	.069 [.063, .076]	.973	.957	-483.26	195.80
Murphy <i>et al.</i> (2020) 2-Factor CFA	1514.44	188	2.11	.117 [.112, .123]	.890	.877	325.44	766.84

WRMR: Weighted Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; AIC: Akaike Information Criterion; Bayesian Information Criterion. *Note.* The hierarchical ESEM model is not presented as it displayed similar fit statistics to the correlated 3-factor ESEM model.

Table 6*Standardized factor loadings and individual item ancillary measures for IAS models.*

		Murphy et al. (2020)													
		One-Factor	Correlated 3-Factor CFA			Correlated 3-Factor ESEM			2-Factor CFA		Bi-Factor ESEM				
		CFA	F1	F2	F3	F1	F2	F3	F1	F2	GF	F1	F2	F3	ECV _{Ind}
IAS1	Heart	.545*	.588*			.460*	.009	.223*	.569*		.578*	-.093	.026	.005	.973
IAS2	Hungry	.636*	.684*			.761*	-.001	-.081	.662*		.664*	-.143	-.279*	-.119	.797
IAS3	Breathing	.599*	.641*			.648*	-.106	.192	.621*		.652*	-.272*	-.006	-.129	.824
IAS4	Thirsty	.582*	.627*			.835*	-.187	-.006	.606*		.621*	-.189	-.296*	.072	.750
IAS5	Urinate	.694*	.745*			.829*	.014	-.111	.719*		.713*	-.064	-.379*	.034	.773
IAS6	Defecate	.705*	.758*			.705*	.130	-.070	.732*		.724*	-.053	-.259*	-.118	.862
IAS7	Taste	.579*		.616*		.278*	.284*	.138*	.600*		.574*	.158*	-.019	.075	.914
IAS8	Vomit	.661*		.703*		.122	.632*	-.022	.681*		.617*	.323*	-.045	-.131	.755
IAS9	Sneeze	.801*		.842*		.025	.866*	-.031		.842*	.707*	.495*	-.055	-.037	.667
IAS10	Cough	.790*		.833*		-.021	.877*	.007		.829*	.691*	.594*	-.040	.072	.570
IAS11	Temperature	.734*		.779*		.408*	.415*	.006	.763*		.718*	.228*	-.167*	.047	.863
IAS12	Sexual Arousal	.674*	.719*			.424*	.314*	.033	.698*		.683*	-.011	-.035	-.289*	.846
IAS13	Wind	.714*		.757*		.017	.722*	.077		.754*	.665*	.273	.136	-.376*	.654
IAS14	Burp	.743*		.787*		.006	.736*	.115		.779*	.685*	.310*	.133	-.254*	.725
IAS15	Muscles	.686*	.731*			.429*	.171	.256*	.706*		.696*	.083	.007	.147*	.944
IAS16	Bruise	.520*			.632*	.280*	-.014	.446*		.550*	.515*	.110	.145	.435*	.544
IAS17	Pain	.683*	.735*			.525*	.134	.171*	.709*		.702*	.066	-.095	.146*	.934
IAS18	Blood Sugar	.452*			.556*	.337*	-.185	.492*		.476*	.460*	-.029	.167	.426*	.502
IAS19	Affective Touch	.591*			.709*	.359*	.029	.390*	.612*		.604*	-.036	.155	.121	.901
IAS20	Tickle	.699*			.817*	-.042	.172	.794*		.737*	.616*	-.047	.658*	-.028	.465
IAS21	Itch	.713*			.841*	.035	.096	.814*		.751*	.628*	-.047	.578*	.083	.535

* $p < .001$. Factor loadings $> .400$ are bolded. $ECV_{Ind} > .500$ are bolded. GF: General Factor. *Note.* The hierarchical ESEM model is not presented as standardized

loadings were practically similar to the correlated 3-factor ESEM model.

Table 7*Psychometric indicators: internal consistency, test-retest reliability, and test-retest agreement*

	BPQ Body Awareness	BPQ Autonomic Reactivity	IAS
ω_{categorical}	.978	.944	.970
99% CIs	[.972, .981]	[.922, .960]	[.951, .982]
Cronbach's Alpha	.960	.900	.903
99% CIs	[.954, .967]	[.875, .922]	[.887, .917]
Test-Retest ICC	.684	.832	.756
99% CIs	[.537, .789]	[.672, .908]	[.642, .837]
Test-Retest Pearson's <i>r</i>	.696	.856	.757
99% CIs	[.561, .795]	[.782, .905]	[.644, .838]
Test-Retest CCC	.682	.831	.754
99% CIs	[.546, .783]	[.750, .887]	[.641, .835]
Spearman Mean-Diff.	.033	-.148	-.099
99% CIs	[-.189, .253]	[-.359, .077]	[-.314, .125]

ICC: Intraclass Correlation Coefficient; CCC: Concordance Correlation Coefficient; Spearman Mean-Diff: Spearman rank correlation coefficients between test-retest mean scores and test-retest difference scores.

Table 8*Correlation matrix for continuous variables*

	BPQ Body Awareness	BPQ Autonomic Reactivity	IAS	TAS-20	Age
BPQ Body	1	.216	.204	-.030	-.157
Awareness		[.105, .322]	[.093, .310]	[-.143, .084]	[-.266, -.045]
BPQ Autonomic	< .001	1	-.267	.372	-.048
Reactivity			[-.369, -.158]	[.270, .466]	[-.160, .066]
IAS	< .001	< .001	1	-.291	.081
				[-.391, -.184]	[-.033, .192]
TAS-20	.500	< .001	< .001	1	-.097
					[-.208, .017]
Age	< .001	.281	< .001	.028	1

Endnotes

- ¹ BPQ Body Awareness is more frequently used as an index of either interoceptive sensibility or interoceptive awareness, an inconsistency that is in itself problematic. More importantly, there are a few studies that compute a total body perception or interoception score using both body awareness and autonomic reactivity items, while others consider both subscales as measures of interoceptive awareness despite calculating separate scores for each domain. This heterogenous operationalization of the BPQ is quite problematic as it may lead to erroneous interpretations of the reported findings within the field.
- ² 1. The original authors clearly state the BPQ assesses two different domains and do not clearly describe any high-order construct that encompasses both of them; 2. The scale presents very distinct instructions for the Body Awareness and Autonomic Reactivity subscales, clearly addressing different constructs; 3. The original authors do not suggest any pooled score comprising both domains and only a few papers have actually computed a total score for analysis; 4. The only psychometric work with the BPQ conducted EFA separately for both subscales. 5. The scales are used independently more frequently than not.
- ³ Additional questionnaires not reported for the purpose of the current manuscript were also included in the online survey. More information in the preregistration (<https://osf.io/5jhcw>).
- ⁴ There were only a few changes to the statistical methods proposed in the preregistration: (1) For the CCC, the formula proposed by Lin (1989, 2000) was used instead of the one proposed by Chinchilli et al. (1996), as the latter was not suitable for the analysis presented here. Also, instead of the McBride (2005) criteria to assess the CCC, the more liberal threshold described by Landis & Koch (1977) was used, as the former was developed for assessing the degree of equivalence between laboratory methods (e.g., chemical concentrations) and it is too conservative for self-report questionnaires; (2) The 99% CIs for the ICC were computed instead of the 95% CIs described in the preregistration. (3) In the analysis excluding other possible confounders, participants with other chronic conditions or that regularly take medications were not excluded as originally planned because they were a very large and heterogeneous group (e.g., asthma, reports of birth control intake, etc) with no rationale to influence interoception-

related measures; (4) Slightly different versions of the statistical software described in the preregistration were used.

- ⁵ A total of 516 responses were collected but one subject who reported less than 18 years of age was excluded.
- ⁶ There were no major differences between the full sample analysis and the follow-up analysis excluding subjects that could constitute potential cofounders. There were only some differences in parallel analysis (number of extracted factors) and factor structure analyses (item loadings and model fit), although the final retained model for each scale was always the bifactor ESEM model. Also, in these control analyses, there was also clear evidence for unidimensionality on all the scales. Comprehensive reports of these follow-up analysis can be found at <https://osf.io/hmtz9/>
- ⁷ The factor structure retained by Cabrera et al. (2018) allowed item 41 (“I feel like vomiting”) to load on two factors. In the CFA models, each item was only attributed to the factor where it loaded the most. Thus, item 41 was only allowed to load on Factor 2.
- ⁸ Hierarchical models were not implemented as they require at least three first order-factors (Kline, 2016).

References

- Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397–438.
<https://doi.org/10.1080/10705510903008204>
- Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, 38(1), 23–32. [https://doi.org/10.1016/0022-3999\(94\)90005-1](https://doi.org/10.1016/0022-3999(94)90005-1)
- Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). The twenty-item Toronto Alexithymia scale—II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research*, 38(1), 33–40. [https://doi.org/10.1016/0022-3999\(94\)90006-X](https://doi.org/10.1016/0022-3999(94)90006-X)
- Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 87–101. <https://doi.org/10.1080/10705511.2014.934850>
- Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154.
<https://doi.org/10.1093/scan/nsw154>
- Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Sage Publications.
- Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations*, 9, 205979911667287. <https://doi.org/10.1177/2059799116672875>
- Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, 65(1), 23–28. <https://doi.org/10.1007/BF02294183>
- Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: A general deficit of interoception. *Royal Society Open Science*, 3(10), 150664. <https://doi.org/10.1098/rsos.150664>

- Cabrera, A., Kolacz, J., Pailhez, G., Bulbena-Cabre, A., Bulbena, A., & Porges, S. W. (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. <https://doi.org/10.1002/mpr.1596>
- Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web survey methodology*. SAGE.
- Campos, C., Murphy, J., Brewer, R., Gabriele, E., Rocha, N., & Barbosa, F. (2020). *Examining the Quadratic Association Between Self-Reported Interoceptive Accuracy and Attention*. <https://osf.io/ktpzc>
- Cavazzana, A., Röhrborn, A., Garthus-Niegel, S., Larsson, M., Hummel, T., & Croy, I. (2018). Sensory-specific impairment among older people. An investigation using both sensory thresholds and subjective measures across the five senses. *PLOS ONE*, 13(8), e0202969. <https://doi.org/10.1371/journal.pone.0202969>
- Cerritelli, F., Galli, M., Consorti, G., D'Alessandro, G., Kolacz, J., & Porges, S. W. (2021). Cross-cultural adaptation and psychometric properties of the Italian version of the Body Perception Questionnaire. *PLOS ONE*, 16(5), e0251838. <https://doi.org/10.1371/journal.pone.0251838>
- Ceunen, E., Van Diest, I., & Vlaeyen, J. W. S. (2013). Accuracy and awareness of perception: Related, yet distinct (commentary on Herbert et al., 2012). *Biological Psychology*, 92(2), 426–427. <https://doi.org/10.1016/j.biopsycho.2012.09.012>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8), 655–666. <https://doi.org/10.1038/nrn894>
- Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, 77(4), 624–638. <https://doi.org/10.1016/j.neuron.2013.02.008>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>

- Donner, A., & Zou, G. (2002). Testing the Equality of Dependent Intraclass Correlation Coefficients. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 51(3), 367–379.
- Dueber, D. M. (2017). *Bifactor Indices Calculator: A Microsoft Excel-Based Tool to Calculate Various Indices Relevant to Bifactor CFA Models*. <https://doi.org/10.13023/edp.tool.01>
- Fotopoulou, A., & Tsakiris, M. (2017). *Mentalizing homeostasis: The social origins of interoceptive inference – replies to Commentaries*.
<https://www.tandfonline.com/doi/full/10.1080/15294145.2017.1307667>
- Gabriele, E., Spooner, R., Brewer, R., & Murphy, J. (2020). *Dissociations between interoceptive accuracy and attention: Evidence from the interoceptive attention scale* [Preprint]. PsyArXiv.
<https://doi.org/10.31234/osf.io/vjgh6>
- Gana, K., & Broc, G. (2019). *Structural equation modeling with lavaan*. ISTE Ltd ; John Wiley & Sons, Inc.
- Garcia-Granero, M. (2005). *Lin's concordance correlation coefficient* [SPSS Macro]. <http://gjyp.nl/marta/>
- Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: New insights link internal physiology to social behaviour. Commentary on: *Social Cognitive and Affective Neuroscience*, 8(3), 231–234. <https://doi.org/10.1093/scan/nss140>
- Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, 104, 65–74. <https://doi.org/10.1016/j.biopsycho.2014.11.004>
- Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology*, 114, 117–126. <https://doi.org/10.1016/j.biopsycho.2015.12.003>
- Garrido, L. E. (2017). *Horn's parallel analysis method with polychoric correlations*. MATLAB Central File Exchange. <https://www.mathworks.com/matlabcentral/fileexchange/64213-horn-s-parallel-analysis-method-with-polychoric-correlations>

- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18*(4), 454–474. <https://doi.org/10.1037/a0030005>
- Garson, G. D. (2012). *Curve Fitting and Nonlinear Regression*. Statistical Associates Publishers.
- Gignac, G. E. (2019). *How2statsbook (Online Edition 1)*. Author.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Grabauskaitė, A., Baranauskas, M., & Griškova-Bulanova, I. (2017). Interoception and gender: What aspects should we pay attention to? *Consciousness and Cognition, 48*, 129–137. <https://doi.org/10.1016/j.concog.2016.11.002>
- Haruki, Y., & Ogawa, K. (2021). Role of anatomical insular subdivisions in interoception: Interoceptive attention and accuracy have dissociable substrates. *European Journal of Neuroscience, 53*(8), 2669–2680. <https://doi.org/10.1111/ejn.15157>
- Herman, A. M., Rae, C. L., Critchley, H. D., & Duka, T. (2019). Interoceptive accuracy predicts nonplanning trait impulsivity. *Psychophysiology, 56*(6), e13339. <https://doi.org/10.1111/psyp.13339>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- IBM Corp. Released. (2020). *IBM SPSS Statistics for Windows* (Version 27.0) [Computer software]. IBM Corp.
- Kelley, K. (2007). Methods for the Behavioral, Educational, and Social Sciences: An R package. *Behavior Research Methods, 39*(4), 979–984. <https://doi.org/10.3758/BF03192993>
- Kelley, K. (2020). *MBESS* (version 4.8.0) [Computer software and manual]. <http://cran.r-project.org>.
- Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*(1), 69–92. <https://doi.org/10.1037/a0040086>

- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., ... Zucker, N. (2018). Interoception and Mental Health: A Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(6), 501–513.
<https://doi.org/10.1016/j.bpsc.2017.12.004>
- Khalsa, S. S., Rudrauf, D., & Tranel, D. (2009). Interoceptive awareness declines with age. *Psychophysiology*, 46(6), 1130–1136. <https://doi.org/10.1111/j.1469-8986.2009.00859.x>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth edition). The Guilford Press.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Lee, I. A., & Preacher, K. J. (2013a). *Calculation for the test of the difference between two dependent correlations with no variable in common* [Computer software]. <http://quantpsy.org>
- Lee, I. A., & Preacher, K. J. (2013b). *Calculation for the test of the difference between two dependent correlations with one variable in common* [Computer software]. <http://quantpsy.org>
- Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1), 255. <https://doi.org/10.2307/2532051>
- Lin, L. I.-K. (2000). A note on the concordance correlation coefficient. *Biometrics*, 56(1), 324–325.
<https://doi.org/10.1111/j.0006-341X.2000.00324.x>
- Machorrinho, J., Veiga, G., Fernandes, J., Mehling, W., & Marmeleira, J. (2019). Multidimensional Assessment of Interoceptive Awareness: Psychometric Properties of the Portuguese Version. *Perceptual and Motor Skills*, 126(1), 87–105. <https://doi.org/10.1177/0031512518813231>

- Marshall, A. C., Gentsch, A., Jelinčić, V., & Schütz-Bosbach, S. (2017). Exteroceptive expectations modulate interoceptive processing: Repetition-suppression effects for visual and heartbeat evoked potentials. *Scientific Reports*, 7(1), 16525. <https://doi.org/10.1038/s41598-017-16595-9>
- Marshall, A. C., Gentsch, A., & Schütz-Bosbach, S. (2018). The Interaction between Interoceptive and Action States within a Framework of Predictive Coding. *Frontiers in Psychology*, 9, 180. <https://doi.org/10.3389/fpsyg.2018.00180>
- MATLAB (Version 2020a). (2020). [Computer software]. The MathWorks Inc.
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 116–139. <https://doi.org/10.1080/10705511.2014.961800>
- Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience*, 23, 45–56. <https://doi.org/10.1016/j.dcn.2016.12.006>
- Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception? Further evidence, the importance of control variables, and the problems with the Heartbeat Counting Task. *Biological Psychology*, 136, 189–197. <https://doi.org/10.1016/j.biopsycho.2018.05.010>
- Murphy, J., Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2020). Testing the independence of self-reported interoceptive accuracy and attention. *Quarterly Journal of Experimental Psychology*, 73(1), 115–133. <https://doi.org/10.1177/1747021819879826>
- Murphy, J., Catmur, C., & Bird, G. (2019). Classifying individual differences in interoception: Implications for the measurement of interoceptive awareness. *Psychonomic Bulletin & Review*, 26(5), 1467–1471. <https://doi.org/10.3758/s13423-019-01632-7>

- Murphy, J., Geary, H., Millgate, E., Catmur, C., & Bird, G. (2018). Direct and indirect effects of age on interoceptive accuracy and awareness across the adult lifespan. *Psychonomic Bulletin & Review*, 25(3), 1193–1202. <https://doi.org/10.3758/s13423-017-1339-z>
- Murphy, J., Viding, E., & Bird, G. (2019). Does atypical interoception following physical change contribute to sex differences in mental illness? *Psychological Review*, 126(5), 787–789. <https://doi.org/10.1037/rev0000158>
- Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22(1), 43–65. [https://doi.org/10.1016/0304-4076\(83\)90093-3](https://doi.org/10.1016/0304-4076(83)90093-3)
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431–462. <https://doi.org/10.1007/BF02294365>
- Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus User's Guide*. (Eighth Edition). Author.
- Ondobaka, S., Kilner, J., & Friston, K. (2017). The role of interoceptive inference in theory of mind. *Brain and Cognition*, 112, 64–68. <https://doi.org/10.1016/j.bandc.2015.08.002>
- Pereira, M. R., Barbosa, F., de Haan, M., & Ferreira-Santos, F. (2019). Understanding the development of face and emotion processing under a predictive processing framework. *Developmental Psychology*, 55(9), 1868–1881. <https://doi.org/10.1037/dev0000706>
- Petzschnner, F. H., Weber, L. A., Wellstein, K. V., Paolini, G., Do, C. T., & Stephan, K. E. (2019). Focus of attention modulates the heartbeat evoked potential. *NeuroImage*, 186, 595–606. <https://doi.org/10.1016/j.neuroimage.2018.11.037>
- Porges, S. W., Kolacz, J., & Holmes, L. (1993). *Body Perception Questionnaire (BPQ) Manual* (University of Maryland). Laboratory of Developmental Assessment.
- Prazeres, N., Parker, J., & Taylor, G. (2000). Adaptação Portuguesa da Escala de Alexitimia de Toronto de 20 Itens (TAS-20). *Revista Iberoamericana de Diagnóstico Y Evaluación Psicológica (RIDEP)*, 9, 9–21.

- Prazeres, N., Taylor, G. J., & Parker, J. D. A. (2008). Escala de Alexitimia de Toronto de Vinte Itens (TAS-20). In *Avaliação psicológica: Instrumentos validados para a população portuguesa: Vol. II* (2nd ed., pp. 95–107). Quarteto.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (version 4.0.3) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rae, C. L., Larsson, D. E. O., Garfinkel, S. N., & Critchley, H. D. (2019). Dimensions of interoception predict premonitory urges and tic severity in Tourette syndrome. *Psychiatry Research*, 271, 469–475. <https://doi.org/10.1016/j.psychres.2018.12.036>
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Salvato, G., Maio, G. D., & Bottini, G. (2019). Interoceptive sensibility tunes risk-taking behaviour when body-related stimuli come into play. *Scientific Reports*, 9(1), 1–5. <https://doi.org/10.1038/s41598-019-39061-0>
- Scarpazza, C., & di Pellegrino, G. (2018). Alexithymia, embodiment of emotions and interoceptive abilities. In *Current developments in alexithymia: A cognitive and affective deficit* (pp. 35–51). Nova Science Publishers.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007. <https://doi.org/10.1098/rstb.2016.0007>

- Simonsohn, U. (2018). Two Lines: A Valid Alternative to the Invalid Testing of U-Shaped Relationships With Quadratic Regressions. *Advances in Methods and Practices in Psychological Science*, 1(4), 538–555. <https://doi.org/10.1177/2515245918805755>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251. <https://doi.org/10.1037/0033-2909.87.2.245>
- Stucky, B. D., & Edelen, Maria Orlando. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki, *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment* (Routledge/Taylor&Francis Group, pp. 183–206).
- Trevisan, D. A., Altschuler, M. R., Bagdasaroy, A., Carlos, C., Duan, S., Hamo, E., Kala, S., McNair, M. L., Parker, T., Stahl, D., Winkelman, T., Zhou, M., & McPartland, J. C. (2019). A meta-analysis on the relationship between interoceptive awareness and alexithymia: Distinguishing interoceptive accuracy and sensibility. *Journal of Abnormal Psychology*, 128(8), 765–776. <https://doi.org/10.1037/abn0000454>
- von Mohr, M., Finotti, G., Ambroziak, K. B., & Tsakiris, M. (2020). Do you hear what I see? An audio-visual paradigm to assess emotional egocentricity bias. *Cognition and Emotion*, 34(4), 756–770. <https://doi.org/10.1080/02699931.2019.1683516>
- Wuensch, K. L. (2006). Curvilinear regression. In N. J. Salkind, *Encyclopedia of measurement and statistics* (pp. 211–215). Sage.
- Yamaoka, K., Nakagawa, T., & Uno, T. (1978). Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *Journal of Pharmacokinetics and Biopharmaceutics*, 6(2), 165–175. <https://doi.org/10.1007/BF01117450>