Mega-analysis of the Interoceptive Accuracy Scale (IAS) Structure and its Subjective Correlates

Ana Neves[1], Magdalena Pfaff[1], Robyn Scharte[1], Raquel Nogueira Arjona[1,2], Giulia Poerio[3], and Dominique Makowski[1,4]

[1]School of Psychology, University of Sussex

[2]Sussex Addictions Research and Intervention Centre, University of Sussex

[3]Sussex Centre for Consciousness Science, University of Sussex

Blabla the abstract blabla.

Keywords: keyword1, keyword2, keyword3

Introduction

Interoception refers to the process of sensing, interpreting and integrating information pertaining to internal organs, such as the heart, the lungs or the gut (Khalsa et al., 2018). While recent research emphasizes a key role of interoception in a variety of processes (e.g., emotion regulation, decision making) and of outcomes (physical and psychological well being), the field remains clouded by concerns about how interoception is assessed.

The Interoceptive Assessment Puzzle

Various measures of interoception have been developed (see Figure 1), forming a combination of "objective" and "subjective" assessments (i.e., physiological tasks such as the heart beat counting or tracking *vs.* questionnaires and subjective scales involving metacognitive judgments), "explicit" and "implicit" paradigms (i.e., directing participants' awareness and attention to interoceptive processes *vs.* measuring interoception unbeknownst to them), various interoceptive modalities (e.g., cardioception, respiroception, gastroception) and theoretical dimensions (e.g., accuracy, sensitivity, awareness). While there is no consensus as to which particular approach provides the most accurate and "pure" measure of interoception and interoceptive abilities (assuming it is a unidimensional construct), it is instead plausible that each measure has strengths and limitations, and a utility dependent on the context and goal at hand (Desmedt et al., 2023; Jahedi & Méndez, 2014).

Although the use of subjective self-report questionnaires to measure deeply embodied functions might seem paradoxical at first, recent redefinitions of interoception emphasize the role of high-level and metacognitive elaboration of interoceptive information. These redefinitions provide theoretical grounding to support the idea that some facets of interoception, including participants' metacognitive beliefs, can be assessed subjectively (Khalsa et al., 2018; Suksasilp & Garfinkel, 2022). Moreover, the notion that self-reports might not reflect the same processes as other interoception tasks might be important to contextualize the apparent lack of convergence between measures in the field (Desmedt et al., 2022). For instance, existing findings typically show weak or no correlations between questionnaires and objective measures, such as the Heartbeat Counting Task (HCT, Schandry, 1981) and the Heartbeat Detection Task (HDT, Kleckner et

---

Ana Neves
Magdalena Pfaff
Robyn Scharte
Raquel Nogueira Arjona
Giulia Poerio
Dominique Makowski

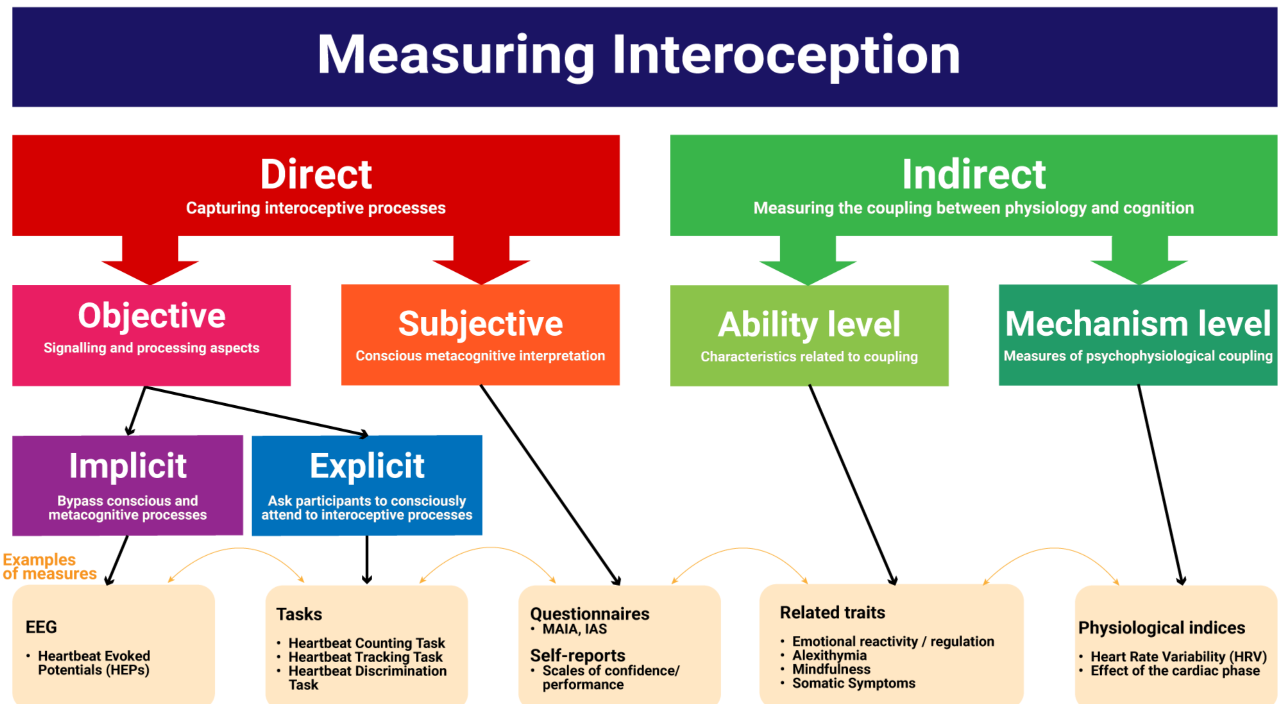
This preprint is a non-peer-reviewed work from the **Reality Bending Lab**.

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Ana Neves: Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing; Magdalena Pfaff: Data curation, Writing – original draft; Robyn Scharte: bla bla; Raquel Nogueira Arjona: bla bla; Giulia Poerio: bla bla; Dominique Makowski: Project administration, Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing

Correspondence concerning this article should be addressed to Dominique Makowski, Email: D.Makowski@sussex.ac.uk

Figure 1

The Interoceptive Assessment Puzzle. The different modalities of interoception (e.g., cardioception) can be assessed directly or indirectly. Direct assessments can further be subjective or objective, depending on whether they involve conscious metacognitive appraisals or more performance-based indices. Interoceptive tasks can be explicit (the participant is aware of the interoceptive nature of the task and must consciously attend to interoceptive signals; e.g., the heartbeat counting task) or implicit (measurements of interoception done unbeknowst to the participant; e.g., heartbeat evoked potentials measured during resting state). Indirect assessments evaluate constructs typically related (and ideally dependent on) to interoceptive processes or ability (or its deficit).



al., 2015), including for measures of the same theoretical dimensions (Arslanova et al., 2022; Brand et al., 2023; e.g., task-based accuracy vs. self-reported accuracy, Murphy et al., 2019). Additionally, even various objective measures assessing in theory the same interoceptive dimension, such as accuracy, either show no or weak correlation (respectively, Brand et al., 2023; Hickman et al., 2020). Perhaps more surprisingly, low correlations have been observed even among questionnaires, suggesting (in parallel to major validity concerns) the potential targeting of different facets related to interoception.

One striking example concerns the assessment of interoceptive sensibility, which is broadly defined as the self-reported tendency to focus on and detect internal sensations (Garfinkel et al., 2015), but more narrowly as the subjective tendency to focus on interoceptive signals, without necessarily implying detection ability (Khalsa et al., 2018). A recent systematic review suggested that various questionnaires designed to assess interoceptive sensibility may, in fact, measure distinct constructs, with the risk of researchers treating them as equivalent despite overall low convergence (Desmedt et al., 2022). Notably, this review adopted a broad definition of sensibility, incorporating both interoceptive sensibility and interoceptive self-report scales, following the eight-facet model by Khalsa et al. (2018). Several widely used questionnaires were included in the review, such as the Multidimensional Assessment of Interoceptive Awareness (MAIA, Mehling et al., 2012; MAIA-2, Mehling et al., 2018), the Body Perception Questionnaire (BPQ, Porges, 1993), the Private subscale of the Body Consciousness Questionnaire (PBCS, Miller et al., 1981), the Body Awareness Questionnaire (BAQ, Shields et al., 1989), and the Eating Disorder Inventory (Garner et al., 1983; EDI, Garner, 1991). The lack of correlations to moderate correlations among these questionnaires highlight the need for greater conceptual clarity regarding what each measure captures, how they relate to different dimensions of interoception, and their potential overlaps with other constructs, such as alexithymia and body awareness.

The Interoceptive Accuracy Scale (IAS)

Focusing on the *accuracy* dimension of interoception, a recently developed scale with a rapidly growing popularity is the Interoceptive Accuracy Scale (IAS, Murphy et al., 2019). The IAS consists of 21 Likert-scale items that query how accurately one can perceive different bodily signals, with one item per physiological modality such as respiration (*"I can always accurately perceive when I am breathing fast"*), heart (*"I can always accurately perceive when my heart is beating fast"*), skin (*"I can always accurately perceive when something is going to be ticklish"*), arousal or bodily functions like coughing (*"I can always accurately perceive when I am going to cough"*) or urinating (*"I can always accurately perceive when I need to urinate"*). Appealingly, the IAS' statements are about specific interoceptive behaviours and "objective" experiences (although still subjective reports of them), which is a distinct difference with other popular interoception questionnaires, such as the MAIA-2, which contains more general and metacognitive items (e.g., *"I trust my body sensations"*, *"I can notice an unpleasant body sensation without worrying about it"*), as well as dimensions related to attention regulation (e.g., Not-distracting) or emotion regulation (e.g., Not-worrying).

The original validation study suggested a two-factor structure for the IAS, one reflecting the perception of general interoceptive signals (urinate, hungry, defecate, thirsty, pain, heart, taste, breathing, temperature, muscles, affective touch, vomit, sexual arousal), and other relating to signals that may be difficult to perceive solely through interoceptive information (itch, tickle, cough, burp, bruise, blood sugar, sneeze, wind). The authors however underlined its acceptable but imperfect fit (Murphy et al., 2019, p. 127), and several follow-up studies have indeed identified different optimal solutions. For instance, Brand et al. (2023) reported a 1-factor solution, while Lin et al. (2023) and Campos et al. (2021) found bifactor solutions (i.e., one general factor above a set of lower-level factors, Rodriguez et al., 2016) to be the best fit. Using a 2-factors Exploratory Factor Analysis (EFA), Koike and Nomura (2023) suggested that the items could be grouped into cutaneous (itching, tickling, coughing, burping, affective touch, bruising, passing gas, sneezing, muscle sensations, sexual arousal, and taste) and visceral sensations (urination, defecation, hunger, thirst, pain, breathing, fatigue/blood sugar, temperature, vomiting, and heartbeat).

Discussions have also been focused on specific items. For instance, Murphy et al. (2019) notes that some items might measure direct interoceptive signals such as cardioception, while others might capture phenomena not perceivable through interoceptive signals alone (e.g., "bruising"; p. 119). Lin et al. (2023) additionally highlights five locally dependent pairs and three items (touch, blood sugar, bruise) with exceptionally high difficulty and low discrimination, and Campos et al. (2021) reported "tickle" to be the only item that reflected more specific factors than the general factor. Furthermore, localization issues also arose, with both "itch" and "tickle" corresponding to the same Chinese character, leading to their collapse into a single item (Lin et al., 2023).

Regarding its validity, the IAS has naturally been compared to other interoception-related measures, and shows a positive correlations with most facets of the MAIA (Gaggero et al., 2021; Mehling et al., 2018), except for the Not-Distracting and Not-Worrying subscales (Brand et al., 2023) - which were previously highlighted as related to non-interoceptive abilities (Ferentzi et al., 2021). Interestingly, findings on the correlation between the IAS and the body awareness dimension of the BPQ, have been mixed: some studies report small positive correlations (Brand et al., 2023; Campos et al., 2021; Koike & Nomura, 2023), while others find small negative correlations (Lin et al., 2023) or no correlation at all (Gaggero et al., 2021; Murphy et al., 2019). Positive correlations have also been observed with the "observation" and "description" subscales of the Five Facet Mindfulness Questionnaire (FFMQ, Baer et al., 2006; Brand et al., 2023; Koike & Nomura, 2023), as well as with the "non-reactivity" and "acting with awareness" subscales (Koike & Nomura, 2023). Additionally, the IAS has shown a positive correlation with the interoceptive awareness subscale of the Eating Disorder Inventory (EDI-IA, Lin et al., 2023) and a negative correlation with the Interoceptive Confusion Questionnaire (ICQ, Brewer et al., 2016), as reported by Brand et al. (2023) and Murphy et al. (2019). Lastly, the correlation with the Interoceptive Attention Scale (IATS, Gabriele et al., 2022) appears rather small (Koike & Nomura, 2023; Lin et al., 2023).

While assessing the predictive validity of an interoception scale can be conceived as theoretically challenging, expected negative associations were observed between the IAS and alexithymia (Brand et al., 2023; Campos et al., 2021; Koike & Nomura, 2023; Lin et al., 2023; Murphy et al., 2019), somatic symptoms (Brand et al., 2023; Koike & Nomura, 2023; Lin et al., 2023), depressive symptoms (Brand et al., 2023; Koike & Nomura, 2023; Lin et al., 2023), anxiety (Brand et al., 2023), neuroticism (Brand et al., 2023) and self-esteem (Murphy et al., 2019). Taken together, these findings support the IAS as measuring an adaptive aspect of interoception, although its pattern of associations with other interoception (or interoception-related) questionnaires points towards some overlap across various theoretical dimensions, casting some doubt on the orthogonal models of interoception and the possibility of its faithful capture by questionnaires.

Given the increasing relevance of interoception across psychology and neuroscience, and the growing use of the IAS as an interesting self-report measure of interoceptive accuracy, the current study aims at 1) clarifying the structure of the IAS with a mega-analytic (which involves a re-analysis at the raw data level by aggregating datasets) approach that leverages existing data and contrast the traditional CFA/SEM factor-based analyses with network-based ones (Exploratory Graph Analysis); 1) provide an overview of the dispositional correlates of the IAS, clarifying its general pattern of associations, which is key to better understand the nature, place and role of interoception questionnaires within a larger context.

Study 1

Study 1 will re-analyse and assess the factor structure of the IAS by taking advantage of the large number of open-access datasets (Arslanova et al., 2022; Brand et al., 2022; Brand et al., 2023; Campos et al., 2021; Gaggero et al., 2021; Lin et al., 2023; Murphy et al., 2019; Poerio et al., 2024; Todd et al., 2022; Von Mohr et al., 2023). While combining these studies might provide a more robust and generalizable understanding of the IAS' factor structure, we also additionally provide an individual analysis (i.e., applying the same method on all samples separately) to add nuance to the general picture, as all studies differ in their sample size, demographic characteristics, language, and procedure.

Methods

Datasets

Our search focused on studies citing the original IAS validation paper (Murphy et al., 2019), identifying 136 papers (as of 01/05/2024). To qualify for inclusion, papers needed to (1) provide accessible data in open-access, (2) employ the IAS as a measure, and (3) report individual IAS items scores. We also included the data of five unpublished studies. A total of 16 datasets was included (see **Table 1**).

The total number of participants was 33,272 participants (Mean = 48.09 ± 13.1, 71.3% Female).

Data Analysis

Psychometrically good items should exhibit various qualities, such as validity, reliability and discrimination, to which one of the contributing factors is the amount of inter-individual variability captured by an item. Items to which all participants' answers are concentrated around one option - i.e., exhibiting a narrow distribution - will be flagged as potentially problematic.

After examining the distributions and correlations of all IAS items, we will test for "redundant" items (e.g., due to multicollinearity or local dependency) using Unique Variable Analysis (UVA, Christensen et al., 2023), a novel method derived from network psychometric designed to identify and merge items that share substantial variance (which can distort the structure estimation). We will use a threshold of 0.30 that detects large to very large overlap to remain conservative and only suggest scale modifications if strongly justified.

Following the analysis of items, we will analyze the factor structure of the IAS using three different approaches, each with particular trade-offs and assumptions, to provide a multi-verse picture of likely solutions. Namely, we will apply traditional exploratory and confirmatory Factor Analysis (EFA/CFA), hierarchical clustering (HCA), and Exploratory Graph Analysis (EGA), to the whole sample, as well as to each dataset separately (details being available in the analysis document at https://github.com/RealityBending/InteroceptionIAS), and our decisions and conclusions will try to take into account both levels of the analysis.

By combining network analysis with psychometric methods, the recently-developed Exploratory Graph Analysis (EGA, H. F. Golino & Epskamp, 2017) framework allows to jointly estimate the number of dimensions (i.e., groups of items), the structure as well as its stability (H. Golino et al., 2020; H. F. Golino & Epskamp, 2017). Evidence has underlined its suitability as an alternative to traditional factor analysis, addressing some of its limitations such as the assumption of a "latent" source of variability, possible biasing in the estimation of the optimal factor numbers depending on sample size, and the poor performance of other methods in complex population structures, while remaining comparable and interpretable (Christensen & Golino, 2021b; Jiménez et al., 2023). At a fundamental level, EGA conceptualizes variables as nodes in a network, with connections (edges) reflecting associations between them. Clustering these nodes reveals distinct communities of related items, in practice akin to traditional latent factors - but without explicitly assuming their presence (Christensen & Golino, 2021b). We used the EGAnet package (Christensen & Golino, 2021a) to fit a hierarchical EGA with the Leiden community detection algorithm.

While EGA offers a robust alternative to traditional factor analysis, factor analysis remains a widely used method for dimensionality assessment. As our goal is to provide a general - yet nuanced - picture, with room to show potential discrepancies emerging from the methods used, we will also include it in the present study. Unlike EGA, factor analysis assumes that a latent source of variability — a common latent variable — underlies the observed set of manifest variables (Cosemans et al., 2022). A critical step in factor analysis is determining the optimal number of factors, for which we will used the Method Agreement Procedure (Lüdecke et al., 2021), which involves a consensus-based decision based on multiple factor estimation methods applied concurrently.

Finally, we will also apply Hierarchical Clustering Analysis (HCA, Murtagh & Legendre, 2014), which differs from factor analysis in that it does not assume any latent source of variability, but instead iteratively groups items based on their similarity (e.g., correlation) into a hierarchy of clusters. The benefits of HCA include its interpretability and ability to capture complex relationships among items without relying on

Table 1

Description of the samples used in the study.

Sample	Reference	Language	N	Difference	Age (Mean ± SD)	Range	Female %	Availability
	Murphy et al., (2020)							osf.io/3m5nh
Sample 1a		English	451		25.8 ± 8.4	18-69	69.4%	
Sample 1b		English	375		35.3 ± 16.9	18-91	70.1%	
Sample 2	Gaggero et al., (2021)	English and Italian	814		24.9 ± 5.3	18-58	60.3%	osf.io/5x9sg
Sample 3	Campos et al., (2022)	Portuguese	515		30.7 ± 10.5	18-72	59.6%	osf.io/j6ef3
Sample 4	Todd et al., (2022)	English	802		48.6.6 ± 14.1*	18-92*	50%*	osf.io/ms354
Sample 5	Arslanova et al., (2022)	English	143		28.5 ± 7.6	18-73	46.8%	osf.io/mp3cy
Sample 6	Brand et al., (2022)	German	619		43.9 ± 14.5	18-78	78.7%	osf.io/xwz6g
	Brand et al., (2023)							osf.io/3f2h6
Sample 7a		German	522		23.4 ± 6.7	18-79	79.5%	
Sample 7b		German	1993		32.0 ± 12.6	16-81	77.7%	
Sample 7c		German	802		27.3 ± 9.3	18-72	68.9%	
	Lin et al., (2023)							osf.io/3eztd
Sample 8a		Chinese	1166	Collapsed "Itch" and "Tickling"	32.5 ± 8.4	16-60	57.0%	
Sample 8b		Chinese	500	Collapsed "Itch" and "Tickling"	37.4 ± 7.4	20-60	56.2%	
Sample 9	VonMohr et al., (2023)	English	21843		56.5 ± 14.4	18-93	73.2%	osf.io/7p9u5
Sample 10	Makowski et al., (2023a)	English	485	Analog scales. No Temperature, Blood sugar and Cough items	30.1 ± 10.1	18-73	50.3%	github.com/RealityBending/IllusionGameReliability
Sample 11	Makowski et al., (2023b)	English	836	Analog scales	25.1 ± 11.3	17-76	53.0%	github.com/DominiqueMakowski/PHQ4R
Sample 12	Makowski et al., (2023c)	English	146	Analog scales	21.1 ± 4.3	18-50	76%	github.com/RealityBending/InteroceptionPrimals
Sample 13	Makowski et al., (2024)	English	737		36.8 ± 14.9	17, 87	57.3%	github.com/RealityBending/InteroceptionScale
Sample 14	Poerio et al., (2024)	English	107		26.8 ± 9.2	18-57	74.8%	osf.io/49wbv
Sample 15	Poerio et al., unpublished	English	131		30.9 ± 12.0	18-60	75.9%	
Sample 16	Arjona et al., unpublished	English	279		26.4 ± 13.2	18-79	67.7%	
Sample 17	Total		33272		48.09 ± 13.1	17-93	71.3%	

*Information taken from the sample description of relevant paper rather than recomputed.

strict assumptions about data distribution or latent variables.

In a typical 2-step fashion, this first analysis run will be followed by a structure refinement with a further selection of items, and the final pool of items will be tested again. Additionally, various solutions (e.g., adding general factors) will be statistically compared using Confirmatory Factor Analysis (CFA).

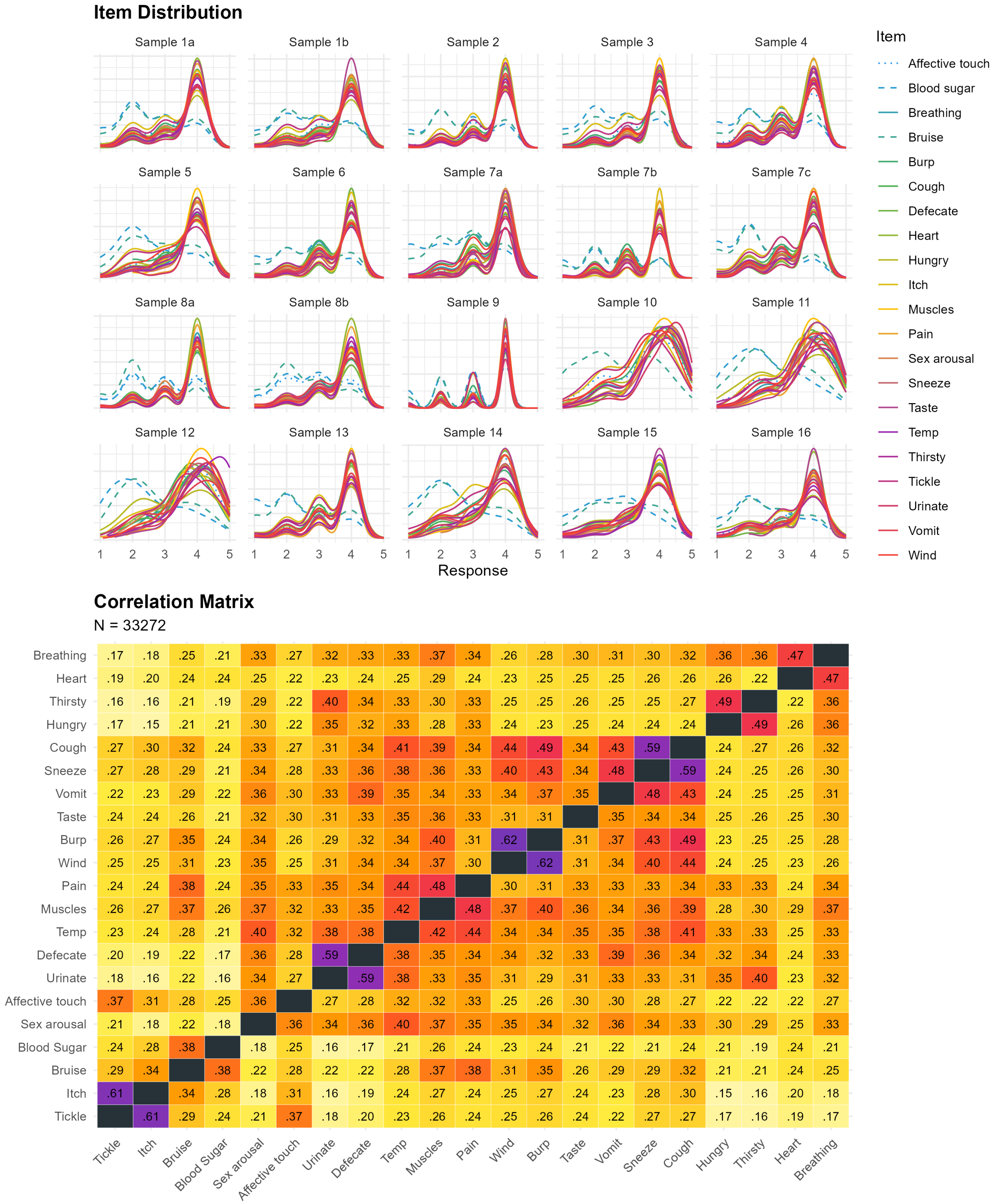
Results

The distribution of the items across samples suggests the presence of a consistent modal value (Figure 2). In other words, participants are most likely to answer 4/5 (i.e., agree) on most items, with the exception of "blood sugar" and "bruise", which exhibit a different distributional pattern with a lower mode (~2/5). While it is not a problem *per se*, the contrasting distribution might be indicative of items with non-homogeneous psychometric "difficulty". This is also the case for "affective touch" in samples 8a and 8b (the Chinese validation samples), which might indicate localisation issues. Additionally, one can note the low occurrence of extreme values (1 and 5), meaning that the bulk of answers varies between 3 values (assuming the IAS is implemented as a 5-point Likert scale following its validation). The samples using an analogue scale (samples 10, 11 and 12 in the figure) display a more continuous and progressive spread of answers, seemingly improving the interindividual variability, although potentially displaying a secondary lower mode at ~2 (which might be suggesting the existence of potential clusters of participants). The correlation matrix between all items shows an overall positive correlation pattern, with highly correlated pairs of items (e.g., Tickle-Itch, Urinate-Defecate, Pain-Wind, Hungry-Thursty) or triplets (Vomit-Sneeze-Cough, Temperature-Muscles-Pain).

UVA flagged two strongly redundant variables, "itch" and "tickle" - suggesting to remove the latter. Several more pairs of items were flagged as moderately redundant ("wind" and "burp"; "urinate" and "defecate") and mildly redundant ("sneeze" and "cough"; "heart" and "breathing"; "hungry" and "thirsty"). These patterns consistently appeared in most samples when considered individually. We removed "tickle" from further analysis due to its high redundancy (and because

Figure 2

Top: Distribution of responses across datasets reveals a consistent modal value, typically around 4 or 5 (indicating agreement), except for 'blood sugar' and 'bruise' - in dashed lines - and 'affective touch' (dotted lines) in the Chinese validation sample, which have lower modes. Most responses cluster around the middle values, with few extreme scores (1 and 5). Samples using an analogue scale (10a, 10b, 10c) show a more continuous distribution and increased interindividual variability. Since most samples use Likert scales (discrete), density plots may not be the most accurate representation but were chosen to clearly highlight variability patterns in the data. Bottom: The correlation matrix between all items shows an overall positive correlation pattern, with correlated pairs (e.g., Wind, Burp) of items or triplets (e.g., Vomit, Sneeze and Cough).



it is absent from some datasets due to translation issues).

The HCA highlighted pairs and triplets of items consistently grouped together across samples, such as “wind” and “burp”, “sneeze” and “cough”, “itch” and “bruise”, “urinate” and “defecate”, and “pain”, “muscles”, and “temperature”. This pattern was largely replicated by the EGA, with the additional presence of a unique cluster comprising “Sex arousal”, “Affective touch”, “Temperature”, “Pain”, “Muscles”, and “Taste”. EFA suggested the optimal number of factors to be 3, yielding one dimension with expulsion-related items (“burp”, “wind”, “cough”, “sneeze”, and “vomit”), a second dimension with visceroceptive items (“heart”, “breathing”, “hungry”, “thirsty”, “urinate”, and “defecate”), and a third dimension with skin-related items (“bruise” and “blood sugar”).

Importantly, this initial structure analysis run highlighted some problematic items: “taste” typically displayed a lone or unstable pattern of associations, “affective touch” exhibited cross-loadings and instability, “vomit” was less strongly associated with other items, and “itch” did not form a consistent cluster. Finally, “temperature” and “sexual arousal” showed redundant patterns of associations but were less reliable. These 6 items were thus removed, and a second run of structure analysis was performed on the remaining 14 items.

HCA and EGA yielded highly consistent results, emphasizing pairs of items, namely Hungry-Thirsty, Bruise-Blood sugar, Urinate-Defecate, Muscles-Pain, Breathing-Heart, Cough-Sneeze, Wind-Burp. HCA also significantly grouped the Urinate-Defecate and Muscles-Pain pairs, as well as expulsion items (Wind-Burp and Cough-Sneeze). EFA suggested once again 3 factors as the optimal solution, with the first factor including expulsion-related items (“burp”, “wind”, “cough”, “sneeze”), the second factor being related to the Urinate-Defecate pair, and the third factor comprising the remaining items.

We then fitted and compared using CFA various candidate structures emerging from the previous analyses, including a 1-factor model (the G-model), a 3-factor model (EFA), a 3+1 model (EFA + general factor), a 5-factor model (HCA), a 5+1 model (HCA + general factor), a 7-factor model (EGA), and a 7+1 model (EGA + general factor). The EGA model with 7 factors of item pairs provided the best fit with the lowest RMSEA (0.026), lowest χ^2 (72.825), highest CFI (0.988). It was followed by the EGA + general factor model (RMSEA = 0.059, χ^2 = 180.403, CFI = 0.065), and the HCA model with 5 factors (RMSEA = 0.062, χ^2 = 182.492, CFI = 0.921). The other models performed poorly, with RMSEA > 0.08 and CFI < 0.90. The unique-factor model yielded the lowest BIC (which takes into account the number of factors), followed by the EGA model. All the other models displayed significantly (BIC-based Bayes Factor < 1/100) lower evidence compared to the EGA model.

This result was relatively consistent across unique datasets with the EGA model providing the best fit. In all samples where CFA converged, it showed excellent performance with low RMSEA (≈ 0.025–0.052), high CFI (≥ 0.95), and substantially lower χ^2 compared to alternative models. Adding a general factor systematically worsened fit, while hierarchical clustering and EFA-based models showed poorer performance (RMSEA > 0.05, CFI < 0.95). Although BIC values occasionally favored alternative specifications, Bayes factor comparisons overwhelmingly supported the EGA solution, confirming its robustness across datasets.

Discussion

Study 1 aimed to systematically evaluate the structure of the IAS. Previous research reported conflicting models, from unidimensional (Brand et al., 2023) to bifactor (Campos et al., 2021; Lin et al., 2023) or two-factor solutions (Koike & Nomura, 2023; Murphy et al., 2019). Concerns about specific items (e.g., low discrimination, local dependencies, and items reflecting phenomena less accessible to interoception) prompted a mega-analytic approach using both factor-analytic and network-based methods.

Analyses revealed that seven item pairs (e.g., Hunger–Thirst, Urinate–Defecate, Cough–Sneeze) provided a robust and replicable structure, leading to a refined 14-item IAS. These pairs consistently emerged across datasets and yielded superior fit indices compared to alternative models. This suggests that the IAS does not measure a broad latent construct of interoceptive accuracy but reflects clusters of tightly coupled items tied to specific bodily sensations.

The presence of redundant or unstable items (e.g., Tickle, Taste, Affective Touch) undermines interpretability and likely explains inconsistencies in prior factor-analytic findings (Campos et al., 2021; Lin et al., 2023). Nevertheless, the structured item pairs correspond to intuitive bodily domains - visceral, expulsion, musculoskeletal - supporting ecological validity. These findings highlight the need for targeted scale refinement and motivate future development of context-sensitive, multidimensional interoception measures.

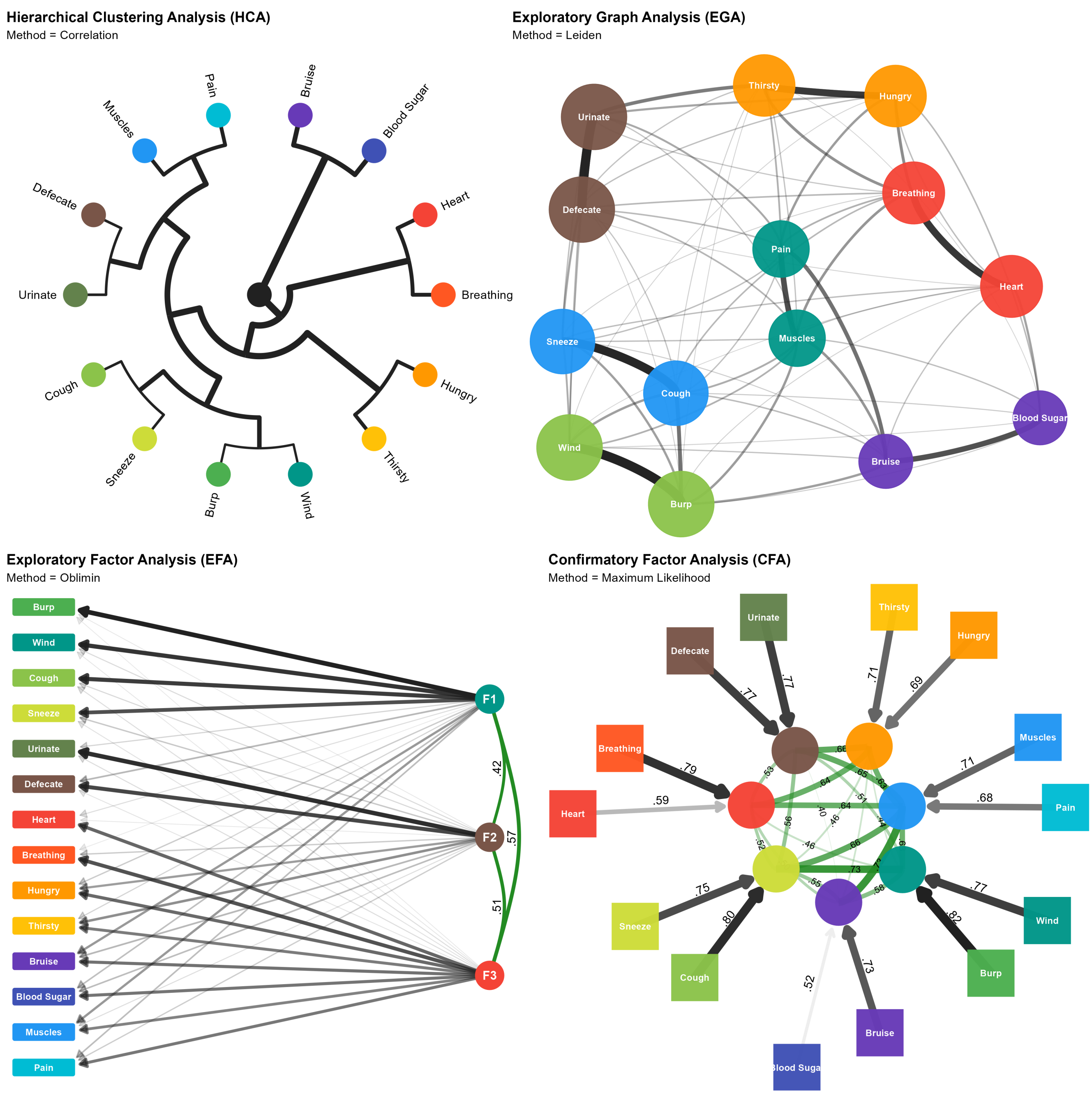
In summary, study 1 suggests that once problematic items are removed, the IAS is best conceptualized as a set of item clusters rather than a unidimensional scale. This structural insight provides the foundation for Study 2, where dispositional correlates are examined to further evaluate the IAS's construct validity.

Study 2

Building on the structural findings from Study 1, the second study turns to the validity of the IAS by examining its dispositional correlates. Although the IAS has been widely used, evidence regarding its associations with personality traits, affective dispositions, and related psychological constructs remains scattered and inconsistent (Brand et al., 2023; Lin et al., 2023; Murphy et al., 2019; Todd et al., 2022). Study 2 therefore had two primary goals. First, to provide a robust

Figure 3

Four structure analysis methods (HCA, EGA, EFA, CFA) were applied and converged on a consistent optimal solution of 14 items formed seven pairs: Hungry-Thirsty, Bruise-Blood sugar, Urinate-Defecate, Muscles-Pain, Breathing-Heart, Cough-Sneeze, Wind-Burp. While EFA suggested 3 factors, CFA confirmed the superiority of the 7-factor model over alternative structures.



overview of the dispositional correlates of the IAS by leveraging the large number of available datasets. Second, to evaluate the practical benefits of using the refined 14-item version of the scale identified in Study 1, compared to the original 21-item unidimensional solution, in order to assess whether structural improvements translate into clearer and more reliable external associations.

Methods

Materials

We selected measures that appeared in multiple datasets from Study 1 and/or were relevant given the scope of the study, i.e., constructs related to physiology, mood, personality, psychopathology, and beliefs and misbeliefs. We merged scores of regular and abridged versions for conciseness where applicable.

Interoception Related. The Multidimensional Assessment of Interoceptive Awareness Version-2 (MAIA-2, Mehling et al., 2018) is a 37-item questionnaire assessing eight dimensions of interoception: Noticing, Not-Distracting, Not-Worrying, Attention Regulation, Emotional Awareness, Self-Regulation, Body Listening, and Trust. Responses are rated on a 6-point Likert scale (e.g., "I notice when I am uncomfortable in my body").

The Body Perception Questionnaire Short Form (BPQ-SF, Cabrera et al., 2018) contains 46 items measuring Body Awareness and Autonomic Reactivity on a 5-point Likert scale (e.g., "My heart often beats irregularly"). The very-short form (BPQ-VSF) contains 12 items from the Body Awareness subscale. For this study, scores from both versions were combined.

The Interoceptive Confusion Questionnaire (ICQ, Brewer et al., 2016) consists of 20 items rated on a 5-point Likert scale (e.g., "I cannot tell when my muscles are sore or tight") assessing difficulties in interpreting non-affective physiological states, such as pain or hunger.

The Toronto Alexithymia Scale (TAS-20, Bagby et al., 1994) contains 20 items rated on a 5-point forced-choice scale (e.g., "I have feelings I can't quite identify"), divided into three dimensions: difficulty identifying feelings, difficulty describing feelings, and externally oriented thinking.

The Bermond–Vorst Alexithymia Questionnaire (BVAQ, Vorst & Bermond, 2001) consists of 40 items across five subscales: fantasising, identifying, analysing, verbalising, and emotionalising. Items are rated on a 5-point Likert scale. The BVAQ also provides two higher-order factors, namely affective and cognitive (e.g., "I like to tell others how I feel").

Mood and Anxiety Related. The Beck Depression Inventory-II (BDI-II, Beck et al., 1996) is a 21-item measure of depressive symptom severity, including psychological (e.g., guilt) and physiological (e.g., loss of energy) components. Items are rated on a 4-point scale (e.g., "I feel sad much of the time").

The Patient Health Questionnaire-4 (PHQ-4, Kroenke et al., 2009) is a 4-item screening tool for depression and anxiety, rated on a 4-point Likert scale (e.g., "Little interest or pleasure in doing things"; "Not being able to stop or control worrying"). The PHQ-2 (PHQ-2, Kroenke et al., 2003) and General Anxiety Disorder-2 (GAD-2, Kroenke et al., 2007) scores were combined with the PHQ-4 scores.

The Patient Health Questionnaire-9 (PHQ-9, Kroenke et al., 2001) includes 9 items assessing depressive symptoms on a 4-point Likert scale (e.g., "Feeling tired or having little energy?").

The Patient Health Questionnaire-15 (PHQ-15, Kroenke et al., 2002) assesses somatic symptom distress with 15 items rated on a 3-point scale (e.g., "Over the last week, how often have you been bothered by back pain?").

The Short Mood and Feelings Questionnaire (SMFQ, Angold et al., 1995) is a 13-item measure of recent depressive symptoms in children aged 6–17, rated on a 3-point scale (e.g., "I felt miserable or unhappy").

The State–Trait Anxiety Inventory Trait Version (STAI-T, Spielberger, 1970) is a 20-item trait anxiety questionnaire rated on a 4-point Likert scale (e.g., "I worry too much about something that really doesn't matter"). Scores from the shorter version of the STAIT-5 (Zsido et al., 2020) were combined with the STAI-T.

The Generalized Anxiety Disorder Scale (GAD-7, Spitzer et al., 2006) is a 7-item measure of generalized anxiety symptoms rated on a 4-point Likert scale (e.g., "Not being able to stop or control worrying").

Personality. The NEO Five-Factor Inventory, neuroticism subscale (NEO-FFI, Costa et al., 1992) is a 12-item measure of neuroticism, assessing the general tendency to experience negative emotions such as fear, sadness, or disgust. Items are rated on a 5-point Likert scale (e.g., "I often feel inferior to others").

The Mini International Personality Item Pool (MINI-IPIP6, Sibley et al., 2011) includes 24 items measuring six broad traits: Agreeableness, Conscientiousness, Neuroticism, Openness, Extraversion, and Honesty-Humility. Items were scored on visual analogue scales in our samples (originally 7-point Likert scales; e.g., "Have a vivid imagination").

The Big Five Inventory Short Form (BFI-S, Rammstedt & John, 2007) is a 10-item measure of the Big Five traits, rated on a 7-point Likert scale (e.g., "I see myself as someone who … is considerate and kind to almost everyone").

Psychopathology. The Personality Inventory for DSM-5 Short Form (PID-5-SF, Thimm et al., 2016) consists of 25 items rated on a 4-point Likert scale, measuring five domains: disinhibition, antagonism, detachment, negative affect, and psychoticism (e.g., "Plenty of people are out to get me").

The Schizotypal Personality Questionnaire – Brief Revised (SPQ-BR, Davidson et al., 2016) contains 32 items rated on a 5-point Likert scale, assessing four primary dimensions: cognitive-perceptual, interpersonal, disorganized, and social anxiety. These are subdivided into nine secondary factors, including constricted affect, eccentricity, magical thinking, lack of close friends, odd speech, referential thinking, social anxiety, suspiciousness, and unusual perceptions (e.g., "Plenty of people are out to get me").

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD, Zanarini, 2003) is a 10-item dichotomous screening measure for borderline personality disorder (present/absent; e.g., "Have any of your closest relationships been troubled by a lot of arguments or repeated breakups?").

The Autism Spectrum Quotient Short Form (ASQ-S, Hoekstra et al., 2011) is a 28-item measure of five autistic traits: social skills, adherence to routines, cognitive flexibility, imagination, and patterns/numbers. Items are rated on a 4-point Likert scale (e.g., "I find it difficult to work out people's intentions").

Beliefs and Misbeliefs. The Generic Conspiracist Beliefs Scale (GCBS, Brotherton et al., 2013) is a 15-item measure assessing five dimensions of conspiracy beliefs: government malfeasance, extraterrestrial cover-up, malevolent global conspiracies, personal wellbeing, and control of information. Items are scored on a 5-point Likert scale (e.g., "Secret organizations communicate with extraterrestrials, but keep this fact from the public").

The Primal World Beliefs Inventory (PI-99, Clifton et al., 2019) is a 99-item measure assessing 26 high-order beliefs about the world. Items are scored on a 5-point Likert scale (e.g., "Nearly everything in the world is beautiful") The PI-18 (Clifton & Yaden, 2021) is a short form of the orginal questionnaire assessing the 4 main world beliefs: the world is Good (vs. Bad) and its three dimensions: Safe (vs. Dangerous), Enticing (vs. Dull), and Alive (vs. Mechanistic). Scores from the PI-18 scores were combined with the PI-99.

The Lie Scale (LIE, Makowski et al., 2023) is a 16-item measure of lying tendencies across four components: Ability, Frequency, Negativity, and Contextuality. Items are scored on visual analogue scales (e.g., "I am a good liar").

Data Analysis

All variables from the 16 open-access datasets will be extracted and standardized (z-scored). For continuous predictors, we will first ensure consistent ranges across datasets; when necessary, ranges will be rescaled to achieve comparability. In cases where scales of different lengths measure the same construct, scores will be harmonized: when possible, means will be computed from overlapping items, while in other cases total scores will be normalized by the number of items. We will collapse redundant subscales (e.g., PHQ-2 and GAD-2 within PHQ-4) into unified composite scores. Similarly, short forms of trait anxiety (e.g., STAI-5-trait) will be collapsed with their longer counterpart (STAI-T), with scores computed as the mean to ensure comparable ranges. Shorter formats of the PI (PI-18) will be combined with scores of its longer version (PI-99). Missing data will be handled via listwise deletion on a per-model basis prior to fitting.

Associations will be estimated using multilevel regression models implemented in the glmmTMB package (McGillycuddy et al., 2025). Each predictor will be entered separately, with the seven IAS item pairs identified in Study 1 (Hungry–Thirsty, Muscles–Pain, Wind–Burp, Urinate–Defecate, Breathing–Heart, Bruise–Blood, Cough–Sneeze) and the original IAS scale serving as outcomes. Models will include random intercepts and slopes by sample to account for between-study heterogeneity; when models fail to converge, a reduced structure with random intercepts only will be used. This approach will treat datasets as clusters, thereby providing pooled estimates of the associations while accounting for both within- and between-sample variability, akin to a meta-analytic framework.

From each fitted model, we will extract standardized regression coefficients (β), 95% confidence intervals, p-values, convergence status, and effective sample size. Participants identifying as "Other" will be excluded from analyses due to limited representation.

Results

Age showed consistent positive standardized associations across IAS correlates (mean $\beta = .13$), with all effects significant. Gender displayed near-zero effects (mean $\beta = -.06$), significant only for the Hungry–Thirsty, Wind–Burp, and Urinate–Defecate pairings.

MAIA subscales showed robust positive standardized effects with the IAS. Noticing yielded the strongest association (mean $\beta = .26$), followed by Emotional Awareness ($\beta = .20$) and related subscales - Body Listening, Attention Regulation, Trusting, and Self-Regulation, each in the .20–.26 range, all significant. Effects were generally strongest for the full IAS, except for Body Listening, which peaked for the Bruise–Blood pairing. Not Distracting was weakly but significantly positive (mean $\beta = .06$), primarily for Hungry–Thirsty, Bruise–Blood, and Muscles–Pain. Not Worrying was significant only for the full IAS and the Hungry–Thirsty and Urinate–Defecate pairings ($\beta \approx -.01$ to $.11$).

Autonomic Reactivity (BPQ; mean $\beta = -.19$) was reliably negative across pairings, reaching significance for all except the Bruise–Blood sugar and Breathing–Heart pairings. In contrast, Body Awareness (BPQ; mean $\beta = .12$) was weakly positive and significant only for Bruise–Blood.

Interoceptive deficits showed consistent negative standardized effects. Alexithymia measures (TAS, BVAQ) were robustly negative ($\beta \approx -.18$ to $-.24$), except for the BVAQ Affective subscale, which was positive for Bruise–Blood ($\beta = .16$). Interoceptive confusion (ICQ) was also strongly negative across all IAS measures (mean $\beta \approx -.42$; range $= -.49$ to

−.20).

Depression (BDI; mean β = −.20) and anxiety (STAI-T; mean β = −.19) showed consistent negative associations across the IAS and pairings, strongest for Hungry–Thirsty. Related constructs diverged somewhat: MFQ was largely nonsignificant; PHQ-9 and other mood indices (PHQ-15, PHQ-4, GAD-7) produced weaker but often reliable negative βs (typically −.10 to −.15), with effects sometimes limited to particular pairings (e.g., PHQ-15 nonsignificant for Wind–Burp and Breathing–Heart). Hungry–Thirsty consistently yielded the strongest standardized effects.

Autistic traits (ASQ) were generally negatively associated with IAS scores. Imagination (mean β = −.18) showed the most consistent pattern across pairings except Cough–Sneeze. Most full-IAS associations were driven by Hungry–Thirsty, Bruise–Blood sugar, and Muscles–Pain, which yielded the largest negative coefficients. Beyond these, effects weakened or became nonsignificant- Patterns and Numbers showed none, and Borderline Personality was associated only with Hungry–Thirsty (mean β = −.14).

Schizotypal traits showed similarly negative coefficients, with Social Anxiety (mean β = −.11) and No Close Friends (mean β = −.12) most consistent. Some facets were unrelated to the overall IAS but linked to specific pairings: Unusual Perceptions with Muscles–Pain and Urinate–Defecate, and Eccentric with Hungry–Thirsty and Urinate–Defecate.

Among Big Five traits, Openness (mean β = .10), Agreeableness (β = .08), and Conscientiousness (β = .11) showed small but consistent positive effects, most evident for the full IAS. Honesty–Humility (mean β = −.10) showed a small negative effect, primarily for Hungry–Thirsty and Bruise–Blood sugar. Neuroticism (mean β = −.13) was negatively associated with the IAS, with the strongest β for Bruise–Blood (β = −.23).

Primal world beliefs were largely null, with small positive standardized coefficients for certain beliefs: "Understandable," "Alive," and "Hierarchical" were weakly positive for the full IAS and Hungry–Thirsty (β ≈ .05–.10), while "Good" was uniquely positive for Muscles–Pain (β ≈ .09). Within the Lying Profile, only Lying Contextuality showed a modest negative coefficient for Bruise–Blood (β = −.12).

Conspiracy Beliefs produced selective small positive effects. Global Conspiracies (GCBS) were associated with the full IAS and several pairings (mean β = .10), Government Malfeasance (β = .05) with Cough–Sneeze, and Extraterrestrial beliefs (β = .03) with Muscles–Pain.

Overall, the full IAS produced the greatest number of significant standardized effects (29), though many were driven by specific pairings rather than uniform scale-wide associations. Hungry–Thirsty (17) and Bruise–Blood sugar (10) consistently yielded the strongest standardized coefficients, especially for interoceptive, mood, and psychopathology measures. Other pairings (Cough–Sneeze, Wind–Burp, Urinate–Defecate) contributed little unique variance and rarely yielded meaningful standardized effects.

Discussion

Study 2 examined the dispositional correlates of the IAS and whether the refined 14-item version offers practical advantages over the original 21-item scale. The findings clarify the conceptual boundaries of self-reported interoceptive accuracy and underscore the domain-specific nature of these standardized associations.

The IAS showed positive associations with most MAIA facets, except for the Not-Distracting and Not-Worrying subscales, replicating prior findings (Brand et al., 2023). Associations with the BPQ Body Awareness subscale were consistent with previous reports ona non-existing association between these measures (Gaggero et al., 2021; Murphy et al., 2019), whereas indices of autonomic reactivity were more strongly and negatively associated with IAS scores, supporting earlier observations that reactivity may reflect aspects of bodily signal processing captured by self-reported interoceptive accuracy (Todd et al., 2022).

In contrast, measures of interoceptive difficulties, including alexithymia and interoceptive confusion, were negatively associated with IAS scores, consistent with prior research (Brand et al., 2023; Koike & Nomura, 2023; Lin et al., 2023; Murphy et al., 2019). These associations were strongest for visceral pairings, such as Hungry–Thirsty and Bruise–Blood sugar, and weaker or nonsignificant for other bodily domains, highlighting the domain-specific nature of interoceptive self-report.

Across mood and interoceptive variables, most item pairings displayed small-to-moderate standardized coefficients, indicating shared variance between interoceptive self-reports and affective tendencies. In contrast, associations with psychopathology were more selective: negative βs were primarily observed for visceral pairings such as Hungry–Thirsty and Bruise–Blood sugar, whereas other domains yielded weaker or nonsignificant effects. This pattern suggests that self-reported interoceptive accuracy is more strongly associated with affective and clinical traits for visceral sensations than for exteroceptive or respiratory ones (Koike & Nomura, 2023; Murphy et al., 2019).

Associations with personality traits followed a similar domain-specific structure. Effects for the full IAS largely reflected contributions from specific pairings: Openness showed its strongest βs for Breathing–Heart, Neuroticism and Honesty–Humility for Hungry–Thirsty and Bruise–Blood sugar, and Conscientiousness for the full scale. These results indicate that associations between interoceptive self-report and personality dimensions may depend on the bodily domain represented.

Associations with belief systems were uniformly small. Weak positive standardized coefficients emerged between the

IAS and certain primal world beliefs (e.g., perceiving the world as comprehensible or alive) and selected conspiracy belief dimensions, particularly Global Conspiracies and Personal Wellbeing. These effects were limited to specific pairings - most often Hungry–Thirsty and Bruise–Blood sugar—indicating that links between interoceptive self-report and broader cognitive or worldview constructs are modest and domain-specific.

The refined 14-item IAS performed similarly to the original 21-item version, reproducing comparable standardized effect patterns while reducing redundancy. This suggests that the shortened scale captures the same latent structure with improved efficiency and interpretability, though further validation across independent samples is warranted.

Overall, standardized effects between the IAS and external constructs were not uniform but varied across bodily domains. Visceral pairings, particularly Hungry–Thirsty and Bruise–Blood sugar, showed the strongest and most consistent standardized relationships with interoceptive, affective, and personality measures. The refined 14-item IAS preserves these domain-specific associations, offering a more parsimonious, though potentially approximateassessment of self-reported interoceptive accuracy.

General Discussion

The present study provides a comprehensive evaluation of the IAS, examining both its structure (study 1) and its dispositional correlates (study 2). These studies offer novel insights into the psychometric characteristics, conceptual interpretation, and practical applications of the IAS. In this discussion, these findings are contextualized within the broader landscape of interoception research.

A short version of the IAS was previously proposed by Lin et al. (2023), who suggested a 12-item version by removing items with local dependency, high difficulty, and/or low discrimination. Their parallel analysis supported a unidimensional structure, although this method tends to underestimate factor numbers in multidimensional data (H. F. Golino & Epskamp, 2017; Markos & Tsigilis, 2024). Consequently, the 12-item IAS was interpreted as reflecting a single underlying factor of interoceptive accuracy.

In contrast, Study 1 applied multiple complementary methods, including EGA, FA, and HCA. Across datasets and methods, the same tightly coupled item pairs consistently emerged, representing distinct bodily domains such as visceral sensations, musculoskeletal cues, and expulsion phenomena. Several items (e.g., Tickle, Taste, Affective Touch) were removed due to instability, redundancy, or weak associations, resulting in a refined 14-item IAS that retained psychometric robustness while improving interpretability and efficiency.

These findings challenge the assumption of IAS unidimensionality (Brand et al., 2023; Murphy et al., 2019). Whereas Lin et al. (2023) 's IRT-based approach treated locally dependent items as artefacts to be removed, our results suggest that such dependencies may reflect theoretically meaningful pairs (e.g., Heart–Breath). Because IRT assumes unidimensionality (Nguyen et al., 2014), it may have obscured these structured item clusters rather than as theoretically relevant structures within interoceptive accuracy.

Study 2 provided further evidence for a domain-specific perspective. Certain item pairings, particularly Hunger-Thirst, Bruise-Blood sugar, and Muscles-Pain, showed consistent negative correlations with depression and anxiety measures, sometimes approaching or even surpassing the strength of associations observed for the full IAS scale. Notably, Hunger–Thirst and Bruise–Blood sugar were also the only pairs, out of the seven examined, that displayed significant positive correlations with the BPQ Body Awareness subscale, whereas the full IAS scale showed no such association.

Additionally, specific item pairings revealed distinct patterns of association with other interoceptive measures and mental health indices that were not captured by the full IAS. For instance, while the full IAS showed no significant associations with Borderline Personality, the Hunger-Thirst pair exhibited negative correlations. These findings collectively suggest that item-level analyses may uncover more nuanced relationships than those identified by the overall scale. This pattern highlights that questionnaires like the IAS may capture multiple facets of interoception rather than a single underlying dimension.

These findings are in line with inconsistencies found in the literature concerning associations between the IAS and the BPQ Body Awareness. While some studies report significant negative associations (Lin et al., 2023), others have observed significant positive associations (Brand et al., 2023; Campos et al., 2021; Gaggero et al., 2021; Todd et al., 2022), and others none at all (Murphy et al., 2019). In our study the full IAS was also not associated with the BPQ Body Awareness. This pattern is consistent with the idea that this subscale may capture maladaptive bodily experiences or autonomic dysfunction, rather than interoceptive accuracy per se.

However, the Body Awareness subscale of the BPQ has also been criticized for its ambiguous instructions. It is often unclear whether respondents should interpret the items as assessing attentional focus on bodily sensations, perceived accuracy in detecting them, or simply the frequency of experiencing such signals. As a result, studies have reported that observed associations vary depending on how the BPQ Body Awareness is interpreted (Campos et al., 2021; Gabriele et al., 2022; Gaggero et al., 2021). As such, it is unclear whether the lack of a relationship between the IAS and the BPQ Body Awareness in this study reflects a genuine divergence between the constructs measured by the two scales, or instead arises from ambiguities in how the BPQ Body Awareness is operationalized and interpreted across studies.

We highlight that some IAS pairings may capture interoceptive self-reported accuracy more effectively than the full IAS measure, raising questions about the interpretability of the full scale. In contrast , other pairings (e.g., Cough–Sneeze, Wind–Burp) were weakly or inconsistently related to other variables, raising concerns about their usefulness. Together, these findings indicate that the strength and relevance of interoceptive self-reported accuracy may vary systematically across physiological domains.

At the same time, the interpretation of such associations is complicated by the lack of contextualisation in the items themselves. For example, a racing heart is often associated with anxiety, yet the same sensation can also arise during exercise or other forms of physiological arousal. Without contextual anchors, it remains unclear which bodily states participants are drawing upon when responding, highlighting the interpretive ambiguity of self-reports. In line with prior work emphasizing the importance of contextual factors in the measurement of interoception (Vlemincx et al., 2023), our results suggest that the value of self-report lies less in broad aggregation across domains and more in the targeted assessment of physiologically and situationally meaningful contexts.

The dispositional correlates of the IAS help delineate the conceptual boundaries of interoceptive accuracy as a construct. Self-reported interoceptive accuracy, as assessed by the IAS, has been robustly associated with mood-related outcomes, particularly depression and anxiety, underscoring its relevance for affective regulation and bodily state monitoring (Benau, 2023; Brand et al., 2023). This aligns with a broader body of research linking affective conditions such as anxiety and depression with altered interoceptive processing (e.g., Garfinkel et al., 2015).

However, recent findings have also reported null or inconsistent associations between interoceptive abilities and mental health symptomatology, suggesting that such relationships may not generalize across all interoceptive domains or measurement modalities (Banellis et al. (2025); Jenkinson et al. (2024)). For instance, a recent review by Jenkinson et al. (2024) found no consistent evidence linking cardiac interoceptive accuracy tasks with anxiety or depression across multiple meta-analyses.Collectively, these findings indicate that relationships between interoceptive accuracy and mental health may vary across bodily domains and measurement modalities. This suggests that the IAS, while informative, may benefit from a more differentiated approach that accounts for domain-specific and contextual influences on self-reported interoceptive accuracy.

In contrast to the relatively stronger associations observed with affective measures, correlations between IAS scores and personality traits were small but directionally consistent. Modest positive associations emerged with Openness, Agreeableness, and Conscientiousness, while Neuroticism and Extraversion showed near-zero relationships, and Honesty–Humility displayed weak negative correlations. This pattern suggests that self-reported interoceptive accuracy, as indexed by the IAS, shares limited but systematic overlap with broad dispositional tendencies -particularly those reflecting intellectual curiosity and creativity, patience and tolerance in interpersonal interactions, and diligence and organizational tendencies (Ashton & Lee, 2007).

Recent evidence supports the view that relationships between interoceptive measures and personality are multifaceted rather than uniform. For instance, Haustein et al. (2024) reported significant associations between interoceptive awareness (MAIA) and all personality traits except Agreeableness, with Openness showing positive correlations with Body Listening and Emotional Awareness, and Neuroticism showing negative correlations with Self- and Attention Regulation. Similarly, Simkute et al. (2025) found that task-based indices of interoceptive sensitivity related negatively to Extraversion. Collectively, these findings indicate that interoceptive constructs show distinct and sometimes opposing patterns of association across measurement modalities, suggesting that the IAS relates to a specific, self-evaluative aspect of interoceptive functioning rather than reflecting a general interoceptive or personality-based trait.

A potential key reason for discrepancies in observed associations, including those seen with the IAS, is that self-report questionnaires in interoception often tap into distinct constructs rather than a common underlying dimension (Desmedt et al., 2022). Measures labeled as interoceptive sensibility, awareness, or self-reported accuracy capture partially overlapping but empirically distinct facets, with low intercorrelations between questionnaires and items. Consequently, differences in associations with psychological outcomes or personality traits may reflect differences in what each questionnaire assesses rather than true variation in interoceptive functioning.

One limitation worth noting is that we did not evaluate the IAS alongside other widely used questionnaires, such as the BPQ or MAIA. Although both are frequently employed in interoception research, they present important limitations. For example, the MAIA-2 contains subscales (e.g., Not-Worrying, Not-Distracting) with weak reliability and minimal relation to a general interoception factor (Ferentzi et al., 2021; Rogowska et al., 2023). Several other subscales primarily reflect higher-level cognitive tendencies, such as mindfulness and attentional regulation, rather than interoceptive body awareness (Todd et al., 2022; Vig et al., 2023). Similarly, the BPQ's Body Awareness and Autonomic Reactivity subscales show conceptual and psychometric concerns, and BPQ scores often overlap substantially with negative affect, raising questions about their specificity as measures of interoceptive sensibility (Campos et al., 2021; Gabriele et al., 2022; Vig et al., 2023).

Because the IAS is a newly developed questionnaire and the only one designed to assess interoceptive self-reported accuracy - a dimension previously assessed solely through objective tasks such as the HDT and HCT - our focus was on evaluating its structure and correlates without introducing established measures with documented shortcomings. Nonetheless, future research should employ novel methods, such as Exploratory Graph Analysis, to clarify the convergent and divergent validity of these questionnaires across interoceptive domains.

Nevertheless, these two studies revealed three main points. First, in the datasets that included analogue-style ratings, these scales captured greater variability, further enhancing the ability to detect individual differences. Second, a shortened, 14-item version of the IAS retained strong psychometric integrity while removing unstable or redundant items. Third, specific item pairings allowed more precise differentiation among individuals and often outperformed global scores in predictive value. Nevertheless, even this refined IAS has important limitations.

While the 14-item version improves interpretability and efficiency, it continues to underrepresent key interoceptive modalities such as cardiac, respiratory, and nuanced somatosensory signals, limiting its ability to capture the full breadth of interoceptive processes. This shortcoming underscores the need for broader coverage in future scale development. At the same time, analogue-style items and item pairings offer clear advantages: they increase response variability, allow more precise differentiation among individuals, and provide domain-specific indicators that often outperform global scores.

Building on these insights, a next-generation interoception scale could balance breadth (multiple items per modality) with depth (domain-specific subscales), retain the benefits of analogue or wider rating ranges (e.g., 1–7 instead of 1–5), and optimize coverage across physiologically and psychologically relevant bodily domains. Such a scale would improve interpretability, sensitivity, and practical utility, ultimately enabling a more nuanced mapping of interoceptive abilities onto affective, cognitive, and health-related outcomes.

Taken together, these studies advance understanding of the IAS and interoceptive self-report more broadly. The refined 14-item IAS provides a more efficient and interpretable measure, capturing domain-specific interoceptive signals that show differential associations with psychological outcomes. Associations were consistently stronger with mood measures and compared to personality and beliefs, indicating that self-reported interoceptive accuracy relates more closely to affective experience than to stable dispositional traits.

At the same time, current findings highlight limitations in modality coverage and item specificity, suggesting that future scale development should incorporate multimodal, context-sensitive, and continuous analogue ratings or wider discrete scales to better capture the richness of interoceptive processes and clarify their role in affective, cognitive, and health-related outcomes (Desmedt et al., 2022; Vlemincx et al., 2023). Ultimately, these studies show that the IAS is most informative when treated as a set of domain-specific, context-sensitive indicators rather than as a broad global index, advancing both the measurement and conceptual understanding of interoception.

References

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*.

Arslanova, I., Galvez-Pol, A., Kilner, J., Finotti, G., & Tsakiris, M. (2022). Seeing through each other's hearts: Inferring others' heart rate as a function of own heart rate perception and perceived social intelligence. *Affective Science*, *3*(4), 862–877.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166.

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27–45.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item toronto alexithymia scale—i. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32.

Banellis, L., Nikolova, N., Fischer Ehmsen, J., Courtin, A. S., Vejlo, M., Tyrer, A., Bohme, R., Bavato, F., Hoogervorst, K., Fardo, F., et al. (2025). Interoceptive ability is unrelated to mental health symptoms: Evidence from a large scale multi-domain psychophysical investigation. *medRxiv*, 2025–2008.

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.

Benau, E. M. (2023). Self-reported interoceptive accuracy and interoceptive attention differentially correspond to measures of visual attention and self-regard. *PeerJ*, *11*, e15348.

Brand, S., Meis, A. C., Tünte, M. R., Murphy, J., Woller, J. P., Jungmann, S. M., Witthöft, M., Hoehl, S., Weymar, M., Hermann, C., & Ventura-Bort, C. (2023). A multi-site German validation of the Interoceptive Accuracy Scale and its relation to psychopathological symptom burden. *Communications Psychology*, *1*(1). https://doi.org/10.1038/s44271-023-00016-x

Brand, S., Petzke, T. M., & Witthöft, M. (2022). The differential relationship between self-reported interoceptive accuracy and attention with psychopathology. *Zeitschrift* *für Klinische Psychologie Und Psychotherapie*.

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

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*, 279.

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.

Campos, C., Rocha, N. B., & Barbosa, F. (2021). *Untangling self-reported interoceptive attention and accuracy: Evidence from the european portuguese validation of the body perception questionnaire and the interoceptive accuracy scale*. http://dx.doi.org/10.31234/osf.io/a7wdj

Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique variable analysis: A network psychometrics method to detect local dependence. *Multivariate Behavioral Research*, *58*(6), 1165–1182.

Christensen, A. P., & Golino, H. (2021a). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A monte carlo simulation and tutorial. *Psych*, *3*(3), 479–500.

Christensen, A. P., & Golino, H. (2021b). On the equivalency of factor and network loadings. *Behavior Research Methods*, *53*(4), 1563–1580.

Clifton, J. D., Baker, J. D., Park, C. L., Yaden, D. B., Clifton, A. B., Terni, P., Miller, J. L., Zeng, G., Giorgi, S., Schwartz, H. A., et al. (2019). Primal world beliefs. *Psychological Assessment*, *31*(1), 82.

Clifton, J. D., & Yaden, D. B. (2021). Brief measures of the four highest-order primal world beliefs. *Psychological Assessment*, *33*(12), 1267.

Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement*, *82*(5), 880–910.

Costa, P., McCrae, R. R., & Revised, N. (1992). Personality inventory (NEO-PI-r) and NEO five-factor inventory (NEO-FFI): Professional manual. *Psychological Assessment Resources, Odessa, FL*.

Davidson, C. A., Hoffman, L., & Spaulding, W. D. (2016). Schizotypal personality questionnaire–brief revised (updated): An update of norms, factor structure, and item content in a large non-clinical young adult sample. *Psychiatry Research*, *238*, 345–355.

Desmedt, O., Heeren, A., Corneille, O., & Luminet, O. (2022). What do measures of self-report interoception measure? Insights from a systematic review, latent factor analysis, and network approach. *Biological Psychology*, *169*, 108289.

Desmedt, O., Luminet, O., Walentynowicz, M., & Corneille, O. (2023). The new measures of interoceptive accuracy: A systematic review and assessment. *Neuroscience & Biobehavioral Reviews*, *153*, 105388. https://doi.org/https://doi.org/10.1016/j.neubiorev.2023.105388

Ferentzi, E., Olaru, G., Geiger, M., Vig, L., Köteles, F., & Wilhelm, O. (2021). Examining the factor structure and validity of the multidimensional assessment of interoceptive awareness. *Journal of Personality Assessment*, *103*(5), 675–684.

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

Gaggero, G., Bizzego, A., Dellantonio, S., Pastore, L., Lim, M., & Esposito, G. (2021). Clarifying the relationship between alexithymia and subjective interoception. *PLoS One*, *16*(12), e0261126.

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.

Garner, D. M. (1991). Eating disorder inventory-2; professional manual. *Psychological Assessment Resources*.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, *2*(2), 15–34.

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035.

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292.

Haustein, M., Thomas, E. B., Scheer, K., & Denburg, N. L. (2024). Interoception, affect, and cognition in older adults. *Experimental Aging Research*, *50*(3), 279–295.

Hickman, L., Seyedsalehi, A., Cook, J. L., Bird, G., & Murphy, J. (2020). The relationship between heartbeat counting and heartbeat discrimination: A meta-analysis. *Biological Psychology*, *156*, 107949.

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., Posthuma, D., & Van Der Sluis, S. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-short). *Journal of Autism and Developmen-*

tal Disorders, *41*, 589–596.

Jahedi, S., & Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organization*, *98*, 97–114. https://doi.org/10.1016/j.jebo.2013.12.016

Jenkinson, P. M., Fotopoulou, A., Ibañez, A., & Rossell, S. (2024). Interoception in anxiety, depression, and psychosis: A review. *EClinicalMedicine*, *73*.

Jiménez, M., Abad, F. J., Garcia-Garzon, E., Golino, H., Christensen, A. P., & Garrido, L. E. (2023). Dimensionality assessment in bifactor structures with multiple general factors: A network psychometrics approach. *Psychological Methods*.

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., et al. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(6), 501–513.

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology*, *52*(11), 1432–1440.

Koike, H., & Nomura, M. (2023). Development and validation of japanese versions of the interoceptive accuracy scale and interoceptive attention scale. *SAGE Open*, *13*(4), 21582440231214639.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, *64*(2), 258–266.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ–4. *Psychosomatics*, *50*(6), 613–621.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–325.

Lin, X.-X., Shen, H.-R., Lin, J.-X., Zhang, Y.-H., Murphy, J., Wang, Y.-Z., Sun, Y.-B., Wang, N., Wang, J.-Y., Wei, G.-X., & Luo, F. (2023). Psychometric validation and refinement of the Chinese Interoceptive Accuracy Scale (IAS) in general population and patients with chronic pain. *Journal of Psychosomatic Research*, *175*, 111541. https://doi.org/10.1016/j.jpsychores.2023.111541

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).

Makowski, D., Pham, T., Lau, Z. J., Raine, A., & Chen, S. A. (2023). The structure of deception: Validation of the lying profile questionnaire. *Current Psychology*, *42*(5), 4001–4016.

Markos, A., & Tsigilis, N. (2024). Dimensionality assessment in ordinal data: A comparison between parallel analysis and exploratory graph analysis. *Frontiers in Psychology*, *15*, 1359111.

McGillycuddy, M., Warton, D. I., Popovic, G., & Bolker, B. M. (2025). Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software*, *112*(1), 1–19. https://doi.org/10.18637/jss.v112.i01

Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). *PloS One*, *13*(12), e0208034.

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The Multidimensional Assessment of Interoceptive Awareness (MAIA). *PLoS ONE*, *7*(11), e48230. https://doi.org/10.1371/journal.pone.0048230

Miller, L. C., Murphy, R., & Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology*, *41*(2), 397.

Murphy, J., Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2019). 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

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, *31*(3), 274–295.

Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient-Patient-Centered Outcomes Research*, *7*(1), 23–35.

Poerio, G. L., Klabunde, M., Bird, G., & Murphy, J. (2024). Interoceptive attention and mood in daily life: An experience-sampling study. *Philosophical Transactions of the Royal Society B*, *379*(1908), 20230256.

Porges, S. (1993). Body perception questionnaire. *Laboratory of Developmental Assessment, University of Maryland*, *10*, s15327752jpa5304_1.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, *41*(1), 203–212.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137.

Rogowska, A. M., Tataruch, R., & Klimowska, K. (2023). Validation of the shortened 24-item multidimensional assessment of interoceptive awareness, version 2 (brief MAIA-2). *Scientific Reports*, *13*(1), 21270.

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, *18*(4), 483–488.

Shields, S. A., Mallory, M. E., & Simon, A. (1989). The body awareness questionnaire: Reliability and validity. *Journal of Personality Assessment*, *53*(4), 802–815.

Sibley, C. G., Luyten, N., Purnomo, M., Mobberley, A., Wootton, L. W., Hammond, M. D., Sengupta, N., Perry, R., West-Newman, T., Wilson, M. S., et al. (2011). The mini-IPIP6: Validation and extension of a short measure of the big-six factors of personality in new zealand. *New Zealand Journal of Psychology*, *40*(3).

Simkute, D., Wang, G. Y., & Griskova-Bulanova, I. (2025). Interoception, personality, and internet use: Preliminary insights into their association. *Plos One*, *20*(7), e0328260.

Spielberger, C. D. (1970). Manual for the state-trait anxiety inventory (self-evaluation questionnaire). *(No Title)*.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Suksasilp, C., & Garfinkel, S. N. (2022). Towards a comprehensive assessment of interoception in a multidimensional framework. *Biological Psychology*, *168*, 108262.

Thimm, J. C., Jordan, S., & Bach, B. (2016). The personality inventory for DSM-5 short form (PID-5-SF): Psychometric properties and association with big five traits and pathological beliefs in a norwegian population. *BMC Psychology*, *4*, 1–11.

Todd, J., Swami, V., Aspell, J. E., Furnham, A., Horne, G., & Stieger, S. (2022). Are some interoceptive sensibility components more central than others? Using item pool visualisation to understand the psychometric representation of interoception. *Plos One*, *17*(12), e0277894.

Vig, L., Ferentzi, E., & Köteles, F. (2023). Self-reported interoception, worries and protective behaviors during the COVID-19 pandemic: A longitudinal study. *Psicologia: Reflexão e Crítica*, *36*(1), 23.

Vlemincx, E., Walentynowicz, M., Zamariola, G., Van Oudenhove, L., & Luminet, O. (2023). A novel self-report scale of interoception: The three-domain interoceptive sensations questionnaire (THISQ). *Psychology & Health*, *38*(9), 1234–1253.

Von Mohr, M., Silva, P. C., Vagnoni, E., Bracher, A., Bertoni, T., Serino, A., Banissy, M. J., Jenkinson, P. M., & Fotopoulou, A. (2023). My social comfort zone: Attachment anxiety shapes peripersonal and interpersonal space. *Iscience*, *26*(2).

Vorst, H. C., & Bermond, B. (2001). Validity and reliability of the bermond–vorst alexithymia questionnaire. *Personality and Individual Differences*, *30*(3), 413–434.

Zanarini, M. C. (2003). Zanarini rating scale for borderline personality disorder (ZAN-BPD): A continuous measure of DSM-IV borderline psychopathology. *Journal of Personality Disorders*, *17*(3), 233–242.

Zsido, A. N., Teleki, S. A., Csokasi, K., Rozsa, S., & Bandi, S. A. (2020). Development of the short version of the spielberger state—trait anxiety inventory. *Psychiatry Research*, *291*, 113223.