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, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, & others, 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, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, & others, 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

---

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

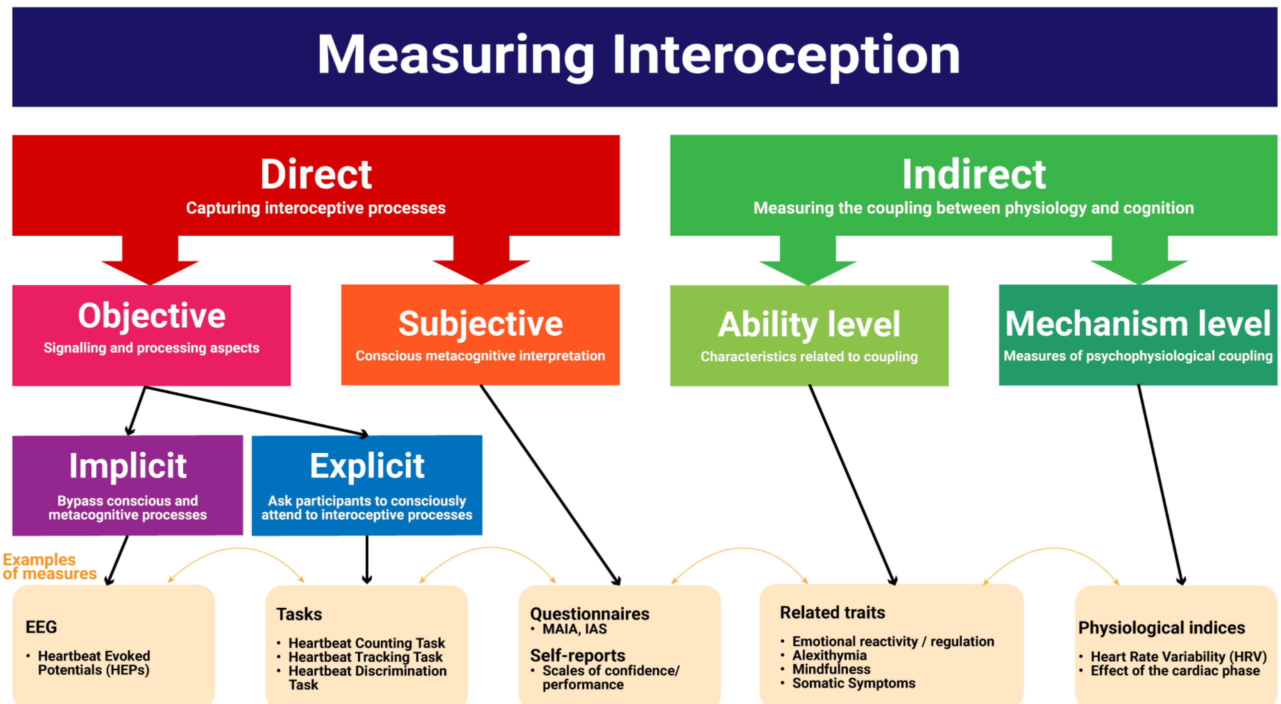
This preprint is a non-peer-reviewed work from the **Reality Bending Lab**.

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: Writing – review & editing; Raquel Nogueira Arjona: Writing – review & editing; Giulia Poerio: Writing – review & editing; 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).



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 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, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, & others, 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, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, and others (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; Gaggero et al., 2021; Koike & Nomura, 2023), while others find small negative correlations (Lin et al., 2023), a quadratic positve relationship (Campos et al., 2021) or no correlation at all (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; Petzke et al., 2024; 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 17 datasets was included (see **Table 1**).

The total number of participants was 33,526 participants (Mean = 47.96 ± 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 Anal-

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	Petzke et al., (2024)	German	254		31.5 ± 10.7	22-69	68.5%	osf.io/seru4
Sample 18	Total		33526		47.96 ± 13.1	17-93	71.3%	

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

ysis (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 strict assumptions about data distribution or latent variables.

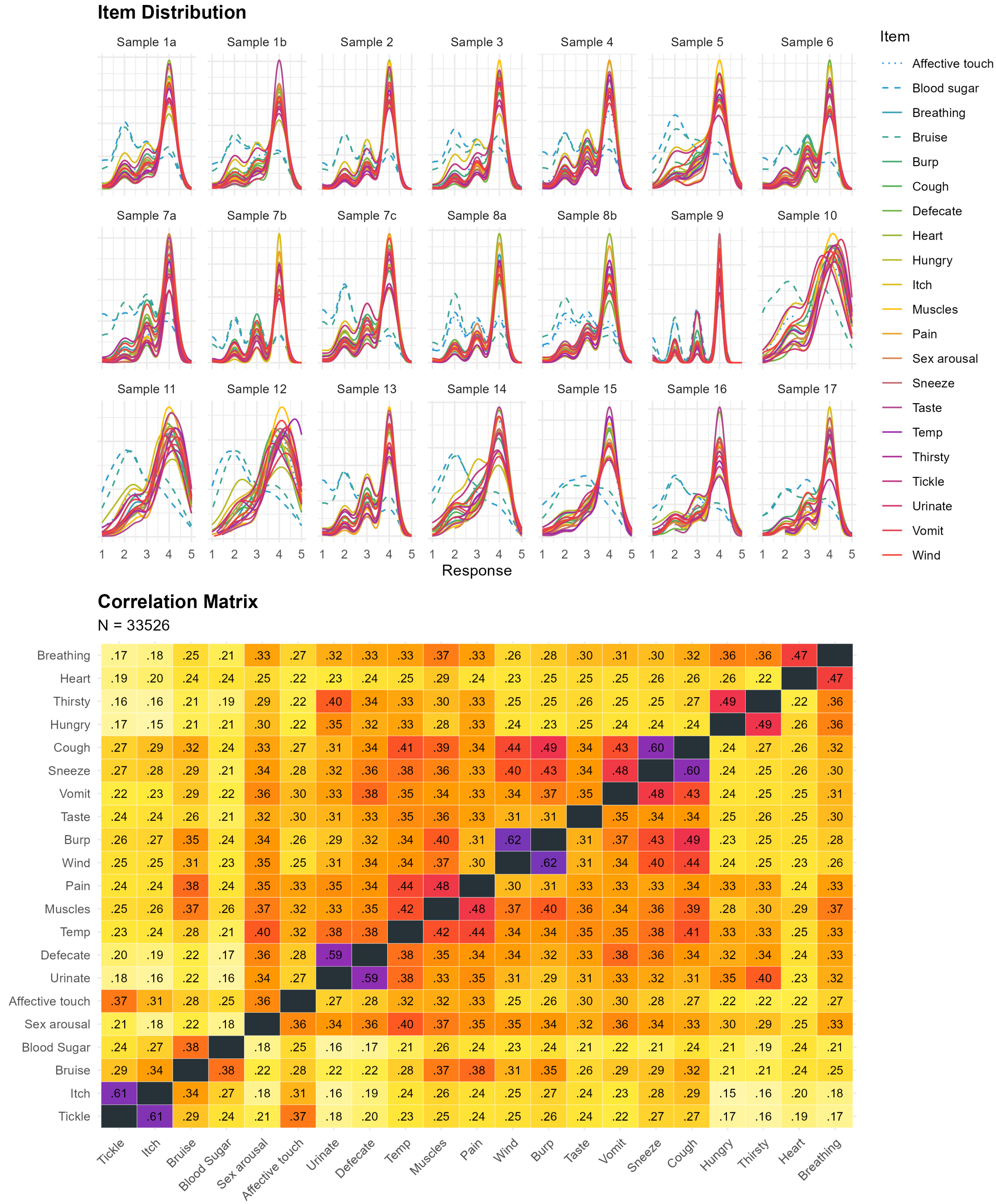
In a typical 2-step fashion, this first analysis run will be followed by a structure refinement with a further selection of items, and the final pool of items will be tested again. Additionally, various solutions (e.g., adding general factors) will be statistically compared using Confirmatory Factor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

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).



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 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 (0.035), lowest χ^2 (2,334.112), highest CFI (0.984). It was followed by the EGA + general factor model (RMSEA = 0.054, χ^2 = 6,876.395, CFI = 0.952), and the HCA model with 5 factors (RMSEA = 0.078, χ^2 = 13,441.878, CFI = 0.906). 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 (≈ 0.025–0.057), 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.

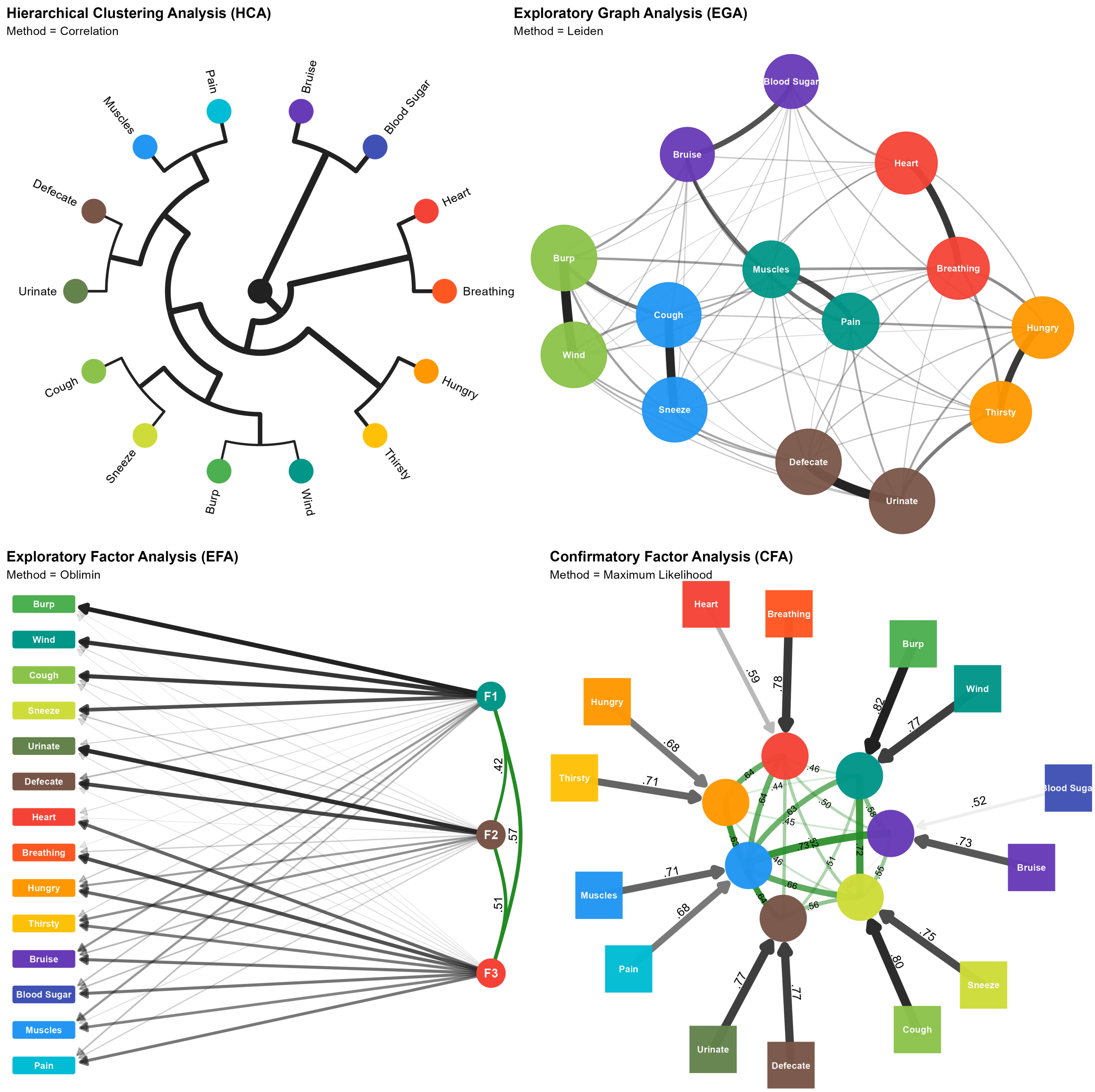
Analyses revealed that seven item pairs (e.g., Hunger–Thirst, Urinate–Defecate, Cough–Sneeze) provided a robust and replicable structure, leading to a refined 14-item IAS. These pairs consistently emerged across datasets and yielded superior fit indices compared to alternative models. This suggests that 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.

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.



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 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. Interoception-related measures included the Multidimensional Assessment of Interoceptive Awareness Version-2 (MAIA-2, Mehling et al., 2018), the Body Perception Questionnaire Short Form (BPQ-SF, Cabrera et al., 2018), and the Interoceptive Confusion Questionnaire (ICQ, Brewer et al., 2016). We also included the Toronto Alexithymia Scale (TAS-20, Bagby et al., 1994), and the Bermond–Vorst Alexithymia Questionnaire (BVAQ, Vorst & Bermond, 2001), due to the close relationship between interoception deficits and alexithymia.

Mood and Anxiety Related. Mood-related measures included the Beck Depression Inventory-II (BDI-II, Beck et al., 1996), the Patient Health Questionnaire-4 (PHQ-4, Kroenke et al., 2009), the Patient Health Questionnaire-9 (PHQ-9, Kroenke et al., 2001), the Patient Health Questionnaire-15 (PHQ-15, Kroenke et al., 2002), the Short Mood and Feelings Questionnaire (SMFQ, Angold et al., 1995), the State–Trait Anxiety Inventory Trait Version (STAI-T, Spielberger, 1970), and the Generalized Anxiety Disorder Scale (GAD-7, Spitzer et al., 2006). We also included the PHQ-2 (Kroenke et al., 2003) and General Anxiety Disorder-2 (GAD-2, Kroenke et al., 2007), which were combined with the PHQ-4 scores, and scores from the shorter version of the STAIT-5 (Zsido et al., 2020), which were combined with the STAI-T. Mood measures were included given the well-established association between interoceptive deficits and mental health conditions (Khalsa, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, Meuret, et al., 2018).

Psychopathology. Personality pathology-related measures included the Personality Inventory for DSM-5 Short Form (PID-5-SF, Thimm et al., 2016), the Schizotypal Personality Questionnaire – Brief Revised (SPQ-BR, Davidson et al., 2016), the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD, Zanarini, 2003), and the Autism Spectrum Quotient Short Form (ASQ-S, Hoekstra et al., 2011). Similarily, these measures were included due to established associations between various psychopathologies and interoceptive disruptions (Leão et al., 2025).

Personality. Personality-related measures included the NEO Five-Factor Inventory, Neuroticism subscale (NEO-FFI, Costa et al., 1992), the Mini International Personality Item Pool (MINI-IPIP6, Sibley et al., 2011), and the Big Five Inventory Short Form (BFI-S, Rammstedt & John, 2007). Although associations between personality and interoception have been less frequently explored than those involving mood or psychopathology, previous research has identified links between neuroticism and extroversion and interoceptive sensibility using the MAIA and BPQ measures. This study therefore included extended personality measures to clarify the broader relationships between interoception and personality traits beyond neuroticism and extraversion (Pearson & Pfeifer, 2022).

Beliefs and Misbeliefs. Beliefs- and misbeliefs-related measures included the Generic Conspiracist Beliefs Scale (GCBS, Brotherton et al., 2013), the Primal World Beliefs Inventory (PI-99, Clifton et al., 2019) and its short form (PI-18, Clifton & Yaden, 2021), and the Lie Scale (LIE, Makowski et al., 2023). These instruments assess individual differences in conspiracist, world, and lying beliefs. This domain represents a novel avenue of research in the context of interoception, as such beliefs have not been previously assessed in relation to interoceptive processes.

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 β = .13), with all effects significant. Gender displayed near-zero effects (mean β = −.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 β = .26), followed by Attention Regulation (mean β = .25) and related subscales - Body Listening, Trusting, Emotional Awareness and Self-Regulation, each in the .19–.24 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 and significantly only for the full IAS, Hungry-Thirsty, Bruise-Blood sugar and the Muscles-Pain pairings (β ≈ −.02 to .06). Not Worrying was significant only for the full IAS and the Hungry–Thirsty and Urinate–Defecate pairings (β ≈ −.01 to .04).

Autonomic Reactivity (BPQ; mean β = −.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 β ≈ .15; range = .10 to .22) was positive and significant across most pairings with the exception of Wind-Burp and Urinate-Defecate.

Interoceptive deficits showed consistent negative standardized effects. Alexithymia measures (TAS, BVAQ) were robustly negative (β ≈ −.13 to −.2), except for the BVAQ Affective subscale, which was positive for Bruise–Blood (β = .16). Interoceptive confusion (ICQ) was also negative across all IAS measures (mean β ≈ −.32; range = −.49 to −.20).

Depression (BDI; mean β = −.20) and anxiety (STAI-T; mean β = −.20) showed consistent negative associations across the IAS and its pairings, with the strongest effects observed for the Hungry–Thirsty dimension. Related constructs showed some divergence: the MFQ was largely nonsignificant except for a notable effect in the Breathing–Heart pairing, while somatic concerns (PHQ-15) were unrelated to the total scale but showed weak negative associations with the Hungry–Thirsty, Bruise–Blood sugar, and Urinate–Defecate pairings (β ≈ −.22 to −.05). Other mood indices (PHQ-9, PHQ-4, GAD-7) produced generally weaker but often reliable negative coefficients (typically −.10 to −.15), though effects sometimes appeared only for specific pairings (e.g., PHQ-4 nonsignificant for Wind–Burp and Breathing–Heart). Across measures, the Hungry–Thirsty pairing consistently yielded the strongest standardized effects.

Autistic traits (ASQ) were generally negatively associated with IAS scores. Among ASQ subscales, Imagination (mean β = −.18) exhibited the most consistent pattern of negative associations across pairings, with the exception of Cough–Sneeze. The strongest full-scale effects were driven by the Hungry–Thirsty, Bruise–Blood sugar, and Muscles–Pain pairings, which yielded the largest negative coefficients. In contrast, associations weakened or became nonsignificant for other pairings.

A comparable trend was observed for maladaptive personality traits. Detachment, Psychoticism, and Negative Affect demonstrated the most reliable, albeit weak, negative associations with overall IAS scores (mean β range = −.08 to −.10), again primarily driven by the Hungry–Thirsty, Bruise–Blood sugar, and Muscles–Pain pairings. Other maladaptive traits were unrelated to IAS scores. Notably, Antagonism was the only trait to show a significant positive association, limited to the Bruise–Blood sugar pairing (β = .14), whereas Borderline Personality was uniquely associated with Hungry–Thirsty (mean β = −.14).

Schizotypal traits showed a similarly negative pattern of associations. Social Anxiety (mean β = −.11) and No Close Friends (mean β = −.12) displayed the most consistent effects. Although some facets were unrelated to the overall IAS score, they were associated with specific pairings: Unusual Perceptions with Muscles–Pain and Urinate–Defecate, and Eccentric with Hungry–Thirsty and Urinate–Defecate.

Among the Big Five traits, Openness (mean β = .10), Agreeableness (mean β = .08), and Conscientiousness (mean β = .11) exhibited small yet consistent positive effects, most evident for the full IAS. Extraversion showed a weak but positive association, driven primarily by its stronger relation to the Bruise–Blood pairing (β = .17). In contrast, Honesty–Humility (mean β = −.10) demonstrated a small negative effect, most evident for the Hungry–Thirsty and Bruise–Blood sugar pairings. Finally, Neuroticism (mean β = −.23) was negatively associated with the IAS, with the strongest effect observed for the Hungry–Thirsty pairing (β = −.22).

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 ≈ .08–.11), 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 three pairings (mean β = .10), Personal Wellbeing with the full scale and the Hungry-Thirsty (mean β = .09), and Information Control (β = .06) with Cough–Sneeze.

Overall, the full IAS produced the greatest number of significant standardized effects (28), 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; Todd et al., 2022). Associations with the BPQ Body Awareness subscale were consistent with previous reports on positive association between these measures (Brand et al., 2023; Campos et al., 2021; Gaggero et al., 2021; Koike & Nomura, 2023), 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, 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 associations 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 somatic or respiratory reflexes (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: Extraversion showed its strongest associations for Bruise-Blood sugar, Neuroticism and Honesty–Humility for Hungry–Thirsty, and Openess, Conscientiousness and Agreeableness 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 associations 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 14-item IAS performed similarly to the original 21-item version, showing comparable patterns of association with mood, psychopathology, and interoceptive measures. Overall, standardized effects between the IAS and external constructs were not uniform but varied across bodily domains. Visceral item pairs, particularly Hungry–Thirsty and Bruise–Blood sugar, showed the strongest and most consistent relationships with interoceptive, affective, and personality measures, whereas other pairs contributed little unique information, particularily in measures of psychopathologies and beliefs. These findings suggest that the shorter scale captures domain-specific patterns that were not evident in the full scale, offering a more efficient and nuanced measure of self-reported interoceptive accuracy, though further validation in independent samples is needed.

General Discussion

The present mega-analysis 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 conceptual interpretation, and practical applications of the IAS. In this discussion, these findings are contextualized within the broader landscape of interoception research.

Study 1 re-analysed the IAS using complementary psychometric and network-based approaches. Across analyses, seven stable item pairings emerged (Hungry–Thirsty, Urinate–Defecate, Muscles–Pain, Breathing–Heart, Cough–Sneeze, Wind–Burp, and Bruise–Blood sugar), yielding a refined 14-item structure with superior fit and cross-dataset stability compared to previously proposed unidimensional or bifactor models. These results challenge interpretations of the IAS as a single latent construct (e.g., Brand et al., 2023; Murphy et al., 2019) and instead indicate that it captures multiple, interrelated but domain-specific clusters of bodily sensations.

Unlike Lin et al. (2023), whose 12-item solution supported unidimensionality, the present findings suggest that local dependencies between items reflect meaningful theoretical coupling among functionally related sensations (e.g., Heart–Breathing). Such clustering likely mirrors the way bodily signals co-occur physiologically or phenomenologically (e.., hunger with thirst, coughing with sneezing) and are therefore perceived and evaluated together. Perceived accuracy in one bodily system may generalize to conceptually or functionally linked sensations, yielding a clustered rather than continuous organization of interoceptive beliefs. Consistent with this view, self-reported interoceptive awareness across related axes (e.g., cardiac and respiratory) tends to show positive associations (Garfinkel et al., 2016), whereas objective measures of interoceptive accuracy across modalities typically show weak or absent correlations (Bruni, 2023; Ferentzi et al., 2018). This dissociation suggests that subjective interoceptive beliefs are structured according to functional bodily domains rather than reflecting a single general capacity.

Study 2 extended these findings by examining the IAS's associations with interoceptive, affective, personality, and belief-related traits. Consistent with prior work, the IAS correlated positively with adaptive aspects of interoceptive sensibility (e.g., MAIA subscales reflecting noticing, attention regulation, and body trust) and negatively with interoceptive difficulties, namely alexithymia and interoceptive confusion (e.g., Brand et al., 2023; Gaggero et al., 2021; Garfinkel et al., 2015).

Across dispositional constructs, individuals reporting higher symptoms of depression and anxiety tended to perceive themselves as less accurate in detecting bodily signals, in line with previous findings (e.g., Brand et al., 2023; Khalsa, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, Meuret, et al., 2018; Nord & Garfinkel, 2022). Crucially, these associations were not uniform across bodily domains: the strongest effects involved visceral sensations (e.g., hunger, thirst), whereas cardiorespiratory and expulsion domains showed weaker or inconsistent relationships. This domain specificity suggests that beliefs about one's visceral accuracy may be more psychologically and clinically relevant than beliefs about reflexive or surface sensations.

This pattern may help clarify inconsistencies in the literature linking interoception and emotion. Reviews and meta-analyses report that performance-based measures of interoceptive accuracy, particularly in cardioception, often show weak or absent correlations with depression and anxiety (Adams et al., 2022; Banellis et al., 2025; Jenkinson et al., 2024). In contrast, self-reported interoceptive measures typically show stronger, often negative, associations with internalizing traits (e.g., Brand et al., 2022; Clemente et al., 2024; Lin et al., 2023). These discrepancies likely reflect both conceptual distinctions between perceptual accuracy and metacognitive self-belief (Garfinkel et al., 2015; Khalsa, Adolphs, Cameron, Critchley, Davenport, Feinstein, Feusner, Garfinkel, Lane, Mehling, Meuret, et al., 2018; Suksasilp & Garfinkel, 2022) and variation in the bodily domains assessed. The present results extend this framework by demonstrating that associations between interoceptive self-beliefs and affective traits are concentrated in psychologically salient visceral domains. Thus, the field's heavy reliance on cardioceptive tasks may have obscured more meaningful relationships in other interoceptive modalities.

Beyond mood, similar though weaker patterns emerged for autistic traits, schizotypy, and maladaptive personality dimensions. These traits tended to correlate negatively with self-perceived interoceptive accuracy, particularly within visceral and musculoskeletal domains, while showing little or no association in other areas. This selective pattern suggests that diminished confidence in bodily perception may contribute to specific social–emotional and personality features, rather than reflecting a general interoceptive deficit. These results align with evidence that interoceptive self-beliefs, more than objective accuracy, track variation in emotional awareness, social connectedness, and psychopathology (Brand et al., 2022; **?**; **?**).

Taken together, these findings suggest that self-perceived interoceptive accuracy is a differentiated construct, varying systematically across bodily systems and psychological traits. Beliefs about visceral sensations, in particular, show the most consistent associations with affective and dispositional features. Although modest in magnitude, these effects underscore the value of considering domain-specificity when linking interoceptive beliefs to emotion and mental health.

Methodologically, combining traditional psychometrics with network-based approaches proved valuable. EGA complemented factor analysis by identifying meaningful inter-item dependencies, producing a more interpretable and stable structure. The refined 14-item IAS balances parsimony and predictive validity, illustrating how network-informed scale refinement, beyond scale development (H. F. Golino & Epskamp, 2017), can enhance construct clarity.

The IAS nonetheless remains limited in scope. Certain interoceptive modalities (e.g., gastric, tactile) are underrepresented, and items lack contextual specification (e.g., physiological vs. affective triggers). For instance, a racing heart could signal anxiety or exercise; without context, it is unclear which state participants referenced from. Such ambiguities may constrain ecological validity and contribute to variability across studies (Vlemincx et al., 2023). Expanding item content and adopting context-specific phrasing will be important for future measurement development.

Clinically, the domain-specific structure of the IAS highlights that interoceptive disruptions in affective and psychopathological conditions may center primarily on beliefs about visceral sensations rather than global interoceptive deficits. This perspective suggests new routes for targeted assessment and intervention focusing on specific bodily domains.

Nonetheless, the current work is not without its limiations. This paper did not evaluate the IAS alongside other questionnaires such as the BPQ or MAIA, whose conceptual overlap and psychometric inconsistencies complicate interpretation (Campos et al., 2021; Ferentzi et al., 2021; Rogowska et al., 2023; Vig et al., 2023). Because the IAS uniquely targets interoceptive self-reported accuracy, a dimension previously assessed only via performance tasks such as the HDT and HCT, our focus was on its structure and correlates rather than introducing established measures with documented shortcomings. Future work should employ contemporary analytic methods, such as EGA and network modeling, to evaluate convergent and discriminant validity across interoceptive domains and clarify how self-belief and accuracy measures jointly shape interoceptive functioning.

References

Adams, K. L., Edwards, A., Peart, C., Ellett, L., Mendes, I., Bird, G., & Murphy, J. (2022). The association between anxiety and cardiac interoceptive accuracy: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *140*, 104754.

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.

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*.

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.

Bruni, P. (2023). Interoception across domains: Assessing the role of age. *Innovation in Aging*, *7*, 755.

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.

Clemente, R., Murphy, A., & Murphy, J. (2024). The relationship between self-reported interoception and anxiety: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *167*, 105923.

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., Drew, R., Tihanyi, B. T., & Köteles, F. (2018). Interoceptive accuracy and body awareness–temporal and longitudinal associations in a non-clinical sample. *Physiology & Behavior*, *184*, 100–107.

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., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160014.

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.

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 Developmental 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.

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

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.

Leão, L. M. M., Rêgo, A. C. M. do, & Araújo-Filho, I. (2025). Interoception: Current knowledge gaps and future directions in neuroscience, psychopathology, and clinical applications. *International Journal of Innovative Research in Medical Science (IJIRMS)*, *10*(03).

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.

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.

Nord, C. L., & Garfinkel, S. N. (2022). Interoceptive pathways to understand and treat mental health conditions. *Trends in Cognitive Sciences*, *26*(6), 499–513.

Pearson, A., & Pfeifer, G. (2022). Two measures of interoceptive sensibility and the relationship with introversion and neuroticism in an adult population. *Psychological Reports*, *125*(1), 565–587.

Petzke, T. M., Köteles, F., Pohl, A., & Witthöft, M. (2024). Somatic symptom distress is not related to cardioceptive accuracy. *Journal of Psychosomatic Research*, *181*, 111655.

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).

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 multi-dimensional 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.