Daily Gas Burden Scale (DGBS)

(DGBS, V1.0): Conceptual Proposal and Rationale for a Unidimensional Scale

Author: Keskitalo, J.A.

Affiliation: Kakamit, Revonlahti, Suomi, Finland

Original Language: Finnish

Abstract

Intestinal gas symptoms are highly prevalent and associated with poorer quality of life at the population level. Several multi-item measures have been developed in clinical research (e.g., IGQ, PROMIS GI, GSRS), which are psychometrically robust but lengthy for repeated digital monitoring in daily life. The goal of this document is to describe the Daily Gas Burden Scale (DGBS), a single-item, five-point Likert scale that measures subjective gas burden over a 24-hour period. We present the rationale for why the use of separate, parallel scales (e.g., frequency and severity) is challenging from the perspectives of psychometrics and usability, and we outline a realistic validation pathway (cognitive validation → test–retest reliability → convergent validity). The DGBS is not a diagnostic test, but a standardized self-assessment for long-term monitoring and research use.

Key References: IGQ development and validation; PROMIS GI scales; GSRS validation; SHS-GI; DQLQ. [1–6]

---

1. Background

Gas symptoms (e.g., flatulence, abdominal bloating, gas accumulation) are extremely common and impair perceived well-being. According to multinational population research, the majority of adults report gas symptoms during a 24-hour period; higher symptom scores correlate with lower quality of life and higher psychological distress [7]. Existing measures—the Intestinal Gas Questionnaire (**IGQ**), **PROMIS GI**, and **GSRS**—demonstrate that constructs related to gas and other GI symptoms can be reliably measured through a systematic development and validation process (qualitative interviews, cognitive debriefing, factor analyses, reliability, and known-group validity) [1–4]. At the same time, these instruments are typically multi-item and vary in their recall period (24 h symptom diary + 7-day impact section in the IGQ) [1,3], which can complicate daily mobile monitoring.

---

2. Why is the Use of Separate Scales (e.g., "Frequency" + "Severity") Challenging?

2.1 Cognitive Load and Usability in Daily Monitoring

Multi-item and parallel assessments increase the respondent's cognitive load and reduce engagement in daily life, especially if the measure is completed daily. In the development of the IGQ, the 43-item pilot version had to be significantly streamlined due to factors such as

floor effects, **cross-loadings**, and **inter-item correlations** to form a statistically clearer structure [3]. Translation and validation work for the PROMIS GI scales show that clarifying the language versions of already standardized multi-item scales requires cognitive debriefing and subtle phrasing to ensure consistent interpretation—this is emphasized when the respondent is asked to distinguish between closely related concepts (such as "frequency" vs. "severity") [4].

2.2 Measurement Error: Halo Effect and Artificial Common Variance

When multiple closely experienced dimensions are inquired about in the same session, the respondent may give similar scores without sufficient independent differentiation (the **halo effect**), which increases common variance between items without genuinely measuring two separate constructs. In the IGQ validation, a need to remove and rephrase items was observed precisely because some had weak or unclear loadings in the factor structure [3]. A similar observation—that a shorter, clearer measure can yield good psychometrics—has also been reported for the **SHS-GI** scale in population data (CFA fits good; Cronbach α ≈ 0.80) [5].

2.3 Ambivalence in Interpretation and Construct Condensation

In gas symptoms, "frequency" and "severity" are not necessarily orthogonal; they may reflect the same perceived "**burden**" in different ways. This leads to interpretation problems in composite measures: different symptom profiles can result in the same score, even though the clinical significance differs. Psychometric literature on GI measures suggests that a clear overall construct (e.g., "gas burden" / "gas & bloating" dimension) is better operationalized and validated than artificially separated, highly correlated sub-dimensions [1–4].

2.4 Lack of Subjective Anchors for "Severity"

Unlike, for example, the **BSFS** scale for stool consistency (1–7 visual categories), there is no established physiological or visual anchor for the "**severity of a single gas episode**." This makes it susceptible to mediating variables (culture, social situations, shame, daily activities) and weakens reproducibility and comparability across populations [2,4]. Furthermore, gas symptoms are dependent on diet and fermentable carbohydrates (FODMAPs), which changes the daily experience and makes it difficult to distinguish constant assessments of "frequency" and "severity" [8,9].

Summary: The psychometric and usability evidence from GI measures clearly supports a well-defined main construct and the use of a carefully anchored, short measure for repeated monitoring [1–6].

---

3. DGBS Compared to Key Validated GI Symptom Measures

Feature	DGBS (v1.0)	IGQ (2021)	PROMIS GI Gas & Bloating	GSRS	SHS-GI (2021)	DQLQ
Measure Purpose	Daily gas burden monitoring	Comprehensive gas symptom assessment	Broad GI symptom mapping	General GI symptom measure	Brief GI distress assessment	GI symptom QOL
Number of Items	1	17	4–6	15	4	12
Recall Period	24 hours	24 h + 7 days	7 days	7 days	7 days	2 days
Psychometric Properties	Not yet validated (plan ready)	$\alpha = 0.85–0.92, ICC = 0.75–0.90$	IRT/Rasch, $\theta > 0.90$	$\alpha = 0.74–0.89$	$\alpha \approx 0.80, ICC \approx 0.85$	$\alpha = 0.93$
MCID Defined	Yes (theoretical: 1 point)	Yes (3–5 points)	Yes	Yes	Yes (≈ 1 point)	Yes
Daily Repeated Use	Primarily designed for this	Not suitable	Not suitable	Not suitable	Possible	Not suitable

Mobile & Digital Monitoring	Excellent	Poor	Moderate	Poor	Good	Moderate
Gas Symptom Specificity	Very High (gas burden only)	Very High (6 dimensions)	High	Low	Low	Low
Cognitive Load	Minimal	High	Moderate–High	High	Low	Moderate
Cultural & Language Versions	AI-Translated (Not Cognitively Validated)	Several	Extensive (PROMIS)	Several	Several	Several

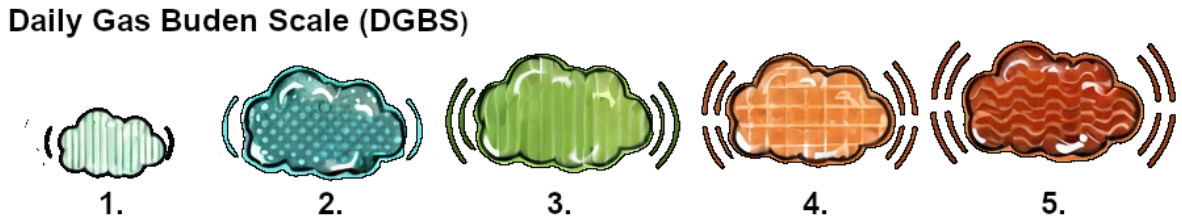
Conclusion from the Table: The DGBS clearly fills an existing gap. None of the current validated measures combine the following features:

- **Ultrashort** (1 item)
- **Gas-specific** (measures gas burden only)
- **24-hour recall period** (allows for precise daily monitoring)
- **Designed for daily mobile monitoring** (minimal cognitive load)

The DGBS complements existing measures—it does not compete with them, but offers a low-burden digital monitoring tool.

---

4. DGBS (V1.0): Scale Definition



Question (24 h recall):

"How much have your intestinal gas symptoms (flatulence, bloating, gas accumulation) disturbed or burdened you during the last 24 hours?"

Likert Anchors (1–5):

1 – No burden

2 – Slight burden

3 – Moderate burden

4 – Significant burden (impaired activity)

5 – Very high burden (clearly impairing life/social situations)

Likert Anchors 1–5 (Refined Version):

1 – No gas symptom burden

No flatulence or associated discomfort. No impact on daily activity.

2 – Slight burden

You noticed occasional flatulence or slight discomfort, but it did not interfere with activity or social situations.

3 – Moderate burden

Gas symptoms were clearly noticeable and occasionally bothersome, but did not prevent usual activities.

4 – Significant burden

Gas symptoms interfered with activity, caused discomfort, or required adjustment of the day's plans or behavior.

5 – Very high burden

Gas symptoms were constantly bothersome and clearly complicated life or social situations (e.g., you avoided situations, interrupted activity, or experienced intense discomfort).

Rationale: One clear, verbally anchored item condenses the construct and is suitable for daily mobile monitoring, analogous to how brief and clearly focused GI scales have yielded good psychometric indices in various populations [3–6].

---

5. MCID for the DGBS Scale: Computational, Theoretical Model

This section includes a new addition: a computational, statistically consistent, and feasible **MCID** model for the DGBS. The model is based on the MCID literature, scale structure, and distribution-based assessment of two validated GI measures (IGQ and SHS-GI). It does not require new data and serves as a theoretical starting value for the pre-validation stage.

5.1 Initial Assumptions

- **IGQ:** MCID typically $\sim$ 0.3–0.5$ SD
- **SHS-GI:** MCID ≈ 1 point (7-point scale $\approx 14\%$ of scale width)
- **DGBS:** 5-point ordinal scale, one point change $= 25\%$ of the scale

5.2 Distribution-Based MCID

The DGBS is a 1-item Likert (1–5). Based on research literature, the SD for 1-item GI scales is typically 0.7–1.0.

A conservative estimate is chosen: $SD \approx 0.85$.

$MCID = 0.5 \times SD \approx 0.425 \to$ rounds to 1 point.

5.3 Anchor-Based MCID to the IGQ Measure

- IGQ's MCID: 3–5 points ($\approx 8–12\%$ of scale width)
- DGBS scale width $1–5 \to$ width $= 4$
- If IGQ MCID $\approx 10\%$ of the scale $\to$ DGBS MCID $= 0.10 \times 4 = 0.4 \to$ rounds to **1 point**.

5.4 Anchor-Based MCID to the SHS-GI Measure

- SHS-GI MCID $\approx 1/7 \approx 14\%$ of the scale
- DGBS: $0.14 \times 4 = 0.56 \to$ rounds to **1 point**.

5.5 Hybrid Model (Distribution + Anchor)

All three methods converge to the same value:

- ⭐ **MCID = 1 point**
- ⭐ **SCMID** (Substantial Clinically Meaningful Change) **= 2 points**

Clinical Interpretation:

- -1 = Significant improvement
- -2 = Large improvement
- 0 = No change
- +1 = Significant worsening

5.6 Validation Plan for MCID Estimate

Once data is collected, the MCID can be validated using:

- **ROC curve / Youden index** (anchor change, e.g., PGIC)
- **SEM-based estimate** ($SD \times \sqrt{(1-ICC)}) \to$ typically $\sim 0.4–0.5 \to 1$ point
- **MIC** (Minimally Important Change) for clinical applications

---

6. Validation Pathway (Step-by-Step)

- **Cognitive Validation (N=15–20):** "Think-aloud" interviews to ensure that "gas burden" and the anchors are interpreted consistently (PROMIS GI translation and debrief methods as reference) [4].
    - 10-minute Rapid Cognitive Interview method (Appendix A)
- **Test–Retest Reliability (N≥100):** 7-day internal repeated measure; **ICC** (two-way model) as the primary reliability measure for a single item. Similar setups have produced excellent ICC levels in GI measures (e.g., DQLQ ICC≈0.89) [2].
- **Convergent Validity:** Correlation with the **BSFS** classification (daily), food diary (**FODMAP** exposure), and possible objective references H_2/CH_4
- in breath tests) [6–9,10].
- **Construct Validity (Exploratory):** If the DGBS is expanded to be multi-item (e.g., calibrating examples/visual anchors), the structure will be evaluated with factor analyses according to IGQ practice [1,3].
- **Cultural and Language Adaptation:** According to FACIT/PROMIS principles (translation, cognitive debrief, pilot) [4].

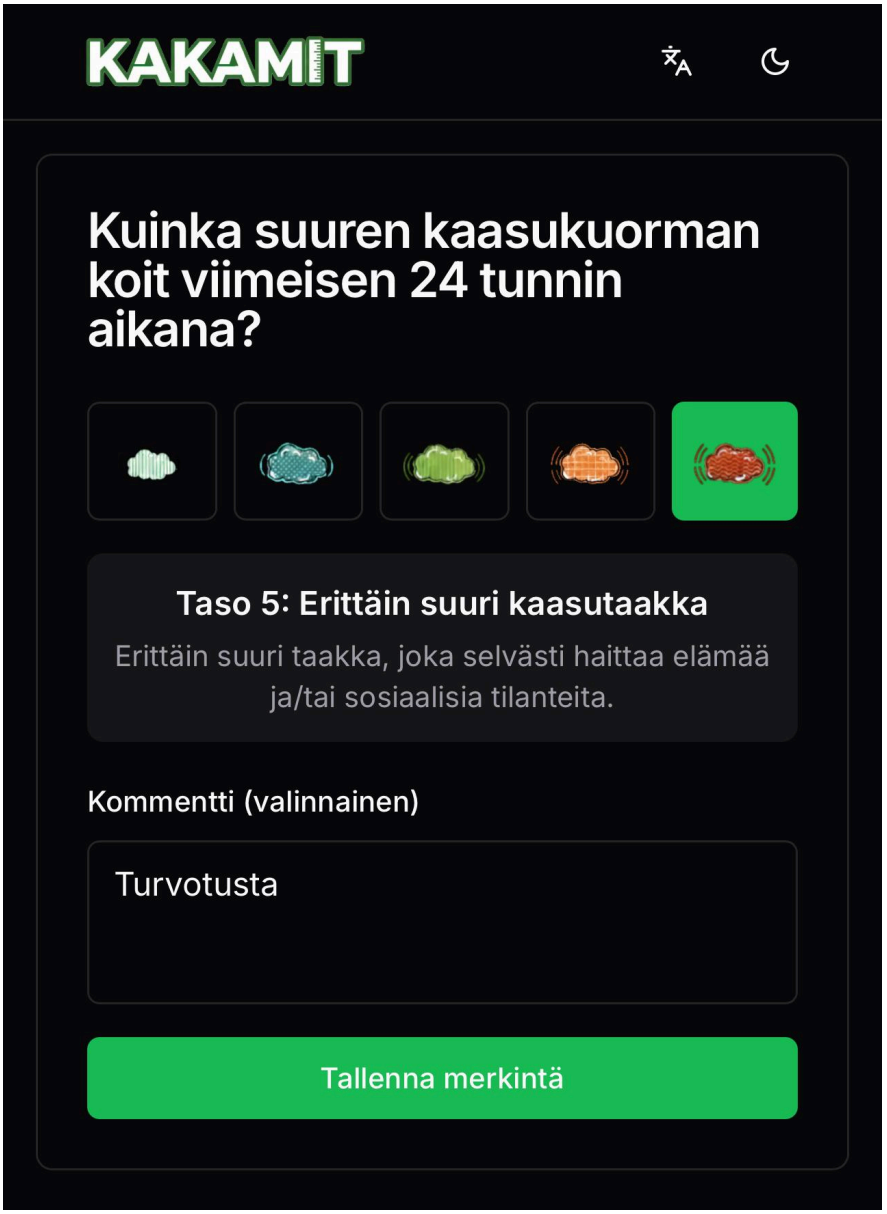
---

7. Statistical Analysis Plan

- **Reliability: ICC** (primary measure for a single item). If multiple items are added, **Cronbach's α** will be reported as a supplementary measure of internal consistency (interpretation limits ≥0.70 typically acceptable) [2,11,12].
- **Convergent Validity:** Spearman/Pearson DGBS $\sim$ BSFS, DGBS $\sim$ food diary FODMAP exposure, DGBS $\sim$ breath gases [6–10].
- **Distributions and Sensitivity:** Floor/ceiling effects; responsiveness in interventions (e.g., low-FODMAP period).
- **Suitability for Digital Monitoring:** Daily response rate and dropout rate; comparison of usability of short vs. long measures (e.g., benefits of SHS-GI-type brevity) [5].

---

8. Digital Implementation (PWA)



- **User Interface:** Single slider (1–5), visible verbal anchors; color gradient (Blue $\to$ green $\to$ red, rasterization).
- **Schedule:** Daily reminder; 24 h recall.
- **Metadata:** Time, free-text note, potential BSFS (1–7), FODMAP flags (food diary quick tags), location
- **Privacy:** IndexedDB + Native browser backup and synchronization if needed. Offline use ensures anonymity (GDPR etc.)
- **Accessibility:** ARIA tags, WCAG AA contrast, haptic feedback.
- **Refinement / Enrichment:** AI, RAG, Machine Learning, self-refining adaptation

---

9. Limitations

- **Subjectivity:** The DGBS measures perceived rather than physiological gas volume; physiological comparisons are required before clinical use [10].
- **Individual and Cultural Differences:** Requires calibration/anchor examples and language adaptation [4].
- **Confounding Variables:** Nutrition (FODMAP), menstrual cycle, medication, psychological factors—must be considered in the analysis [8,9,6].

---

10. Conclusion

The psychometric and usability evidence from GI measures supports a clear, unidimensional self-assessment when the goal is daily digital monitoring. The DGBS condenses the "perceived gas burden" into a clinically understandable and research-applicable number. After validation, the DGBS can serve as a consistent core measure that links with BSFS, nutrition, and objective measurements in large datasets.

References

Duracinsky M, Thonon F, et al. The Intestinal Gas Questionnaire (IGQ): Psychometric validation… Neurogastroenterol Motil. 2021. (Open-access PDF) https://www.researchgate.net/publication/352539820_The_Intestinal_Gas_Questionnaire_IGQ_Psychometric_validation_of_a_new_instrument_for_measuring_gas-related_symptoms_and_their_impact_on_daily_life_among_general_population_and_irritable_bowel_syndrome/fulltext/60cde525458515dc1792688c/The-Intestinal-Gas-Questionnaire-IGQ-Psychometric-validation-of-a-new-instrument-for-measuring-gas-related-symptoms-and-their-impact-on-daily-life-among-general-population-and-irritable-bowel-syndrome.pdf

Beke M, Burns AM, et al. Validation of a novel quality of life questionnaire: the DQLQ. Health Qual Life Outcomes. 2022;20:53. https://hqlo.biomedcentral.com/articles/10.1186/s12955-022-01956-4

ISPOR Presentation (PGI47). Psychometric validation of the Intestinal Gas Questionnaire (IGQ)… ISPOR Europe 2019. https://www.ispor.org/heor-resources/presentations-database/presentation/euro2019-3122/98611

van der Ende-van Loon M, Korteling D, et al. Dutch–Flemish translation and validation of the PROMIS® GI Symptom Scales. J Patient-Reported Outcomes. 2023;7:125. https://jpro.springeropen.com/articles/10.1186/s41687-023-00662-z

Walter S, Jones MP, et al. Validation of the Short Health Scale for GI symptoms (SHS-GI). Scand J Gastroenterol. 2021;56(12):1406–1413. https://www.tandfonline.com/doi/pdf/10.1080/00365521.2021.1974087

Kulich KR, Madisch A, et al. Reliability and validity of the GSRS in dyspepsia (six-country study). Health Qual Life Outcomes. 2008;6:12. https://hqlo.biomedcentral.com/articles/10.1186/1477-7525-6-12

United European Gastroenterology (UEG) – Press release. Gas-related symptoms affect nearly all adults… 2021. https://ueg.eu/a/281

NIDDK (NIH). Eating, Diet & Nutrition for Gas in the Digestive Tract. 2024 (päivitetty). https://www.niddk.nih.gov/health-information/digestive-diseases/gas-digestive-tract/eating-diet-nutrition

Judkins TC, Dennis-Wall JC, et al. Stool frequency/form and GI symptoms by menstrual cycle day. BMC Women's Health. 2020;20:136. https://bmcwomenshealth.biomedcentral.com/articles/10.1186/s12905-020-01000-x

Pitchumoni CS, Goldstein DR, Vuittonet CL. Gas, Belching, Bloating, and Flatulence: Pathogenesis, Evaluation, and Management. Springer Living Reference. 2020. https://link.springer.com/rwe/10.1007/978-3-319-90761-1_44-1

Martínez-González AE, Montoro-Pérez N, et al. Psychometric Properties of the Gastrointestinal Symptom Severity Scale. J Clin Med. 2024;13(6):1662. https://www.mdpi.com/2077-0383/13/6/1662

Santos JRA. Cronbach's Alpha: A Tool for Assessing the Reliability of Scales. Journal of Extension. 1999;37(2). https://open.clemson.edu/joe/vol37/iss2/15/

Appendix A. Rapid Cognitive Interview (RCI) – Method for Cognitive Assessment of DGBS

A1. Purpose

The **Rapid Cognitive Interview (RCI)** is a concise, 10-minute interview method for the cognitive validation of the DGBS measure, used either:

- as a preliminary stage for the main cognitive validation (Section 4.1)
- *or* as a lightweight checking method suitable for mobile use before piloting.

The goal of the RCI is to quickly ensure that the DGBS item's wording, time frame, and anchor levels are understandable and consistent in interpretation for respondents.

A2. Item to be Tested

"How much have your intestinal gas symptoms (flatulence, bloating, gas accumulation) disturbed or burdened you during the last 24 hours?"

Response scale 1–5 (No burden $\to$ Very high burden).

A3. Method and Duration

- **Method:** Think-aloud + three core sets of structured probing questions
- **Delivery:** Phone, video link, mobile device, or brief face-to-face interview
- **Duration:** 10 min / participant
- **Recommended Sample:** 6–10 participants (rapid saturation), separate N=15–20 for full cognitive interviews

A4. RCI Interview Structure (10 min)

1. Introduction (1 min)
   Tell the participant that the goal is to understand how they interpret the question—not to evaluate health. No right or wrong answers.
2. Think-aloud Section (2 min)
   Ask the participant to read the item and describe in their own words what they think about the question and what it includes.
   - **Goal:** Identify potential differences in understanding and interpretation.
3. **Three Core Probing Questions (4 min)**
   - "What all did you include in the phrase 'total gas symptom burden'?"
   - "How did you decide which level (1–5) matched your experience?"
   - "Was any word or part unclear?"
   - These map the interpretation of the construct, the decision-making process, and linguistic ambiguities.
4. Scale Differences – Mini-Probe (2 min)
   Depending on the respondent's choice:
   - If the answer is 2–3: "What is the difference between slight and moderate burden for you?"

    - If the answer is 4–5: "How do you know the burden is 'significant' and not just 'moderate'?"
    - **Goal:** Ensure a differentiating understanding of the anchor points.
5. Final Assessment (1 min)
   Two quick questions:
    - "Did you fully understand the question?" (Yes / No) — if not, what remained unclear?
    - "How easy would it be to answer this daily?" (1–5)

A5. Analysis (RCI Coding)

The analysis briefly records:

- Misunderstanding (Y/N)
- Unclear points (word/section)
- What components the respondent included in the "total burden"
- Understanding of the scale point (good / partial / poor)
- Ease of daily use (1–5)

The results of the RCI will be combined later with the observations from the main cognitive validation (Section 4.1).

A6. RCI's Role in the Validation Process

The RCI serves as a rapid preliminary study and supports the full cognitive interview in the following situations:

- Mobile prototype iteration
- Fine-tuning of wording before a larger pilot
- Testing of UX and micro-anchors
- Preliminary check of cultural and language versions

It does not replace formal cognitive validation but can shorten its iteration cycles.

Appendix A ends.