2018

The translation, validity and reliability of the German version of the Fremantle Back Awareness Questionnaire

Katja Ehrenbrusthoff
Cormac G. Ryan
Christian Gruneberg
Benedict Wand
*The University of Notre Dame Australia, benedict.wand@nd.edu.au*

Denis J. Martin

Follow this and additional works at: [https://researchonline.nd.edu.au/physiotherapy_article](https://researchonline.nd.edu.au/physiotherapy_article)

Part of the Physical Therapy Commons, and the Physiotherapy Commons

This article was originally published as:


Original article available here:

10.1371/journal.pone.0205244

This article is posted on ResearchOnline@ND at
[https://researchonline.nd.edu.au/physiotherapy_article/149](https://researchonline.nd.edu.au/physiotherapy_article/149). For more information, please contact researchonline@nd.edu.au.
The translation, validity and reliability of the German version of the Fremantle Back Awareness Questionnaire

Katja Ehrenbrusthoff, Cormac G. Ryan, Christian Grüneberg, Benedict M. Wand, Denis J. Martin

1 Health and Social Care Institute, Teesside University, Middlesbrough, Tees Valley, United Kingdom,
2 Hochschule für Gesundheit, Department of Applied Health Sciences, Bochum, Germany,
3 School of Physiotherapy, The University of Notre Dame Australia, Fremantle, Western Australia, Australia

These authors contributed equally to this work.
† These authors also contributed equally to this work.
* katja.ehrenbrusthoff@hs-gesundheit.de

Abstract

Background

The Fremantle Back Awareness Questionnaire (FreBAQ) claims to assess disrupted self-perception of the back. The aim of this study was to develop a German version of the FreBAQ (FreBAQ-G) and assess its test-retest reliability, its known-groups validity and its convergent validity with another purported measure of back perception.

Methods

The FreBaQ-G was translated following international guidelines for the transcultural adaptation of questionnaires. Thirty-five patients with non-specific CLBP and 48 healthy participants were recruited. Assessor one administered the FreBAQ-G to each patient with CLBP on two separate days to quantify intra-observer reliability. Assessor two administered the FreBaQ-G to each patient on day 1. The scores were compared to those obtained by assessor one on day 1 to assess inter-observer reliability. Known-groups validity was quantified by comparing the FreBAQ-G score between patients and healthy controls. To assess convergent validity, patient’s FreBAQ-G scores were correlated to their two-point discrimination (TPD) scores.

Results

Intra- and Inter-observer reliability were both moderate with ICC$_{3,1}$ = 0.88 (95%CI: 0.77 to 0.94) and 0.89 (95%CI: 0.79 to 0.94), respectively. Intra- and inter-observer limits of agreement (LoA) were 6.2 (95%CI: 5.0–8.1) and 6.0 (4.8–7.8), respectively. The adjusted mean difference between patients and controls was 5.4 (95%CI: 3.0 to 7.8, p<0.01). Patient’s FreBAQ-G scores were not associated with TPD thresholds (Pearson’s r = -0.05, p = 0.79).
Conclusions
The FreBAQ-G demonstrated a degree of reliability and known-groups validity. Interpretation of patient level data should be performed with caution because the LoA were substantial. It did not demonstrate convergent validity against TPD. Floor effects of some items of the FreBAQ-G may have influenced the validity and reliability results. The clinimetric properties of the FreBAQ-G require further investigation as a simple measure of disrupted self-perception of the back before firm recommendations on its use can be made.

Introduction
Low back pain is a major cause of disability worldwide [1] and is associated with substantial health care costs [2]. Many interventions attempt to normalise assumed peripheral structural pathology [3]. However, current treatment strategies provide limited and short-term pain relief [4].

The cortical body representation is distorted in people with persistent pain which may play an important role in the development and/or maintenance of pain [5]. Early work by Flor et al [6] using brain imaging identified somatosensory disorganization in patients with chronic low back pain (CLBP). More recent imaging studies in people with CLBP have shown structural and functional alterations in cortical and subcortical areas, associated with the processing of sensory information [7–9].

These changes can present clinically as alterations in a person’s body perception [10–13], decreased ability to distinguish/interpret peripheral sensory stimuli [14–16] and impaired lumbopelvic motor control [17]. Interventions targeting these perceptual distortions may present novel approaches to managing persistent low back pain [18–23]. Due to the growing research and clinical interest in cortical body representation, there is an increasing need for valid and reliable body perception measurement tools which are quick and easy to deliver clinically. A recent systematic review highlighted the lack of such measures and the need for further work in this area [24].

The Freemantle Back Awareness Questionnaire (FreBAQ) is a simple tool that claims to assess back-specific altered body perception [25]. It comprises 9 items with a five-point Likert scale attempting to investigate neglect-associated features, proprioceptive acuity, and a person’s perceived body image [26, 27].

In people with CLBP, the FreBAQ has been associated with a number of clinical characteristics such as pain duration (Pearson correlation $\rho = 0.357$, $p = 0.01$) and pain intensity (Spearman’s $r_{ho} = 0.40$, $p = 0.004$) [26] though others have found no such relationship [28]. As a measure of body perception it has demonstrated evidence for known-groups validity (median difference between healthy controls and people with CLBP = 11, Mann-Whitney test, $p < 0.001$) and reliability (ICC2,1agreement = 0.652 (95% CI: 0.307 to 0.848), ICC2,1consistency = 0.667 (95% CI: 0.317 to 0.857) [26] but did not provide evidence of convergent validity with other body perception measures in a different pain population [29]. The FreBAQ was recently translated and validated into Japanese (FreBAQ-J) [27] and Dutch [28]. However, a German version does not yet exist within the peer-reviewed literature.

The aim of this study was to produce a German version of the FreBAQ (FreBAQ-G) and assess its test-retest reliability, its known-groups validity and its convergent validity as a measure of disrupted self-perception of the back.
Methods

Translation and face validity

The FreBAQ translation was conducted following international guidelines for the transcultural adaptation of self-reported measures [30]. This process attempts to maintain content validity by ensuring that similar issues are covered, taking into consideration language differences and potentially differing sociocultural backgrounds [31].

Firstly, two native German speakers translated the original English FreBAQ independently of each other into German. One translator (K.E.) was a physiotherapist and researcher well acquainted with the subject area and one was a university graduated translator and occupational therapist but uninformed regarding the subject area (M.D.). Consensus regarding discrepancies was reached through discussion between both translators. Secondly, the revised German version was back-translated into English by two different translators, neither of whom had any specialist knowledge in the subject area of physiotherapy or chronic pain; one (J.J.H.) was a university lecturer in psychology and a native English speaker fluent in German while the other (A.J.) was a German English teacher, fluent in English. Again, consensus was reached regarding differences in wording through discussion between translators. Thirdly, the whole translation process was documented by K.E. and discussed with the developer of the original English version (B.W.) arriving at a pre-final German version (FreBAQ-G) (Appendix S1).

Once this translation process was complete, the FreBAQ-G was provided to a group of individuals with CLBP and healthy controls to assess its face validity. Face validity can be defined as “the degree to which a measurement instrument, looks as though it is an adequate reflection of the construct to be measured” [32]. As there are currently no standards concerning its measurement or quantification [33], three major aspects were assessed; completeness of content, comprehensibility and time to complete. Both groups provided feedback on completeness of content (“Do you think that this questionnaire covers the most important aspects of altered back related perception? [Yes/No]”; “If “NO” which aspects would you incorporate?”), comprehensibility (“Are the questions sufficiently comprehensibly worded? [Yes/No]”; “If “No” which items are not sufficiently comprehensible?”) and time to complete (“Is the time needed for filling in the questionnaire appropriate?) scored on a 0–10 scale with 0 representing “unacceptably long” and 10 “completely ok”). These questions were deemed to provide information on overall usability and whether it potentially needed revision due to content or linguistic ambiguities [28].

Floor and ceiling effects of the questionnaire were investigated in the CLBP group on an item level and by assessment of the total scores. Ceiling and floor effects occur when a considerable proportion of subjects score highest or lowest on a scale, demonstrating the measure unsuitable to discriminate between subjects at either extreme of the scale [34, 35]. Ceiling or floor effects were considered present if more than 15% of respondents achieved the highest or lowest possible score, respectively [36].

Participants

35 patients with non-specific CLBP were recruited consecutively from physiotherapy practices in Bochum, Germany between June 2013 and December 2014. Participants had to meet the following inclusion criteria: age ≥18 years; non-specific CLBP with or without leg pain (for those with leg pain, the back pain had to be dominant); duration of symptoms ≥6 months; sufficient cognitive and German language ability to understand both oral and written instructions, provide feedback and informed consent. Participants were excluded if they were pregnant or less than 6 months post-partum, had signs and symptoms indicating serious spinal
pathologies (i.e. red flags), thus differentiating them clinically from people with non-specific CLBP [37].

In addition, a sample of 48 healthy participants, recruited from staff and students (lower age limit: 18 years) at the University of Applied Health Sciences in Bochum, Germany were recruited. According to the original FreBAQ protocol by Wand et al [26] healthy participants had to meet the following criteria: currently back pain free, no episode of back pain within the last two years restricting them from work or leisure activities, sufficient cognitive and German language ability to understand both oral and written instructions, provide feedback and informed consent. Exclusion criteria were pregnancy or less than 6 months post-partum or significant spinal deformities. The study was approved by Teesside University’s School of Health and Social Care Research Governance and Ethics Board (Study No 186/12) and the Ethics committee of the German National Physiotherapists Society (Ethics committee submission number: 2013–02). Before study commencement, all participants provided written informed consent to participate in the study.

The patient population provided basic demographic data and clinical characteristics as well as a battery of outcome measures recommended for back pain research, including measures of symptom severity and frequency, physical function, general well-being and current work disability [38]. Demographic information comprised: age, sex, height, weight, body mass index (BMI), and current working status. Clinical characteristics comprised: duration of symptoms, FreBAQ-G, Brief Pain Inventory Short form (BPI) [39], (pain intensity and interference); Roland Morris Disability Questionnaire (RMDQ) [40] (function); Hospital Anxiety and Depression Scale (HADS) [40] (anxiety and depression) and Euroquol 5D-3L [41] (quality-of-life).

The control group provided the same basic demographic data and also completed the HADS and the FreBAQ-G. Regarding the FreBAQ-G instructions the wording was slightly adopted in that the phrase “other patients” was replaced by “other people” and the section concerning current pain experience was replaced by “please indicate to which degree your back feels like this”.

Within the data analysis a FreBAQ-G item which was not answered was categorised as ‘not endorsed’, in keeping with Wand et al [26] and scored as zero, representing “never feels like this”. All analyses were conducted using SPSS, version 24 (IBM, Armok, USA) or Microsoft Excel 2010, version 14 (Microsoft, Redmond, USA).

**Relationship to clinical status**

To quantify the association between the FreBAQ-G and the clinical characteristics of the patients a series of Pearson’s or Spearman’s correlations were conducted dependent upon the normality of the data. An r value of 0.10, 0.30 and 0.5 represented small, medium and large correlations respectively [42]. We hypothesised that people scoring higher on the FreBAQ-G, indicating a more disturbed self-perception of the back, would achieve poorer scores on other clinical outcome measures, assessing different constructs, such as pain and physical function.

**Reliability**

One assessor (KE) provided the FreBAQ-G to each participant with CLBP on two separate days. The participants were asked to complete the questionnaire independently in the presence of the assessor in a quiet room of the University’s outpatient department. The assessor did not provide any assistance with the completion of the questionnaire. Day 1 and day 2 were on average one week apart. FreBAQ-G scores between day 1 and day 2 (collected by assessor 1), were compared to quantify intra-observer reliability over one week. To quantify inter-observer
reliability, a second assessor provided the questionnaire to each participant on day 1 approximately two hours after it was provided by KE, and this was compared to the scores obtained by the first assessor’s administration on day 1. Assessors were blind to previous FreBAQ-G scores, as the questionnaires were immediately filed in a folder and only analysed upon the participant’s completion of the study. On day 2, the participant was not provided with any information regarding their previous scores to reduce the risk of recall bias.

The data obtained by assessor one on day one was used to quantify the frequencies of responses per item. The systematic bias (mean (95% CI) between data collected from two assessors on the same day and from one assessor from two sessions was determined using a paired t-test. Within-subjects standard deviations, defined as the standard error of measurement, coefficients of variation, limits of agreement and a random-error only intraclass correlation coefficient (ICC), model 3.1, were calculated to quantify the random error component within and between assessors.

The within-subjects SD was then used within a statistical power calculation to estimate whether the random measurement error identified in this study was small enough to detect a clinically relevant change in FreBAQ scores with a feasible sample size. As no MCID for the FreBAQ exists a value of 10 was chosen as the MCID within the power calculation based upon previous data quantifying the difference in FreBAQ between people with back pain and healthy controls [26]. ICC\(_{3,1}\) scores of <0.75 were considered to demonstrate poor reliability, 0.75–0.89 moderate and ≥0.90 excellent reliability [43]. Statistical significance was set at \(p\leq0.05\) [44].

**Validity**

**Internal consistency.** Internal consistency of the FreBAQ-G was assessed by calculating the Cronbach’s alpha coefficient. A correlation coefficient of at least 0.7 was defined as indicative of adequate inter-relatedness of items [33].

**Convergent validity.** Convergent validity is defined as a positive correlation between instruments assuming to measure the same underlying construct [33]. To assess the convergent validity the patient’s FreBAQ-G scores were correlated to their two-point discrimination (TPD) scores measured by assessor one on day 1. TPD is a simple clinical test of tactile acuity which measures the minimum distance between two points on the skin that can be obviously detected with smaller distances indicating better acuity [45]. It has been shown to be a valid measure of cortical reorganization when compared against the gold standard measure of fMRI [46] and other clinical tests which purport to measure body awareness indirectly such as movement control tests [47]. The TPD collection method and data have been published previously [48]. The measurement tool was a two-point discrimination caliper (Nexgen Medical Systems, Florida, USA) with a 1 mm precision. To minimise the risk of assessor bias assessor 1 was not aware of the FreBAQ results when undertaking the TPD assessment. In principal, we hypothesised a positive correlation between FreBAQ total scores and TPD results. The direction of this hypothesis was based on findings from previous studies, demonstrating an association between an altered body image and tactile acuity, measured by TPD [14, 16, 23].

**Known-groups validity.** Known-groups validity is defined as an instrument’s ability to differentiate between individuals with a specific condition and healthy individuals [33] or its ability to differentiate between two groups on a construct on which they theoretically should differ [49]. To investigate the known-groups validity of the FreBAQ-G the total score on the questionnaire was compared between the group of patients and healthy controls. There was no attempt to match groups regarding characteristics such as age and sex, which may affect body perception. However, when assessing the difference between groups an ANCOVA was used...
which adjusted for age, sex and BMI. We hypothesised that people with CLBP would differ from healthy controls on the construct of self-perception of the back, as assessed by the FreBAQ-G, in that healthy people would on average score lower compared to people with CLBP, demonstrating better self-perception of the back in healthy controls.

**Results**

**Translation and face validity**

The majority of participants in both groups found the FreBAQ-G to be a complete and comprehensible measure, which could be completed within an appropriate period of time (Table 1).

In the patient group, additional questions about content covering aspects of night sleep, stair climbing, current awareness of posture, morning stiffness and current sensory abnormalities hampering body awareness were suggested for inclusion. Regarding comprehensibility it was stated by one patient that the double negative expressions in question 4, 5, and 6 could be misleading. With respect to questions 2 and 3 one individual suggested to provide examples of which specific activities were meant.

Given the qualitative feedback, and as all three scores of the feedback form were well below the preset threshold of 50% negative responses, it was judged that the translation process revealed no obvious cultural adaptations necessary for a German speaking population.

**Participant characteristics and questionnaire responses**

The participant characteristics for each group are shown in Table 2. On average the control group was 16 years younger and there were small differences in sex and BMI between groups. Thus age, sex and BMI were adjusted for as covariates in the comparison between groups.

For the patient group, pain severity at time 1 and 2, were 3.6 and 3.5 respectively, defined as mild severity [50]. The average back related physical function was 7.5 at time 1, defined as a mild-to-moderate functional impairment [51]. In addition, both groups had similar HADS anxiety scores (mean (SD) CLBP group: 5.2 (3.4), control group: 4.8 (2.2)), both of which could be interpreted as normal [52].

Of the 35 participants with CLBP, two did not answer item 2 at timepoint one, one did not answer item 8 at timepoint one and two and one participant did not answer item 7 at timepoint two. In all, missing items account for 1.6% of the data, thus it is unlikely that they had a significant impact on the overall results.

The frequencies of FreBAQ-G responses, as well as the mean and median scores for the patient group are displayed per item in Table 3. All nine items were at least endorsed at some level, although reported frequencies differed across items. Items 2 and 9 were the most often endorsed. In contrast, items 6, 7 and 8 were the most infrequently endorsed, with more than

<table>
<thead>
<tr>
<th>Item</th>
<th>CLBP group (n = 35)</th>
<th>Control group (n = 48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Completeness of Contents N (%)</td>
<td>25 (71.4)</td>
</tr>
<tr>
<td>2</td>
<td>Comprehensibility N (%)</td>
<td>29 (82.9)</td>
</tr>
<tr>
<td>3</td>
<td>Appropriateness of time to complete Mean (SD)</td>
<td>9.9 (0.8)</td>
</tr>
</tbody>
</table>

N = proportion of participants stating complete agreement or acceptance with item; CLBP = chronic low back pain

https://doi.org/10.1371/journal.pone.0205244.t001
80% of participants stating that their back never feels shrunk (item 8). In contrast, items 3, 4, 5 and 6 were not endorsed on the upper end of the Likert scale (‘always feels like this’).

### Relationship to clinical status

The associations between the FreBAQ-G and the clinical characteristics in the patient group were moderate for all characteristics except for duration of symptoms, which was unrelated to the FreBAQ-G (see Table 4).

### Reliability

**Intra-observer reliability.** The mean value for the FreBAQ-G scores obtained from the participants by assessor 1 on day 1 was 8.8 (SD 6.1) and 7.8 (SD 7.0) on day 2. The mean FreBAQ-G difference score within one week for assessor 1 was 1.06 (95% CI: -0.03 to 2.14, p = 0.055). The ICC_{3,1} values for absolute agreement and consistency were 0.88 (95%CI: 0.77 to 0.94) and 0.89 (95%CI: 0.79 to 0.94), respectively. The Bland and Altman plot for the individual differences between day 1 and 2 for assessor 1 is shown in Fig 1.
Inter-observer reliability. The mean value for the FreBAQ-G scores obtained from the participants by assessor 1 and 2 on day 1 was 8.8 (SD 6.1) and 7.4 (SD 7.2) respectively. The mean FreBAQ-G difference score on the same day between assessors 1 and 2 was 1.4 (95% CI: 0.36 to 2.45, p = 0.01). The ICC values for absolute agreement and consistency were 0.88 (95%CI: 0.75 to 0.94) and 0.90 (95%CI: 0.81 to 0.95), respectively. The Bland and Altman plot for the individual differences between assessor 1 and 2 is shown in Fig 2.

Table 3. Frequency of responses to each FreBAQ-G item in the patient group (n = 35).

<table>
<thead>
<tr>
<th>Item</th>
<th>Response category</th>
<th>Never feels like that</th>
<th>Rarely feels like that</th>
<th>Occasionally or some of the time feels like that</th>
<th>Often, or a moderate amount of time feels like that</th>
<th>Always, or most of the time feels like that</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scoring</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of Responses</td>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>My back feels as though it is not part of the rest of my body</td>
<td>19 (54.3)</td>
<td>6 (17.1)</td>
<td>4 (11.4)</td>
<td>5 (14.3)</td>
<td>1 (2.9)</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>I need to focus all my attention on my back to make it move the way I want it to</td>
<td>7 (20.0)</td>
<td>9 (25.7)</td>
<td>8 (22.9)</td>
<td>9 (25.7)</td>
<td>2 (5.7)</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>3</td>
<td>I feel as if my back sometimes moves involuntarily, without my control</td>
<td>15 (42.9)</td>
<td>10 (28.6)</td>
<td>8 (22.9)</td>
<td>2 (5.7)</td>
<td>0 (0)</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>When performing everyday tasks, I don’t know how my back is moving</td>
<td>13 (37.1)</td>
<td>11 (31.4)</td>
<td>8 (22.9)</td>
<td>3 (8.6)</td>
<td>0 (0)</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>I am not sure exactly what position my back is in</td>
<td>15 (42.9)</td>
<td>9 (25.7)</td>
<td>8 (22.9)</td>
<td>3 (8.6)</td>
<td>0 (0)</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>I can’t perceive the exact outline of my back</td>
<td>20 (57.1)</td>
<td>7 (20.0)</td>
<td>5 (14.3)</td>
<td>3 (8.6)</td>
<td>0 (0)</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>7</td>
<td>My back feels like it is enlarged (swollen)</td>
<td>23 (65.7)</td>
<td>1 (2.9)</td>
<td>7 (20.0)</td>
<td>3 (8.6)</td>
<td>1 (2.9)</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>My back feels like it has shrunk</td>
<td>28 (80.0)</td>
<td>3 (8.6)</td>
<td>2 (5.7)</td>
<td>1 (2.9)</td>
<td>1 (2.9)</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>9</td>
<td>My back feels lopsided (asymmetrical)</td>
<td>12 (34.3)</td>
<td>5 (14.3)</td>
<td>6 (17.1)</td>
<td>7 (20.0)</td>
<td>5 (14.3)</td>
<td>2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

N = absolute frequency responses, FreBAQ-G = Fremantle Back Awareness Questionnaire German

https://doi.org/10.1371/journal.pone.0205244.t003

Table 4. Univariate correlations between FreBAQ-G total scores and clinical characteristics in the patient population (n = 35).

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety (HADS-A)a</td>
<td>0.37</td>
<td>0.03*</td>
</tr>
<tr>
<td>Depression (HADS-D)b</td>
<td>0.52</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Duration of Symptoms (Years) b</td>
<td>0.06</td>
<td>0.72</td>
</tr>
<tr>
<td>Pain severity (BPI)  a</td>
<td>0.32</td>
<td>0.07</td>
</tr>
<tr>
<td>Pain interference (BPI)b</td>
<td>0.47</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Back related disability (RMDQ)b</td>
<td>0.46</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

HADS-A = Hospital Anxiety and Depression Scale- Anxiety; HADS-D = Hospital Anxiety and Depression Scale- Depression; BPI = Brief Pain Inventory; RMDQ = Roland-Morris-Disability Questionnaire.

*a Correlation is significant at the 0.05 level (2-tailed).
*bPearson correlation coefficient.
*bSpearman correlation coefficient.

https://doi.org/10.1371/journal.pone.0205244.t004
The systematic bias for inter- and intra-observer reliability in both cases equaled approximately one unit on the 0–36 FreBAQ-G scale. All data quantifying the systematic and random error components of the reliability analysis are displayed in Table 5.

Fig 1. Limits of agreement for intra-observer reliability. For intra-observer reliability, the FreBAQ-G difference scores for assessor 1 at day 1 and 2 are plotted against their mean scores. Mean session differences (systematic bias) are displayed by solid lines and limits of agreement by dashed lines.

https://doi.org/10.1371/journal.pone.0205244.g001

The systematic bias for inter- and intra-observer reliability in both cases equaled approximately one unit on the 0–36 FreBAQ-G scale. All data quantifying the systematic and random error components of the reliability analysis are displayed in Table 5.

Fig 2. Limits of agreement for inter-observer reliability. For inter-observer reliability, the FreBAQ-G difference scores for assessor 1 and 2 are plotted against their mean scores. Mean session differences (systematic bias) are displayed by solid lines and limits of agreement by dashed lines.

https://doi.org/10.1371/journal.pone.0205244.g002
Validity

Internal consistency. A Cronbach’s alpha value of 0.91 indicated an adequate internal consistency of all items of the German version of the FreBAQ in people with chronic low back pain. Table 6 displays Cronbach’s Alpha values, given that one out of nine items was deleted, as well as inter-item correlations and total-item correlations.

Moreover, the internal consistency score was not severely affected by deletion of one item and correlations greater 0.7 were found between each item and the total score except for item 7 (r = 0.57), item 8 (r = 0.43) and item 9 (0.66).

Known-groups validity. The FreBAQ-G total scores in the patient group ranged from 0–21, the mean score (SD) was 8.8 (6.1) and the median score 7.0. In the control group, the total FreBAQ-G score ranged from 0–13, the mean score (SD) was 4.0 (3.3) and the median score was 3.0. FreBAQ-G scores were, on average, higher in the CLBP group compared to the control group \[\text{unadjusted mean difference (95%CI)} 4.8 (2.55 to 7.15), p < 0.01\].

There was a statistically significant effect between groups regarding the FreBAQ-G scores after adjusting for age, gender and BMI \[F (1.78) = 20.39, p < 0.001, \text{adjusted } R^2 = 0.22\]. The adjusted mean scores for the patient group was 9.1 (95%CI: 7.77 to 10.87) and 3.7 (95%CI: 2.31 to 5.18) for the control group with an adjusted mean difference of 5.4 (95%CI: 3.02 to 7.79, p < 0.01).

Table 5. Intra- and Inter-observer reliability data.

<table>
<thead>
<tr>
<th></th>
<th>Intra-observer</th>
<th>Inter-observer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean session difference</td>
<td>1.06 (-0.03–2.14)</td>
<td>1.40 (0.36–2.44)</td>
</tr>
<tr>
<td>SD of session differences</td>
<td>3.15 (2.55–4.13)</td>
<td>3.04 (2.46–3.98)</td>
</tr>
<tr>
<td>Within-subjects SD (SEM)</td>
<td>2.23 (1.80–2.92)</td>
<td>2.15 (1.74–2.82)</td>
</tr>
<tr>
<td>Limits of agreement (LOA)</td>
<td>6.18 (-5.12–7.24)</td>
<td>5.96 (-4.56–7.36)</td>
</tr>
<tr>
<td>ICC\textsubscript{3.1} absolute agreement (95%CI)</td>
<td>0.88 (0.75–0.94)</td>
<td>0.90 (0.79–0.94)</td>
</tr>
</tbody>
</table>

SD = Standard Deviation; SEM = Standard Error of Measurement; CI = Confidence Interval, ICC\textsubscript{3.1} = Intraclass Correlation coefficient (absolute agreement). Ninety-five percent confidence intervals are presented in brackets for all metrics.

https://doi.org/10.1371/journal.pone.0205244.t005

Validating the German Fremantle Back Awareness Questionnaire in people with chronic low back pain

Table 6. Internal consistency of the German Fremantle back awareness questionnaire in people with chronic low back pain.

<table>
<thead>
<tr>
<th>FreBAQ-G item</th>
<th>Cronbach’s Alpha if item deleted</th>
<th>Item-total Correlation</th>
<th>Inter-Item Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Item 1</td>
<td>Item 2</td>
<td>Item 3</td>
</tr>
<tr>
<td>Item 1</td>
<td>0.90</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>0.90</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.90</td>
<td>0.76</td>
<td>0.53</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.90</td>
<td>0.73</td>
<td>0.59</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.89</td>
<td>0.84</td>
<td>0.67</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.89</td>
<td>0.81</td>
<td>0.72</td>
</tr>
<tr>
<td>Item 7</td>
<td>0.91</td>
<td>0.57</td>
<td>0.47</td>
</tr>
<tr>
<td>Item 8</td>
<td>0.91</td>
<td>0.43</td>
<td>0.30</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.90</td>
<td>0.66</td>
<td>0.52</td>
</tr>
</tbody>
</table>

FreBAQ-G = Fremantle Back Awareness Questionnaire-German version

https://doi.org/10.1371/journal.pone.0205244.t006
Convergent validity. The total FreBAQ scores were not associated with the mean TPD thresholds (Spearman’s \( \rho = -0.05 \), \( p = 0.79 \)) in the patient group (see Fig 3).

Discussion
Participants found the FreBAQ-G demonstrated completeness of contents, comprehensibility and could be completed within an acceptable amount of time. These findings are in line with the results of the cross-cultural adaptation of the Dutch version of the FreBAQ-Q [28], in which participants (n = 22) with CLBP reported an overall acceptable comprehensibility of 77% and an acceptable level of completeness of contents of 82%. Quantitatively, the mean score of 8.8 for the patients on the FreBAQ-G, was similar to those reported for the original English version (10.8) [26] and the Dutch version (11) [28]. This adds confidence to the translation process and cross-cultural validity of the FreBAQ-G. Three out of 35 participants in the patient group scored 0 in total, equaling 9% of the total scores. This was below our predefined criteria of 15%, suggesting that floor/ceiling effects were not an issue for the questionnaire as a total score. However, from the frequency of responses per item potential floor effects of the FreBAQ-G could be deduced while there was no evidence of ceiling effects. These item specific floor effects could have artificially enhanced the level of reliability and whilst have a detrimental effect on the validity of the FreBAQ-G reported in this study.

The FreBAQ-G demonstrated adequate internal consistency, with a Cronbach’s alpha of 0.91 being slightly higher compared to other translated versions [27, 28]. However, as our sample size was smaller than those of the other validation studies, these results need to be interpreted cautiously.

The FreBAQ-G showed moderate intra and inter-observer reliability with ICC\(_{3,1}\) values of \( \geq 0.88 \). These values were similar (or higher) to those reported for the Japanese [27] (ICC\(_{3,1}\) of 0.81 (95% CI: 0.67–0.89)), Dutch [28] (ICC\(_{2,1} = 0.69\) (95%CI: (0.51–0.82)) and original English...
version [26] (ICC\(_{2,1} = 0.65\) (95% CI: 0.307–0.848)). A systematic bias of one unit between time 1 and 2 for observer 1 indicated some small learning effects, thus a familiarization session may be warranted when using this questionnaire.

The SEM (intra- and inter-observer) in our study was ~2 units, below the SEM of 3.5 reported by Janssens et al [28]. However, with 95% limits of agreement ~6 units, this indicates an individual patient with CLBP could change by as much as 6 units due to normal variation. In addition, a random error component of ~26% (CV) suggests the FreBAQ-G may be more appropriately used on a group level rather than an individual patient level.

To understand if the FreBAQ-G has sufficient reliability for research purposes it can be useful to use the estimated variability of the measure and its minimally clinically important difference (MCID) to calculate sample sizes for different study designs. There is no existing empirically derived MCID for the FreBAQ. Using 0.5 of a standard deviation as a clinically worthwhile change one could estimate an MCID of approximately 3.0 units for power calculation purposes. Assuming the SD of change is 3.15 (see Table 5) it can be estimated that \( n = 14 \) would be required for a single arm pre-post study (two-tailed significance level \(< 0.05\), statistical power = 90%) to detect the difference between a null hypothesis mean of 0.0 and an alternative mean of 3.0 units. Within an RCT design, under the same conditions, a sample size of \( n = 25 \) in each arm would be required. Both estimated sample sizes could be considered achievable within a musculoskeletal research context, supporting the potential of the FreBAQ in research.

The convergent validity of the FreBAQ-G was assessed by correlating it with TPD. There was no correlation between the FreBAQ-G and the TPD, in contrast to our initial hypothesis. These results were in keeping with Wand et al [29] who found no correlation between the English version of the FreBAQ and TPD in a sample of 34 pregnant women. This questions the assumption whether both assessments measure the same construct, although previous studies have demonstrated a relationship between body image drawings and tactile acuity in patients with CLBP [13, 14]. However, outlining or drawing one’s perceived body image and answering dedicated questions regarding one’s perceived body awareness might require different cognitive and self-reflective skills. Moreover, TPD testing constitutes a direct measurement requiring touch. Hence, TPD could be seen as a test to investigate peripheral innervation density and/or intact neural sensory pathways rather than a person’s perceived body image [53].

Body perception as measured by the FreBAQ-G was correlated with a number of the clinical outcomes assessed. This implies that the FreBAQ-G may have clinical utility and body perception may be a clinically relevant construct in this patient population. In our study sample, disturbed body perception was associated with pain interference scores (BPI-I), but not with symptom duration. In addition, FreBAQ-G scores showed moderate correlations to back related disability (RMDQ) and anxiety and depression scores (HADS). It may be possible that an altered self-perception of the back, in particular motor neglect aspects, might contribute to motor control impairments, resulting in higher back related disability scores [54, 55]. In addition, a growing body of evidence supports the notion that anxiety and depression negatively affect an individual’s confidence in an adequate loading of the back and might hence contribute to the distortion of the self-perception of the back [56, 57].

Our findings are partly in line with both English study samples, where statistically significant correlations to pain severity were found (Pearson’s \( r = 0.40, p = 0.04 \)) [26] and (Pearson’s \( r = 0.27, p < 0.001 \)) [25]. The strength of the relationship between the FreBAQ-G and pain severity in our sample was similar to those studies (\( r = 0.32, p = 0.07 \)). In addition, the Japanese sample [27] showed only correlations to back pain intensity in motion whereas the Dutch sample did not demonstrate any correlations to pain intensity at all [28]. Differences between our findings and those of other studies may due to differences in methodology. The differences
here could be attributable to the greater anxiety and depression scores in our study sample compared to the Japanese study and that our sample showed higher values in pain scores interfering with daily function (BPI-pain interference scores) in contrast to the pain intensity in motion scores in the Japanese sample. However, all existing versions of the FreBAQ showed a correlation between back related disability and disturbed body perception [26–28]. This could be explained by the fact that an inability to adequately perform activities of daily living might be associated with reduced sensorimotor lumbopelvic control [17, 55].

In contrast to Wand et al [26] and Nishigami et al [27], our sample showed an association between anxiety and depression scores and disturbed body perception. This finding may be attributable to the notion that cognitive emotional aspects of pain drive central nervous adaptation, such as central sensitization, which may in turn modulate sensorimotor control and body perception [58].

The FreBAQ-G demonstrated a degree of known-groups validity, identifying a difference of ~5 units between individuals with CLBP and health participants, after adjusting for age, gender and BMI. The difference between groups was half that previously reported (11.0 units) using the original FreBAQ [26]. This difference may have been due to sample differences in both the clinical and control participants between that study and our study.

**Strengths and limitations**

Regarding the translation process, an initial pre-testing phase in a smaller sample of patients with CLBP could have been utilized to reveal and resolve any difficulties regarding comprehensibility and completeness of contents before commencing the study. However, patients were satisfied with all usability aspects. In addition, we did not measure the exact amount of time it took patients to complete the questionnaire though participants reported that the time to complete was appropriate in their opinion. The current version of the FreBAQ-G demonstrated evidence of floor effects on an item level. This might have adversely affected reliability and validity scores. However, regarding sum scores, the percentage of respondents scoring 0 were below the pre-defined cut-off value of 15%. Hence, our main criterion demonstrated that floor effects did not appear to be an issue in our sample.

Although all the patient participants in our study were patients accessing a health care setting for treatment of their CLBP they were on the low end of the spectrum for the range of clinical measures that were used, especially regarding anxiety and depression scores. Thus, our findings may not be generalisable to the wider CLBP population, especially those scoring higher on the clinical spectrum.

Our final sample size of 35 patients was lower than current recommendations of 40 participants or more for reliability studies [59]. Initially, 51 individuals were contacted. Ten people did not respond to any further communication and six did not meet the inclusion criteria. In addition, for the known-groups validity testing, the design would have been improved if groups were matched on key characteristics such as age, sex and BMI, however these were adjusted for statistically in the analysis.

To assess the convergent validity of the FreBAQ-G, its scores were correlated to TPD performance which claims to measure the same/or similar construct. The choice of comparator measure to assess convergent validity was difficult as there is no gold standard measure for the construct of self-perception of the back. A recent systematic review published by our group [24] found there were no existing measures of sensory motor perception that have demonstrated adequate levels of validity and reliability. However, the review did identify TPD as one of the most promising measures. In addition, TPD is one of the most commonly used measures of back perception within the literature [60–62] and it has been previously used as
comparator for other measures of sensorimotor back function [17]. Thus it was chosen as the comparator in this study but the findings should be interpreted cautiously. Finally, while components of the validity of the FreBAQ-G have been assessed, definitive evidence that the FreBAQ-G measures the construct back self-perception is lacking. This is likely attributable to the fact that self-perception is a complex construct to define and, as previously stated, no definitive gold standard measure exists. Further exploration of the validity of the FreBAQ-G is warranted.

Clinical implications

The translation and assessment of the German FreBAQ is an important step in the use of this questionnaire in people with CLBP, as is makes it available to a German speaking population of 118 million people [63]. The FreBAQ-G constitutes a time efficient, low-cost and safe assessment tool, provisionally demonstrating acceptable levels of reliability for research purposes though it is unclear if the level of reliability is sufficient to be used at the individual patient level. There is evidence of small learning effects, thus a familiarisation session would appear warranted.

The FreBAQ-G is not proposed as an alternative outcome measure to established clinical measures such as pain and function. However, if a researcher/clinician wishes to assess the specific construct of self-perception of the back very few instruments are available and the clinimetric properties of those measures are limited [24]. If self-perception of the back is a construct of interest the FreBAQ-G could be a potentially useful tool. However, it should be employed knowing that the current level of validity is unclear and its level of reliability is not yet sufficient to be used on an individual patient level. Further research is required before firm recommendations on the use of the FreBAQ-G can be made.

Conclusion

Main results

We created a German translation of the FreBAQ. The FreBAQ-G demonstrated a degree of reliability and known-groups validity, while it did not demonstrate convergent validity against a measure, which purports to assess the same construct. These findings are broadly in keeping with other language versions of the questionnaire. The clinimetric properties of the FreBAQ-G require further investigation as a simple measure of self-perception of the back.

Practical tips

Given the degree of measurement error the FreBAQ-G could potentially be employed for research purposes to assess back self-perception but it may be too variable to monitor change in individual patients. To minimize learning effects, a familiarisation trial should be considered. The validity of the FreBAQ-G requires further exploration.

Supporting information

S1 File. The Fremantle Back Awareness Questionnaire-German (FreBAQ-G).

(PDF)

S1 Table. Supporting Information–Raw Data set. FreBAQ = Fremantle Back Awareness Questionnaire; TPD = Two-Point Discrimination; HC = Healthy controls.

(DOCX)
Acknowledgments

The authors gratefully acknowledge Dr J. Honisch and Ms A. Janßen for their support in the translation process. The authors thank J. Emmert, M. Giesen, K. Heidel, A. Heller, M. Moscheik, L. Palici, L. Steinbrunner and L. Weissert who participated as students in this project.

Author Contributions


Data curation: Katja Ehrenbrusthoff.


Funding acquisition: Christian Grüneberg.

Investigation: Katja Ehrenbrusthoff.


Project administration: Katja Ehrenbrusthoff, Cormac G. Ryan, Christian Grüneberg.

Resources: Christian Grüneberg.


Validation: Cormac G. Ryan, Benedict M. Wand, Denis J. Martin.

Visualization: Katja Ehrenbrusthoff, Denis J. Martin.

Writing – original draft: Katja Ehrenbrusthoff.

Writing – review & editing: Katja Ehrenbrusthoff, Cormac G. Ryan, Christian Grüneberg, Benedict M. Wand, Denis J. Martin.

References


8. Vrana A, Meier ML, Holz-Boendermaker S, Humphreys BK, Scholkmann F. Cortical Sensorimotor Processing of Painful Pressure in Patients with Chronic Lower Back Pain: An Optical Neuroimaging Study


