Analysis of Test-Retest Reliability, Construct Validity, and Internal Consistency of the Brazilian Version of the Pelvic Girdle Questionnaire

Luan Simões, MS,[a] Luci Fuscaldi Teixeira-Salmela, PhD,[b] Lívia Magalhães, PhD,[b] Britt Stuge, PhD,[c] Glória Laurentino, PhD,[a] Elaine Wanderley, MS,[a] Raphaela Barros, PT,[a] and Andrea Lemos, PhD[a]

Abstract

Objective: The purpose of this study was to evaluate test-retest reliability, construct validity, and internal consistency of the Brazilian version of the Pelvic Girdle Questionnaire (PGQ-Brazil).

Methods: Analysis of the measurement properties was carried out in 4 steps. Step 1 was the pilot study, on which basis 4 hypotheses were formulated. These hypotheses were tested during the next step (construct validity, step 2) by completion of the questionnaire by the 2 groups (in pain [n = 105] and not in pain [n = 52]). For implementation of the PGQ-Brazil in the group with pain, we calculated the internal consistency (step 3) and, 7 days later, test-retest reliability (step 4) by re-application of the instrument in this group.

Results: First, the PGQ-Brazil was able to discriminate between these groups (construct validity). Second, test-retest reliability (intraclass correlation coefficients for Activities subscale [0.97 with 95% confidence interval of 0.95-0.98] and Symptoms subscale [0.98 with 95% confidence interval of 0.97-0.98] and κ coefficient between 0.50 and 0.89 for the items) was found to be good; the Bland-Altman test indicated satisfactory agreement. The Rasch analysis indicated good internal consistency, and the instrument's ability to divide the participants into at least 3 levels of skills was confirmed. In contrast, a ceiling effect was observed, as 24% of pregnant women exhibited skills superior to what the PGQ-Brazil could evaluate.

Conclusions: The PGQ-Brazil had good internal consistency, test-retest reliability, and construct validity in assessment of limitations in activities and symptoms of pregnant women with pelvic girdle pain. (J Manipulative Physiol Ther 2018;41:425-433)

Key Indexing Terms: *Disability Evaluation; Pain; Pelvic Girdle Pain; Pregnancy*

Introduction

Lumbopelvic pain during pregnancy is a major complaint, not only because of the high frequency (21%-81%) of women affected, but also because of the intensity of pain and discomfort reported.[1-3] However, there is growing evidence that pelvic disorders constitute a distinct subgroup of low back pain with a unique clinical presentation and the need for specific treatment, meaning there is a clinical classification that distinguishes the 2 specific conditions.[4-9] Pelvic girdle pain (PGP) related to pregnancy is experienced between the posterior iliac crest and the gluteal fold, particularly near the sacroiliac joints from which it may radiate to the posterior thigh.[6,9-12]

Commonly, studies that assess the functional capacity of pregnant women with PGP use tools developed for a nonpregnant population with low back pain.[4,12-17] To fill this gap, in 2011, a Norwegian research group proposed a condition-specific questionnaire for PGP called the Pelvic Girdle Questionnaire (PGQ).[18] The instrument contained 25 items in 2 subscales (Activities subscale: 20 items; Symptoms subscale: 5 items) with 4 response options (0-3 points). The overall score is calculated by dividing the sum of the subscale scores by 75, resulting in percentage scores ranging from 0 (no disability) to 100 (severe disability). Three points are deleted for each "not applicable" response to an item. A major strength of the questionnaire is that data were collected from a representative sample of pregnant and postpartum women

[a] Department of Physiotherapy, Federal University of Pernambuco, Recife, Pernambuco, Brazil.
[b] Physical Education and Physiotherapy School, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[c] Division of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway.
Corresponding author: Andrea Lemos, PhD, Avenida Jorn, Anibal Fernandes, s/n Cidade Universitária, CEP: 50740-560 Recife, PE-Brazil. Tel.: +81 992325707.
(e-mail: *andrealemos4@gmail.com*).

Paper submitted April 18, 2017; in revised form September 30, 2017; accepted October 19, 2017.
0161-4754

https://doi.org/10.1016/j.jmpt.2017.10.008

with clinically confirmed PGP. Furthermore, the questionnaire is simple to administer and is feasible for use in both research and clinical practice.[18]

During its development, the PGQ was found to have good internal consistency and good test-retest reliability for both subscales and satisfactory construct validity, thereby indicating the validity of this instrument for the assessment of function in pregnant and postpartum women diagnosed with PGP related to pregnancy.[18,19]

For use of this instrument in a different country, however, it was necessary to evaluate its measurement properties to ensure their quality in a different sociocultural environment. Thus, the aim of this study was to analyze the test-retest reliability, construct validity, and internal consistency of an interview-based version of the Brazilian version of the PGQ (PGQ-Brazil).

Methods

The PGQ-Brazil is a socioculturally adapted version of the PGQ for Brazilian women. The PGQ was adapted by translation, back-translation, an expert committee, a Delphi study, and a pretest[20] (Appendix A). Its measurement properties were studied based on the recommendations of the Consensus-Based Standards for the Selection of Health Measurement Instruments and Terwee et al.[21] This phase of the study consisted of 4 steps: pilot study, construct validity, test-retest reliability, and internal consistency.

Before analysis of the PGQ-Brazil's measurement properties, a pilot study was conducted on a sample of 21 pregnant women, 9 without complaints of pain and 12 diagnosed with PGP, whose data were not included in the analysis of the study. The purpose of this pilot study was to train the evaluator in application of the specific tests for the diagnosis of PGP, as well as to formulate hypotheses to be tested during the construct validity step (hypothesis testing).

From the pilot study, 4 hypotheses were developed:

1. The mean score on the Activities subscale for the pregnant women with pain will be greater than that for pain-free pregnant women.
2. The mean score on the Symptoms subscale for pregnant women with pain will be greater than that for pain-free pregnant women.
3. The mean total score for pregnant women with pain will be greater than that for pain-free pregnant women.
4. The instrument can distinguish between pregnant women with pain and pregnant women without pain reaching a mean difference in the total score pre-established in the pilot study of 34.76 ± 7.40% within the 95% confidence interval (CI) of 50.26 to 19.25.

These hypotheses were tested by comparing 2 different groups: a group of 52 pregnant women without pain complaints and a group of 105 pregnant women diagnosed with PGP. All participants, regardless of group, could be primiparous or multiparous and had to be 18 years or older and at more than 18 weeks of gestation during the evaluation. Among the pregnant women with PGP, those who also had low back pain were eliminated. In the group without pain, patients who had back pain related to pregnancy were excluded from the study. Women with neuromuscular, rheumatic, or urogenital diseases were excluded from both groups.

The same group of 105 pregnant women with a diagnosis of PGP related to pregnancy involved in the construct validity step also participated in the test-retest reliability and internal consistency steps. Patients were selected according to the criteria described above and were recruited sequentially from 6 Family Healthcare Units in the city of Recife, state of Pernambuco (Brazil), in the period January 2013 to January 2014.

Sample size was based on the Consensus-Based Standards for the Selection of Health Measurement Instruments and specific criteria for studies of measurement properties, which recommend minimums of 50 individuals to study reliability and 100 individuals to study internal consistency through Rasch analysis.[21-25]

Pelvic girdle pain was diagnosed based on the presence of pain as assessed with pain provocation tests recommended by the European guidelines[6] (posterior pelvic pain provocation test, pubic symphysis provocation test by modified Trendelenburg test, palpation of the long dorsal sacroiliac ligament, and palpation of the pubic symphysis) and 1 functional test (the active straight leg raise test). The diagnosis was later confirmed when the posterior pelvic provocation test and/or active straight leg raise (right and/or left sides) and any of the other 3 tests were also positive.[4,6,7] Pain location was assessed by having participants mark, on a drawing, the body areas in which they had pain, in 4 different regions the pelvis: symphysis, left sacroiliac joint, right sacroiliac joint, and over the sacrum.

After inclusion of the pregnant women in the study, demographic and clinical data were collected using a standardized form. For analysis of internal consistency and performance of the initial verification of test-retest reliability, the women were interviewed by the researcher using the PGQ-Brazil. After 7 days, test-retest reliability was measured a second time by re-application of the questionnaire to the same pregnant women.

It is worth mentioning that the 2 interviews were conducted by the same examiner, under the same environmental and clinical conditions. Also, to ensure the clinical stability of pregnant women, the visual analogue scale (0-10) was used, excluding those women whose scores changed ≤2 points in the second evaluation.

Data Processing and Analysis

The sample was characterized using descriptive statistics. Hypothesis testing was calculated by comparing averages using the Mann-Whitney test for independent samples. To determine test-retest reliability, the intraclass correlation

coefficient (ICC) was used for continuous variables, and the κ coefficient for ordinal variables. These analyses were performed in the SPSS Statistical Package for Windows, Version 20.0 (IBM, Armonk, New York).

We performed Bland-Altman analysis to check for the occurrence of systematic or random change in averages and calculated the standard errors of measurement (SEMs = standard deviation × $\sqrt{1 - ICC}$) of the test and retest. Then, we calculated the individual minimal detectable change (MDC_{ind}) and group minimal detectable change (MDC_{group}) for the 95% CI: $MDC_{ind} = ASE \times 1.96 \times \sqrt{2}$ and $MDC_{group} = MDC_{ind} / \sqrt{n}$ of the group.

Internal consistency was assessed by Rasch analysis using Winsteps Software 3.80.1. The primary objective of Rasch analysis is to calibrate the difficulty of the items and the ability level of individuals measured in the same simple linear continuum, divided into equal intervals called *logits*. The items are arranged in order of difficulty (easiest to hardest), and the individuals in order of ability (least skilled to most skilled).[26]

After calibration, the mean square (MnSq), which represents the relationship between the expected model score and the obtained score, was calculated (in infit and outfit formats).[14] A result of 1 means that the relationship is in accordance with the assumptions of the model; a slight variation of ±0.3 is allowed. The MnSq infit format signals fluctuations in scores, and the MnSq outfit format signals the presence of extreme scores.[21,26,27]

Mean square values >1.3 associated with $Z = \pm 2$ indicate that the scores for the item are erratic (very variable); that is, individuals with high abilities received low scores, or individuals with low abilities received high scores. In contrast, MnSq values <0.7 with $Z < -2$ indicate that the item does not discriminate people with different ability levels, but are not considered erratic. The presence of 5% erratic items indicates that the range of items does not combine to form a 1-dimensional concept, which threatens the construct validity of the instrument.[27-29]

The separation of people index, which is the capacity of items to divide the population into different levels of abilities, was also calculated. A more accurate calculation of the number of ability levels is obtained using the following formula: Number of distinct levels = $(4G + 1)/3$, where G is the separation index. The items of an instrument are expected to divide the population into at least 2 ability levels.[30]

Still using the Rasch analysis, the presence of a "floor" or "ceiling" effect was verified based on estimation of the abilities of individuals. A "ceiling" effect is present when more than 20% of the sample has the ability to complete all items of the instrument, and a "floor" effect is present when more than 20% of the sample do not have the ability to complete any of the items.[26,28]

Ethical Considerations

This study was approved by the Ethics Committee in Research Involving Human Beings of the Federal University of

Table 1. *Sociodemographic and Clinical Characterization of the Group of Pregnant Women Diagnosed With and Without Pelvic Girdle Pain for the Analysis of Measurement Properties*

Characteristic	Group With Pain (n = 105)	Group Without Pain (n = 52)
Age, y, mean (SD)	24.94 (4.97)	26.12 (4.47)
Pregnancies		
1	35 (33.3)	20 (38.5)
2	34 (32.4)	13 (25)
3	20 (19)	10 (19.2)
>3	16 (15.3)	9 (17.2)
Gestational week, mean (SD)	27.88 (6.19)	27.83 (5.96)
Classification of pain		
Unilateral sacroiliac syndrome	56 (53.3)	N/A
Bilateral sacroiliac syndrome	36 (34.3)	N/A
Pelvic girdle syndrome	12 (11.4)	N/A
Symphysiolysis	1 (1)	N/A
Civil status		
Single	53 (50.5)	27 (51.9)
Married	52 (49.5)	25 (48.1)
Years of study		
<12 y	34 (32.4)	13 (25)
≥12 y	71 (67.6)	39 (75)
Current occupation		
Employed	29 (27.6)	14 (26.9)
Unemployed	76 (72.4)	38 (73.1)
Family income,[a] mean (SD)		
<US $31 800	17 (16.2)	12 (23.1)
≥US $31 800	88 (83.8)	40 (76.9)
PGQ score, mean (SD)		
Total	36.97 (19.36)	3.42 (5.09)
Activities subscale	35.04 (18.99)	3.40 (5.32)
Symptoms subscale	44.02 (25.32)	3.57 (9.17)

PGQ, Pelvic Girdle Questionnaire; *SD,* standard deviation.
Values are expressed as the number (%) unless otherwise noted.
[a] Minimum wage (July 2014) equivalent to US $31 800.

Table 2. *Description of Hypotheses Tested for Construct Validity in Groups of Pregnant Women With and Without Diagnosed Pelvic Girdle Pain*

Hypothesis	Score, Mean (SD)		Average Difference, Mean (SD)	95% CI	Hypothesis Confirmed?
	Group With Pain	Group Without Pain			
The mean score on the Activities subscale for pregnant women with pain will be greater than that for pain-free pregnant women.	35.04 (18.99)	3.40 (5.32)	31.63 (2.68)	26.32-36.95	Yes
The mean score on the Symptoms subscale for pregnant women with pain will be greater than that for pain-free pregnant women.	44.02 (25.32)	3.57 (9.17)	40.44 (3.62)	33.27-47.61	Yes
The mean total score for pregnant women with pain will be greater than that for pain-free pregnant women.	36.97 (19.36)	3.42 (5.09)	33.54 (2.73)	38.94-28.13	Yes

The instrument can distinguish between pregnant women with pain and pregnant women without pain reaching a mean difference in the total score pre-established in the pilot study of 34.76 ± 7.40% within the 95% CI 50.26-19.25. *CI*, confidence interval; *SD*, standard deviation.

Pernambuco under Protocol CAAE 07215712.3.0000.5208 of the Health Sciences Center. All participants were informed about the study objectives, and those who wished to participate signed the informed consent form.

Results

The pilot study included 21 pregnant women divided into 2 groups: a group of 12 pregnant women diagnosed with PGP and a group of 9 pregnant women without the diagnosis of PGP.

For the group of pregnant women with PGP, the mean age was 26 (standard deviation [SD] = 5.2) years, 66.7% were married, 41.7% had <12 years of education, 66.7% were unemployed, 58.3% had a family income above US $31 800, 16.7% were primiparous, and mean duration of gestation was 35.4 weeks (SD = 4.54). Unilateral sacroiliac pain was the most common clinical condition (58.3%). The mean total score on the PGQ-Brazil for these women with pain was 52.52%.

For the pregnant women without pain, the mean age was 26.8 (SD = 5.8) years, 88.9% were married, 11.1% had <12 years of education, 66.7% were unemployed, 33.3% had a family income above US $31 800, 44.4% were primiparous, and mean duration of gestation was 27.4 weeks (SD = 8.6). The total mean score on the PGQ-Brazil for this group was 17.75%. The results obtained indicate even the women without pain experienced discomfort during activities, but the questionnaire was sensitive enough to detect differences between the 2 groups.

For the analysis of measurement properties, the sample consisted of 157 pregnant women, 105 diagnosed with PGP and 52 with no complaints of pain. Pregnant women without pain participated only in the construct validity step (hypothesis testing). The characteristics of these participants are summarized in Table 1.

Implementation of the PGQ-Brazil in the different groups (pregnant women with pain and those without pain) revealed that all hypotheses formulated during the pilot study were confirmed (Table 2).

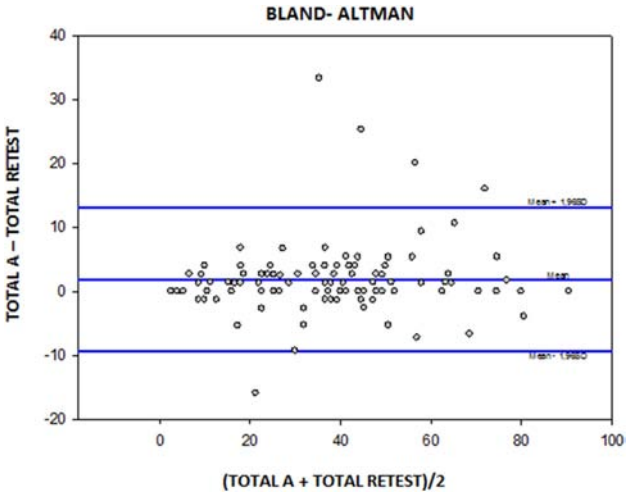

Fig 1. *Bland-Altman plot of test-retest measurement error.*

The ICC for the reliability analysis of the Activities subscale was 0.97 (95% CI = 0.95-0.98), and that for the Symptoms subscale was 0.98 (95% CI = 0.9-0.98). The ICC for the total score on the PGQ-Brazil was 0.97 (95% CI = 0.96-0.98). Each of the 25 PGQ-Brazil items was analyzed using the κ coefficient, and 2 items had κ values <0.70 (item 9 [0.57] and item 11 [0.50]).

Clinical stability achieved with the average visual analogue scale score was 6.59 cm (SD = 1.85) for the test and 6.58 cm (SD = 1.86) for the retest. There were no losses between the test and retest. To determine agreement and measurement error between the test and retest, a Bland-Altman plot was constructed (Fig 1). The agreement parameters were 13.1 (95% CI = 11.22-15.07) and –9.4 (95% CI = –11.2 to –7.47). However, the SEM calculated for the test-retest considering the total score was 3.83, with an MMD_{ind} of 10.80 (Table 3).

Rasch analysis calibration results, MnSq values, and *Z* values for each of the 25 items of the instrument are listed in

Table 3. *Test-Retest Average, SEM, MDC_{ind}, MDC_{group}, and ICC*

	Average (SD)						
Instrument	Test	Retest	Difference	SEM	MDC_{ind}	MDC_{group}	ICC (95% CI)
Entire scale	38.8 (19.9)	36.9 (19.3)	1.87 (5.7)	3.83	10.80	1.05	0.97 (0.96-0.98)
Activities subscale	35.0 (18.9)	37.2 (19.6)	2.2 (6.4)	5.95	16.79	1.63	0.97 (0.95-0.98)
Symptoms subscale	44.0 (25.3)	45.4 (25.7)	1.4 (6.5)	6.16	17.38	1.69	0.98 (0.97-0.99)

CI, confidence interval; *ICC,* intraclass correlation coefficient; MDC_{group}, group minimal detectable change; MDC_{ind}, individual minimal detectable change; *SD*, standard deviation; *SEM,* standard error of measurement.

Table 4. The average calibration of participants calibration was –0.46 (SD = 0.77).

Item 11 on the Activities subscale ("How hard is it for you to carry light objects?") was considered the easiest; that is, it had the lowest crude score (9) and a calibration of 2.29 logits. On the Symptoms subscale, item 23 ("To what extent will your legs fail?") was the easiest (crude score = 88) with a calibration of 0.69 logits. Items 06 ("How hard is it for you to sit for more than 60 minutes?") on the Activities subscale and item 22 ("How much pain do you feel at night?") on the Symptoms subscale were considered the most difficult, because these items had the highest crude scores (196 and 200) and calibrations of –0.89 and –0.88 logits, respectively (Table 4). Only item 5 ("How hard is it for you to sit for less than 10 minutes?") had an erratic pattern of response, with a residual adjustment interval –1.5 to 1.5.

The sample separation index was 3.85, which was calculated as the number of distinct levels ([4*G* + 1]/3), indicating that those items separated people into at least 3 of disability. The estimated calibration stability obtained was 0.89.

Analysis of the questionnaire by dividing it into 2 subscales (Activities and Symptoms) revealed that the Activities subscale comprising 20 items had a participant separation index of 3.38 with an estimated calibration stability of 0.84. In contrast, for the Symptoms subscale, comprising 5 items, the participant separation index was 1.16 with an estimated calibration stability of 0.57.

Figure 2 is a map of the difficulty level of items with respect to functionality of individuals. This map, based on calibration of the instrument items, illustrates the functionality of the continuous sample on the left and the continued difficulty of the items on the right. At the top of the continuum are some people for whom no items aligned to the right are present, whereas at the bottom is a very easy item evaluating 1 person. Pregnant women are on the left. Symbols relate to the number of pelvic joints affected: # = 1 joint, @ = 2 joints, § = 3 joints. It is observed that 25 people are arranged at the lower end of the continuum, corresponding to 24% of the sample, and indicating the presence of the "ceiling" effect.

It was also established by Rasch analysis that PGQ-Brazil response scores (0-3) were arranged in ascending order of difficulty, where 0 was considered the easiest score (–1.42), followed by 1 (–0.37), 2 (0.38), and 3, the most difficult (1.41).

Discussion

The PGQ-Brazil was specifically developed to assess limitations in activities and symptoms in pregnant women affected by pelvic girdle pain. The results of this study indicate the good test-retest reliability, internal consistency, and construct validity of the PGQ-Brazil. The original PGQ was developed for use as a self-report form. However, during the Delphi study phase, it was suggested that the PGQ be applied in an interview, considering the sociocultural context of Brazil.

Although this study reports test-retest reliability scores higher than 0.70 as recommended in the literature,[21] it is not possible to compare these findings with the results of the original instrument,[18] which considered only the subscales separately (20 items of the subscale of activities and 5 of the subscale of symptoms) and not the entire instrument. This made it difficult to compare the 2 studies. However, the reliability scores of the subscales were similar to that for the entire instrument, indicating good reliability between the test and retest.

Equally, the test-retest correlation analyzed with the Bland-Altman plot was in good agreement, indicating that the scores obtained in the first assessment are consistent with those for the second assessment. Despite the higher and lower limits of agreement indicating greater variation than expected, the ICC values in Table 3 are quite high and the SEM values are quite low. These figures are indicative of good test-retest reliability, indicating a low probability of random and systematic error. It is expected that an SEM of 3.83 indicates only 1 possible measurement error and not an improvement or worsening of the patient between the test and retest. Complementing this, a MMD_{ind} of 10.80, being higher than the SEM, also indicates that there has been no change in the measured parameter, but simply a measurement error, thus confirming the results observed in the total score and Activities and Symptoms subscale scores of the original study.[19]

From the construct validity results (hypothesis testing), we observed that the PGQ-Brazil, in its 2 subscales, has the

Table 4. *Calibration Distribution, Z, and MnSq Values for Items in the Brazilian Version of the Pelvic Girdle Questionnaire*

			Infit		Outfit	
Item [a]	Calibration	Error	MnSq	*Z*	MnSq	*Z*
Activities subscale						
11. Carry light objects	2.24	0.31	1.06	0.3	0.36	–1.3
14. Push a shopping cart	0.90	0.15	1.03	0.2	0.85	–0.4
07. Walk for less than 10 min	0.88	0.15	1.24	1.1	1.04	0.2
05. Sit for less than 10 min	0.70	0.13	1.48	2.2	1.61	1.9
02. Stand for less than 10 min	0.59	0.13	0.99	0.0	0.94	–0.1
20. Push something with a foot	0.53	0.12	1.14	0.8	1.03	0.2
01. Get dressed	0.12	0.11	0.89	–0.8	0.83	–0.8
13. Stand up/sit down	0.01	0.10	0.89	–0.9	0.83	–0.8
17. Lie down	–0.05	0.10	0.10	0.0	0.88	–0.6
19. Have a normal sex life	–0.05	0.11	1.18	1.3	1.32	1.5
04. Bend down	–0.23	0.10	1.01	0.2	0.93	–0.3
10. Do housework	–0.36	0.10	0.91	–0.8	0.81	–1.1
09. Climb stairs	–0.50	0.10	0.99	0.0	1.02	0.2
15. Run	–0.65	0.10	0.75	–2.4	0.68	–2.1
18. Roll over in bed	–0.71	0.10	0.82	–1.6	0.82	–1.1
12. Carry heavy objects	–0.81	0.10	1.22	1.8	1.28	1.5
08. Walk for more than 60 min	–0.83	0.10	0.97	–0.2	0.84	–0.9
03. Stand for more than 60 min	–0.88	0.10	1.15	1.2	1.39	2.0
06. Sit for more than 60 min	–0.89	0.10	1.18	1.4	1.24	1.3
Symptoms subscale						
23. Do your leg(s) fail?	0.69	0.13	1.09	0.7	1.30	1.4
21. How much pain do you feel in the morning?	0.26	0.12	1.03	0.3	0.98	0.0
25. Is your sleep interrupted?	–0.04	0.11	1.18	1.3	1.04	0.3
24. Do you do things more slowly?	–0.08	0.11	0.87	–1.0	0.96	–0.2
22. How much pain do you feel at night?	–0.83	0.12	0.87	–0.9	0.84	–1.0

MnSq, mean square value.
[a] Back-translated to English from the Portuguese version.

ability to discriminate women diagnosed with PGP from women without a PGP diagnosis. Thus, it was found that pregnant women affected by PGP tend to have higher scores because of their level of disability compared with pregnant women who had no pain complaints. The same result was obtained in the original instrument, which

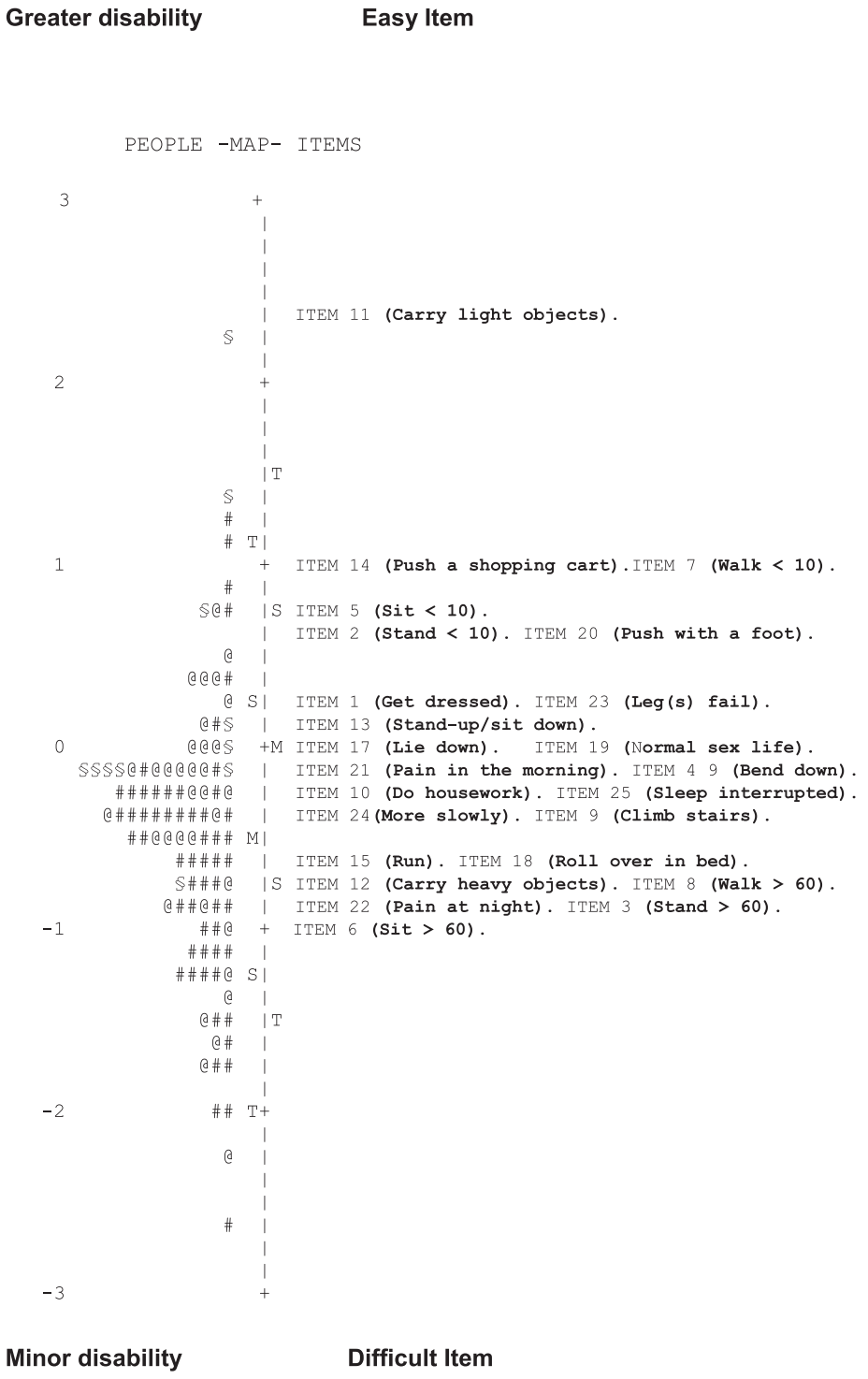

Fig 2. *Map of the distribution of representative individuals and items according to levels of difficulty and ability. #, Pregnant with 1 affected joint; @, pregnant with 2 affected joints; §, pregnant with 3 affected joints.*

indicated the ability of PGQ to distinguish pregnant and postpartum women with PGP, as well as the number of joints affected.[19]

As in the original study, there was an appropriate coefficient of reliability when estimated calibration stability was analyzed separately for individuals by Rasch analysis, considering the 25 items of the entire PGQ-Brazil and the 20 items of the Activities subscale. This finding ensures that measures can be reproduced in subsequent applications of the instrument, and indicates

the possibility that the Activities subscale could function independently. However, the Symptoms subscale did not have good reliability, indicating that it could not be used independently.[18]

It is possible to ensure from the sample separation index that the 25 items of the PGQ-Brazil have an excellent level of separation.[30] Thus, the greater the index, the more distinct the levels will be, suggesting that the responders are divided into at least 3 levels of disability: low, moderate, and high.

Despite the ability of the questionnaire to differentiate responders, 24% of the participants were found to be distributed at the bottom end of the continuum, indicating a ground effect. This result reveals that the instrument may not be able to assess pregnant women with higher capacity. In considering these findings in the light of the results of the analysis of the original instrument, we observed that this did not identify floor and ceiling effects. However, it is noteworthy that there were differences in the populations in which the instruments were applied. This study included pregnant women at low risk with no restriction on daily activities, whereas in the original study consisted of pregnant and puerperal postpartum women receiving care in physical therapy clinics and, therefore, more severely afflicted.[18]

Similarly, the different characteristics of the samples may be related to the differences between the PGQ and the original PGQ-Brazil with respect to the calibration sequence of the Activities subscale items. Although not following the same calibration sequence, 4 items considered easy on the PGQ-Brazil are among the first 5 easy items from the original PGQ Activities subscale, whereas only 2 difficult items on the PGQ-Brazil are among the 5 most difficult on the original instrument. However, both the calibration of the Symptoms subscale and the sequential increase in difficulty of response options (0-3) were equal to those in the original study.[18]

Only item 5 ("How hard is it for you to sit for 10 minutes?") elicited erratic responses, in 4% of the sample. Furthermore, the identified remainders (residues, ie, out-of-normal values for MnSq and Z) are within the recommended parameters recommended in the literature,[29] also ensuring that the instrument measures a unidimensional concept.

We emphasize that although some terms of the original instrument were modified during the back-translation because of the need for cultural adaptation, we do not believe that these changes influenced the content of the questionnaire.

Limitations

The absence of interexaminer assessment and responsiveness are study limitations. Another limitation is the failure to include other sections of the population. There is a need for further studies to analyze the interexaminer reliability and responsiveness of the PGQ-Brazil, as well as interpretability as an additional tool for the instrument. Other populations, such as pregnant adolescents and women with high-risk pregnancies, should also be studied.

Conclusion

The PGQ-Brazil was determined to have adequate test-retest reliability, internal consistency, and construct validity (hypothesis testing), which allowed for discrimination between women diagnosed with PGP and women without a PGP diagnosis. In this way, it stands out as a specific tool for assessing the functioning of Brazilian pregnant women diagnosed with PGP related to pregnancy.

Funding Sources and Conflicts of Interest

The Coordination for the Improvement of Higher Level Personnel (CAPES) provided a scholarship for this study. No conflicts of interest were reported for this study.

Contributorship Information

Concept development (provided idea for the research): L.S., G.L., A.L.
Design (planned the methods to generate the results): L.S., G.L., A.L.
Supervision (provided oversight, responsible for organization and implementation, writing of the manuscript): L.F.T.-S., B.S., A.L.
Data collection/processing (responsible for experiments, patient management, organization, or reporting data): L.S., R.B.
Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): L.S., L.F.T.-S., L.M.
Literature search (performed the literature search): L.S., E.W., R.B.
Writing (responsible for writing a substantive part of the manuscript): L.S., A.L.
Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): L.S., L.F.T.-S., B.S., A.L.

References

1. Wu WH, Meijer OG, Uegaki K, et al. Pregnancy-related pelvic girdle pain (PPP), I: terminology, clinical presentation, and prevalence. *Eur Spine J.* 2004;13(7):575-589.
2. Hilde G, Stuge B, Vøllestad N. Physical therapy for pregnancy-related low back and pelvic pain: a systematic review. *Acta Obstet Gynecol Scand.* 2003;82(11):983-990.
3. Saccomanni B. Low back pain associated with pregnancy: a review of literature. *Eur Orthop Traumatol.* 2011;1(5):169-174.

Practical Applications

- The PGQ-Brazil has good test-retest reliability, internal consistency, and construct validity.
- The minimal detectable change results indicated the good sensitivity of the instrument in identifying possible changes in the study population.
- The ground effect revealed that the instrument may not be able to assess pregnant women with higher capacity.

4. Robinson HS, Mengshoel AM, Bjelland EK, Vøllestad NK. Pelvic girdle pain, clinical tests and disability in late pregnancy. *Man Ther.* 2010;15(3):280-285.
5. O'Sullivan PB, Beales DJ. Diagnosis and classification of pelvic girdle pain disorders–part 1: a mechanism based approach within a biopsychosocial framework. *Man Ther.* 2007;12(2):86-97.
6. Vleeming A, Albert HB, Ostgaard HC, Sturesson B, Stuge B. European guidelines for the diagnosis and treatment of pelvic girdle pain. *Eur Spine J.* 2008;17(6):794-819.
7. Ostgaard HC, Zetherström G, Roos-Hansson E. The posterior pelvic pain provocation test in pregnant women. *Eur Spine J.* 1994;3(5):258-260.
8. Kanakaris NK, Roberts CS, Giannoudis PV. Pregnancy-related pelvic girdle pain: an update. *BMC Med.* 2011;9:15.
9. Ostgaard HC, Andersson GB, Wennergren M. The impact of low back and pelvic pain in pregnancy on the pregnancy outcome. *Acta Obstet Gynecol Scand.* 1991;70(1):21-24.
10. Johansson G, Norén L, Ostgaard HC, Ostgaard S. Lumbar back and posterior pelvic pain during pregnancy: a 3-year follow-up. *Eur Spine J.* 2002;11(3):267-271.
11. Bastiaanssen JM, de Bie RA, Bastiaenen CH, Essed GG, van den Brandt PA. A historical perspective on pregnancy-related low back and/or pelvic girdle pain. *Eur J Obstet Gynecol Reprod Biol.* 2005;120(1):3-14.
12. Gutke A, Ostgaard HC, Oberg B. Pelvic girdle pain and lumbar pain in pregnancy: a cohort study of the consequences in terms of health and functioning. *Spine (Phila Pa 1976).* 2006;31(5):E149-E155.
13. Damen L, Buyruk HM, Güler-Uysal F, Lotgering FK, Snijders CJ, Stam HJ. Pelvic pain during pregnancy is associated with asymmetric laxity of the sacroiliac joints. *Acta Obstet Gynecol Scand.* 2001;80(11):1019-1024.
14. Mens JM, Vleeming A, Snijders CJ, Ronchetti I, Ginai AZ, Stam HJ. Responsiveness of outcome measurements in rehabilitation of patients with posterior pelvic pain since pregnancy. *Spine (Phila Pa 1976).* 2002;27(10):1110-1115.
15. Mogren IM. Physical activity and persistent low back pain and pelvic pain post-partum. *BMC Public Health.* 2008;8:417.
16. Padua L, Padua R, Bondi R, et al. Patient-oriented assessment of back pain in pregnancy. *Eur Spine J.* 2002;11(3):272-275.
17. Wedenberg K, Moen B, Norling A. A prospective randomized study comparing acupuncture with physiotherapy for low-back and pelvic pain in pregnancy. *Acta Obstet Gynecol Scand.* 2000;79(5):331-335.
18. Stuge B, Garratt A, Jenssen HK, Grotle M. The pelvic girdle questionnaire: a condition-specific instrument for assessing activity limitations and symptoms in people with pelvic girdle pain. *Phys Ther.* 2011;91(7):1096-1108.
19. Grotle M, Garratt AM, Jenssen HK, Stuge B. Reliability and construct validity of Self-report questionnaires for patients with pelvic girdle pain. *Phys Ther.* 2012;92(1):111-123.
20. Simões LCF, Teixeira-Salmela LF, Wanderley ELS, et al. Adaptação transcultural do "Pelvic Girdle Questionnaire" (PGQ) para o Brasil. *Acta Fisiatr.* 2016;23(4):166-171.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60(1):34-42.
22. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to Questionnaire development, evaluation, and refinement. *Qual Life Res.* 2007;16(suppl 1):5-18.
23. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539-549.
24. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22.
25. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology and definitions of measurements properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737-745.
26. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
27. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med.* 2003;35(3):105-115.
28. Miller KL, Slade AL, Pallant JF, Galea MP. Evaluation of the psychometric properties of the upper limb subcales of the motor assessment scale using a Rasch analysis model. *J Rehabil Med.* 2010;42(4):315-322.
29. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health.* 2004;7(suppl 1):S22-S26.
30. Duncan PW, Bode RK, Min Lai S, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil.* 2003;84(7):950-963.