The Pelvic Girdle Questionnaire: A Condition-Specific Instrument for Assessing Activity Limitations and Symptoms in People With Pelvic Girdle Pain

Britt Stuge, Andrew Garratt, Hanne Krogstad Jenssen, Margreth Grotle

B. Stuge, PT, PhD, Department of Orthopaedics, Oslo University Hospital, Kirkeveien 166, NO-0407 Oslo, Norway. Address all correspondence to Dr Stuge at: britt.stuge@medisin.uio.no.

A. Garratt, PhD, National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

H. Krogstad Jenssen, PT, Department of Physical Medicine and Rehabilitation, Oslo University Hospital.

M. Grotle, PT, PhD, FORMI (Communication Unit for Musculoskeletal Disorders), Oslo University Hospital, and National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital.

[Stuge B, Garratt A, Krogstad Jenssen H, 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;91:1096–1108.]

Background. No appropriate measures have been specifically developed for pelvic girdle pain (PGP). There is a need for suitable outcome measures that are reliable and valid for people with PGP for use in research and clinical practice.

Objective. The objective of this study was to develop a condition-specific measure, the Pelvic Girdle Questionnaire (PGQ), for use during pregnancy and postpartum.

Design. This was a methodology study.

Methods. Items were developed from a literature review and information from a focus group of people who consulted physical therapists for PGP. Face validity and content validity were assessed by classifying the items according to the World Health Organization's *International Classification of Functioning, Disability and Health*. After a pilot study, the PGQ was administered to participants with clinically verified PGP by means of a postal questionnaire in 2 surveys. The first survey included 94 participants (52 pregnant), and the second survey included 87 participants (43 pregnant). Rasch analysis was used for item reduction, and the PGQ was assessed for unidimensionality, item fit, redundancy, and differential item functioning. Test-retest reliability was assessed with a random sample of 42 participants.

Results. The analysis resulted in a questionnaire consisting of 20 activity items and 5 symptom items on a 4-point response scale. The items in both subscales showed a good fit to the Rasch model, with acceptable internal consistency, satisfactory fit residuals, and no disordered threshold. Test-retest reliability showed high intraclass correlation coefficient estimates: .93 (95% confidence interval=0.86–0.96) for the PGQ activity subscale and .91 (95% confidence interval=0.84–0.95) for the PGQ symptom subscale.

Limitations. The PGQ should be compared with low back pain questionnaires as part of a concurrent evaluation of measurement properties, including validity and responsiveness to change.

Conclusions. The PGQ is the first condition-specific measure developed for people with PGP. The PGQ had acceptably high reliability and validity in people with PGP both during pregnancy and postpartum, it is simple to administer, and it is feasible for use in clinical practice.

Pregnancy-related back pain and pelvic girdle pain (PGP) are common across countries, irrespective of socioeconomic factors.[1] In general, the literature does not make a clear distinction between pregnancy-related low back pain (LBP) and PGP. There is, however, growing evidence that PGP disorders comprise a distinct subgroup with a unique clinical presentation and a need for specific management.[1–5] According to European guidelines, PGP generally arises in relation to pregnancy and pain is experienced between the posterior iliac crest and the gluteal fold, particularly in the vicinity of the sacroiliac joints. The diagnosis of PGP can be reached after the exclusion of lumbar causes, and the pain or functional disturbances must be reproducible by specific clinical tests.[1]

For adequate evaluation of interventions, reliable and valid outcome measures are needed. It is not clear whether available measures are entirely appropriate for use in patients with PGP in clinical practice and research.[6] There are clinical grounds for classifying LBP and PGP as 2 different conditions; therefore, outcome measures that have been validated for LBP are not necessarily the most appropriate ones for PGP. It has also been reported that back pain increases during pregnancy and seems to differ from that in the general population.[7] The Pregnancy Mobility Index was developed to reflect these concerns, but the activity items were based on literature research and clinical experience and did not include the views of patients.[7] Current outcome measures may not adequately capture the specific problems and consequences that patients describe, and there is a discrepancy between patients' scores and their feedback. Thus, there is a need for an outcome measure that is reliable, valid, and responsive to change for patients with PGP.[1] In addition to research applications, such as randomized trials, health care professionals require appropriate measures that are feasible for use in clinical practice.

Disability and functioning are core issues in the assessment and treatment of patients with PGP, and items comprising outcome measures must reflect these important domains. Item response theory and Rasch analysis have been increasingly applied in the field of patient-reported outcomes for assessment of the unidimensionality (the extent to which items measure a single construct, such as pelvic function), item difficulty (the relative difficulty of items when compared with one another), and person separation (the extent to which items distinguish between distinct levels of functioning) of a measure.[8] This study was designed to develop a condition-specific measure, the Pelvic Girdle Questionnaire (PGQ), for use both during pregnancy and postpartum in research and clinical practice. After the application of Rasch analysis, the final questionnaire was assessed for test-retest reliability and construct validity.

Method

Figure 1 shows the various stages in the development and testing of the PGQ with different groups of people. Following is a description of the stages.

Development of the PGQ

A structured literature search of databases, including CINAHL and MEDLINE, revealed no existing condition-specific measures for patients with PGP; therefore, development work was undertaken. In September 2008, pregnant and postpartum patients with PGP were invited to take part in a focus group that was semistructured and led by the primary author (B.S.). Items that were included in frequently used outcome measures (the Roland-Morris Disability Questionnaire,[9] the Oswestry Disability Index,[10] and the Disability Rating Index[11]), that were related to the pelvic girdle areas, and that were considered clinically relevant for patients with PGP were added to those obtained from the focus group. Physical therapists working

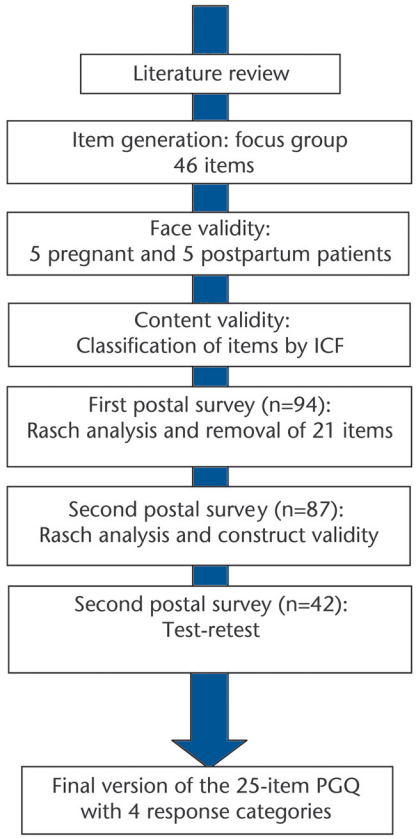

Figure 1.
Stages of the study. ICF=*International Classification of Functioning, Disability and Health*. The Norwegian version of the Pelvic Girdle Questionnaire (PGQ) was tested.

with patients with PGP were asked to record activities that they considered important for patients with PGP. The process of item development was designed to lend the new measure content validity and face validity.

Face validity and content validity were assessed by classifying the items according to the World Health Organization's *International Classification of Functioning, Disability and Health* (ICF),[12] and the initial version of the PGQ was assessed in a pilot study of 5 pregnant and 5 postpartum participants. A 5-point scale with the responses "not at all," "to a limited extent," "to a moderate extent," "to a large extent," and "to a very large extent" was used for the items.

Psychometric Testing: Validity and Reliability

The participants included in this study were recruited consecutively by physical therapists working with patients with PGP in Oslo, Norway. All potential participants were examined by a physical therapist and assessed with the following recommended inclusion criteria[1,2]: PGP located distal, lateral, or both in relation to the L5–S1 area in the buttocks, in the symphysis, or both, with pain onset during pregnancy or within 3 weeks after delivery. Fulfillment of the diagnostic criteria was based on the following tests: posterior pelvic pain provocation test, active straight-leg-raising test, pain provocation of the long dorsal sacroiliac ligament, and pain provocation of the symphysis by palpation and by a modified Trendelenburg test. The posterior pelvic pain provocation test or the active straight-leg-raising test had to be positive on the right side, left side, or both sides, and at least 1 of the other 3 tests had to be positive. Participants were asked to complete the initial version of the PGQ at home and return it along with sociodemographic information and the Patient-Specific Functional Scale in a prepaid reply envelope.[13] Rasch analysis was used for item reduction of the initial version of the PGQ.

After the pilot study in autumn 2008, the Norwegian version of the PGQ was administered by means of a postal questionnaire to 94 participants with PGP (52 pregnant and 42 not pregnant). The PGQ was revised and included in a second postal survey (spring 2009) of 87 participants with PGP (43 pregnant and 44 not pregnant). With the exception of 13 postpartum participants who also completed the first questionnaire, the participants in the second survey were new participants. Test-retest reliability was assessed with a random sample of 42 participants (21 pregnant and 21 not pregnant) who completed the revised questionnaire within 1 week. Finally, 5 participants with PGP were asked to comment on the final version of the questionnaire, including the time needed to complete it. Written informed consent was obtained from all participants.

Statistical Analysis

Items for which a third of the data or more were missing or for which "not applicable" responses were made were considered for removal from the questionnaire. Data were analyzed with SPSS (version 14)* and RUMM2020 (www.rummlab.com).

The questionnaire was assessed for unidimensionality, item fit, redundancy, and differential item functioning (DIF) with the Rasch model,[8] which implements an unconditional maximum-likelihood procedure. This model assumes that the probability of a person affirming a trait, for example, functioning, in an item of a measurement depends on the person's level of that trait (θ) and the level of functioning required by the item (b). A model with several response categories is expressed as follows[8]:

$$ln(P_{ni}/1 - P_{ni}) = \theta_n - b_i - \tau_i.$$

In this equation, P_{ni} is the probability that a person n will affirm the item, θ_n is the person's level of the trait, b_i is the level of the trait expressed by the item, and τ_i represents the .5 probability point (threshold) between adjacent response levels for that item. In this case, it is assumed that as a person's disability increases, the probability of a maximum score on the item increases.

Participants and item scores were used to "calibrate" items on a logit scale, which indicated difficulty level. Items at one end of the scale were "easier," and items at the other end were more "difficult." The difficulty of individual items was determined by the frequency of endorsement. In the present analysis, items with a negative calibration were less difficult, that is, indicated less disability. A logit is the natural log-odds of the level of difficulty of a particular item in relation to all other items in the scale (hierarchy).

Rasch analysis also constructs a hierarchy of the respondents ordered by their level of disability. *Unidimensionality* refers to the single underlying construct assessed by items that form a scale within a measure or questionnaire. The adequacy of the fit of each item to the Rasch model is assessed by the overall model fit, individual person fit, item fit, thresholds of the response categories, DIF and, finally, a check for local independence (multidimensionality).

The likelihood ratio test was used to assess whether a rating scale model or an unrestricted model was appropriate. The overall model fit for the

* SPSS Inc, 233 S Wacker Dr, Chicago, IL 60606.

scale was given by Bonferroni-adjusted chi-square item-trait interaction statistics. A nonsignificant probability value indicated that there was no substantial deviation from the model and that the hierarchical ordering of the items was consistent across all levels of the underlying trait.

The person separation reliability index, which is equivalent to the Cronbach alpha, indicated the power of a scale to discriminate among respondents with different levels of the trait being measured; values of .8 and .9 indicated that the scale could statistically discriminate among at least 2 and 3 groups, respectively.

Misfit of the model can be due to both misfitting respondents and misfitting items. The individual respondent fit and item fit were assessed by inspecting the means and standard deviations of the separate fit residuals. If the items and respondents fit the model, a mean value of 0 and a standard deviation of 1 were expected. Misfitting items were identified by fit residuals of greater than ± 2.5 or a significant chi-square probability value.[8] For each item, the ordered set of response thresholds and the category probability curves were inspected. Disordered thresholds were corrected by rescoring the items.

For the assessment of potential bias across groups of respondents, DIF was analyzed for the final version of the questionnaire in relation to age (<30, 30–37, or ≥ 38 years), pregnancy (yes or no), work status (employed or not), and level of education (≤ 12 years or >12 years). Two types of DIF, uniform and nonuniform, could be identified. Uniform DIF indicated that 1 subgroup displayed a consistently greater ability to confirm an item than another subgroup (analysis of variance [ANOVA] main effect). Nonuniform DIF indicated that the ability differences among the subgroups were inconsistent (ANOVA interaction effect).

Principal components analysis of 2 subsets of the residuals was used to assess potential multidimensionality. The absence of any meaningful pattern in the residuals supported the assumption of local independence and unidimensionality of the scale. The component loadings of 2 subsets of items were compared, and a paired *t* test was used to determine whether the associated person estimates were significantly different from that derived from all of the items. Differences in the person estimates between the subsets and the full-item scale would indicate a breach of the assumption of local independence and unidimensionality.

Test-retest reliability was assessed with the intraclass correlation coefficient.[14] An intraclass correlation coefficient of .80 or more was considered to be evidence of test-retest reliability. Repeatability was assessed by calculating the standard error of measurement between test and retest scores (calculated as the square root of the mean square error or residual in the ANOVA table), and the 95% confidence interval for repeatability was calculated by multiplying the standard error of measurement by 2.77.[14] This value is the smallest amount of change between 2 scores that must be observed before the change can be considered to exceed the measurement error and is frequently referred to as the minimum detectable change.[14]

Role of the Funding Source

Grant support was provided by The Norwegian Fund for postgraduate education in physiotherapy.

Results

Development of the PGQ

The focus group and physical therapists working with patients with PGP provided 46 items for the initial Norwegian version of the PGQ. The pilot study resulted in minor changes of wording in the difficulty scaling. Table 1 shows the 46 items classified according to the ICF. The majority reflected different domains of the activity/participation component of the ICF, whereas 6 items were related to body functions. Because of the differences in these theoretical constructs, the domains were analyzed separately.

Data Collection

All of the people invited to participate in the 2 surveys responded. A total of 94 people responded to the first survey, and 87 responded to the second survey. The characteristics of the participants are shown in Table 2. The mean age was 34 years (SD=5.6) in the first survey and 35 years (SD=5.0) in the second survey. The initial 46-item measure had low levels of missing data, and none of the participants reported the items to be not applicable.

Statistical Analysis: First Survey

In the initial analysis of the 40 activity items, the likelihood ratio test was highly significant ($P<.0000001$); thus, the unrestricted model was applied. The first Rasch analysis indicated that the 40 initial items were not unidimensional because there was a misfit with the Rasch model (chi-square value of 128.9, with Bonferroni-adjusted level of significance of $P=.0004$). The initial person separation reliability index was high (.95). Eight participants had extreme scores outside the fit residual limits of ± 2.5, and 1 item ("travel by car") was redundant (fit residual=2.91). Several of the items (13/40) had disordered thresholds, suggesting that for many of the questions, the participants were not able to dis-

Table 1.
Initial 46 Items Classified According to the *International Classification of Functioning, Disability and Health* (ICF) and Results of First Rasch Analysis[a]

Item Description	Item No.	Item Difficulty (Logits)	Standard Error for Location	Fit Residual	Probability	Disordered Thresholds	Reason for Exclusion
Activity/participation component							
Dress yourself	1	1.984	0.12	0.262	0.58575		
Stand <10 min	2	1.35	0.135	0.124	0.885801		
Stand >10 min	3	0.237	0.118	0.931	0.918673		Location similar to that of other item
Stand >30 min	4	−1.168	0.127	0.135	0.791615		Location similar to that of other item
Stand >60 min	5	−2.522	0.164	−0.706	0.315443		
Bend down	6	−0.332	0.108	2.143	0.013287		
Bend forward and rotate	7	−0.349	0.113	0.916	0.175393		Redundant in second analysis
Pick up object from the floor	8	0.065	0.117	−0.77	0.274219	Yes	Location similar to that of other item
Sit <10 min	9	2.472	0.158	1.745	0.00363	Yes	
Sit >10 min	10	1.026	0.119	0.644	0.151281	Yes	
Sit >30 min	11	0.391	0.115	1.161	0.033629		Location similar to that of other item
Sit >60 min	12	−0.824	0.12	1.778	0.003841		Location similar to that of other item
Walk <10 min	13	2.318	0.136	−0.79	0.249745		
Walk >10 min	14	0.719	0.109	−0.312	0.663528		Location similar to that of other item
Walk >30 min	15	−0.28	0.113	−0.275	0.222125	Yes	Location similar to that of other item
Walk >60 min	16	−1.248	0.132	−0.779	0.019065		
Climb stairs	17	0.188	0.122	0.158	0.68767		
Perform light work	18	0.709	0.134	−1.523	0.037681	Yes	Location similar to that of other item
Perform heavy work	19	−1.232	0.141	−1.413	0.117438	Yes	Location similar to that of other item
Vacuum	20	−0.866	0.113	0.891	0.163026		Location similar to that of other item
Clean floor	21	−1.047	0.124	−0.664	0.902375	Yes	
Make a bed	22	−0.123	0.105	0.392	0.767503	Yes	Location similar to that of other item
Carry light objects	23	1.096	0.138	−0.019	0.531649	Yes	
Carry heavy objects	24	−1.045	0.135	−0.078	0.078765	Yes	
Lift heavy objects	25	−1.693	0.16	−0.455	0.171335		
Squat	26	−0.022	0.113	0.167	0.705647		Location similar to that of other item
Sit on the floor	27	−0.45	0.11	1.286	0.205488	Yes	
Get up/sit down	28	0.236	0.117	0.744	0.937505		
Travel by car	29	1.157	0.134	2.908	0.085447		Redundant in first analysis
Travel by bus	30	0.542	0.116	0.505	0.495207		Location similar to that of other item
Push a shopping cart	31	0.882	0.118	0.514	0.674318		
Push a baby carriage	32	0.511	0.119	0.949	0.161218		Location similar to that of other item
Push something with one foot	33	−0.231	0.111	1.13	0.007791	Yes	Location similar to that of other item
Run	34	−1.479	0.15	−0.911	0.143384	Yes	
Cycle	35	−0.292	0.105	1.794	0.388337		Location similar to that of other item
Carry out sporting activities	36	−2.338	0.146	0.48	0.486996		
Lie down	37	0.855	0.128	1.304	0.643377		
Turn in bed	38	−0.047	0.116	0.738	0.862746		

(Continued)

Table 1.
Continued

Item Description	Item No.	Item Difficulty (Logits)	Standard Error for Location	Fit Residual	Probability	Disordered Thresholds	Reason for Exclusion
Have a normal sex life	39	0.238	0.116	0.45	0.884547		
Participate in social life	40	0.613	0.136	−0.194	0.169369		Location similar to that of other item
Body function component							
Morning pain	41	0.734	0.121	−0.344	0.243286		
Evening pain	42	−0.632	0.131	−0.035	0.874025		
Leg(s) giving way	43	1.324	0.121	0.054	0.512072		
Stress	44	0.107	0.111	3.157	0.001288		Redundant in first analysis
Moving more slowly	45	−1.592	0.126	1.74	0.637549		
Sleep disturbances	46	0.059	0.107	−1.611	0.029646		

[a] The Norwegian version of the Pelvic Girdle Questionnaire was tested. Item 29 was redundant (with a fit residual of >2.5), and 13 items had disordered thresholds; these results suggested that for many of the questions, the women were not able to distinguish among 5 response levels of disability. Among the 6 initial body function items, item 44 was redundant, but none of the items had disordered thresholds. During further Rasch analyses, items with similar locations (reflecting degree of difficulty) were excluded to achieve an instrument with fewer items.

Table 2.
Participant Characteristics[a]

Characteristic	First Survey (n=94)	Second Survey (n=87)
Age, y		
<30	14 (15)	10 (11)
30–37	50 (53)	52 (60)
≥38	30 (32)	25 (29)
Education of >12 y	82 (87)	76 (88)
Work status, employed	81 (86)	64 (73)
Pregnant	52 (55)	43 (49)
No. of children		
0	14 (15)	7 (8)
1	39 (42)	26 (30)
2	21 (22)	29 (33)
≥3	9 (10)	11 (13)
Did not respond to the item	11 (11)	14 (16)
Pain-free periods		
Never	12 (13)	15 (17)
Sometimes	59 (63)	54 (62)
Often	23 (24)	17 (20)
Did not respond to the item		1
Pain duration, mo, $\bar{X}$ (SD)	26.8 (52.6)	27.1 (6.1)

[a] Data are reported as number (percentage) unless otherwise indicated.

tinguish among 5 response levels of disability (Tab. 1).

On the basis of further Rasch analyses, items with similar locations (reflecting degree of difficulty), redundant items, or both were excluded to achieve an instrument with fewer items (Tab. 1). The initial 40-item scale was thus subsequently reduced to 20 items. The 20-item solution showed an excellent fit to the Rasch model (chi-square value of 37.8, with a Bonferroni-adjusted level of significance of $P=.57$); the item mean was 0.00 (SD=1.67), with a fit residual of 0.32 (SD=0.81), and the person's mean location was 0.17 (SD=0.88), with a fit residual of −0.02 (SD=0.92). The person separation reliability index was .89. However, 5 of the 20 items (items 5, 9, 24, 25, and 34) had disordered thresholds, and the category probability curves showed that there were difficulties in discriminating among 5 response levels. Hence, a 4-level response category was adopted for the 20-item version.

The unrestricted model was also applied to the 6 symptoms items. The first Rasch analysis of these 6

Table 3.
Fit Statistics for the 20 Activity and 5 Symptom Items Shown in Order by Location (n=87)[a]

Item Description	No. of Items	Item Difficulty (Logits)	Standard Error for Location	Fit Residual	Probability
Activity subscale					
Run	15	−2.87	0.211	−0.706	0.711874
Carry heavy objects	12	−2.722	0.2	−0.503	0.637292
Do housework	10	−2.204	0.197	0.616	0.271427
Walk >60 min	8	−1.537	0.179	−1.087	0.046608
Carry out sporting activities	16	−1.507	0.168	0.024	0.153782
Stand >60 min	3	−1.327	0.182	0.273	0.209659
Sit >60 min	6	−0.607	0.149	1.748	0.025986
Push something with one foot	20	−0.59	0.149	1.278	0.903588
Bend down	4	−0.09	0.154	0.569	0.463872
Roll over in bed	18	0.227	0.135	−0.497	0.200295
Climb stairs	9	0.295	0.142	−1.41	0.184629
Have a normal sex life	19	0.303	0.14	−1.395	0.159314
Push a shopping cart	14	0.453	0.147	−0.519	0.881982
Get up/sit down	13	0.58	0.148	0.092	0.602385
Lie down	17	0.753	0.14	2.273	0.016312
Carry light objects	11	1.288	0.16	0.367	0.972182
Stand <10 min	2	1.468	0.165	0.735	0.766967
Dress yourself	1	2.077	0.164	−0.443	0.453703
Walk <10 min	7	2.892	0.172	−1.243	0.675974
Sit <10 min	5	3.122	0.191	−0.055	0.776339
Symptom subscale					
Evening pain	2	−2.323	0.182	0.65	0.527535
Doing things more slowly	4	−1.542	0.191	−0.291	0.474492
Interrupted sleep	5	0.271	0.143	−2.019	0.039457
Morning pain	1	0.53	0.151	0.571	0.588506
Leg(s) giving way	3	3.064	0.163	1.709	0.006886

[a] The Norwegian version of the Pelvic Girdle Questionnaire was tested. Location was the logit measure of difficulty (items near the top of the list were the most difficult).

items revealed a borderline fit to the Rasch model (chi-square value of 25.7, with a Bonferroni-adjusted level of significance of $P=.01$), and the person separation reliability index was acceptable (.80). However, 6 participants had extreme scores, and 1 item ("stress") did not fit the model (fit residual=3.16). After exclusion of the misfitting respondents and item, the remaining 5 symptom items showed a good fit to the Rasch model (chi-square value of 8.0; $P=.63$), indicating a unidimensional construct of the symptom subscale. The person separation reliability index was .73. Furthermore, the item mean was 0.00 (SD=1.28), with a fit residual of 0.41 (SD=1.38), and the persons' mean location was 0.42 (SD=1.11), with a fit residual of −0.22 (SD=0.94). None of the 5 items had disordered thresholds.

On the basis of these results, 20 and 5 items were retained in the activity and symptom subscales, respectively. The 2 PGQ subscales were harmonized through the use of a 4-point response scale in the second survey. The items in the Patient-Specific Functional Scale selected by more than 10% of the participants were "walk" (61%), "housework" (53%), "sit" (23%), "carry" (22%), "bend down" (19%), "exercise" (19%), "stand" (17%), "run" (15%), "push object" (12%), "lift" (11%), and "walk stairs" (11%). All of these items were included in the 20-item activity subscale of the PGQ.

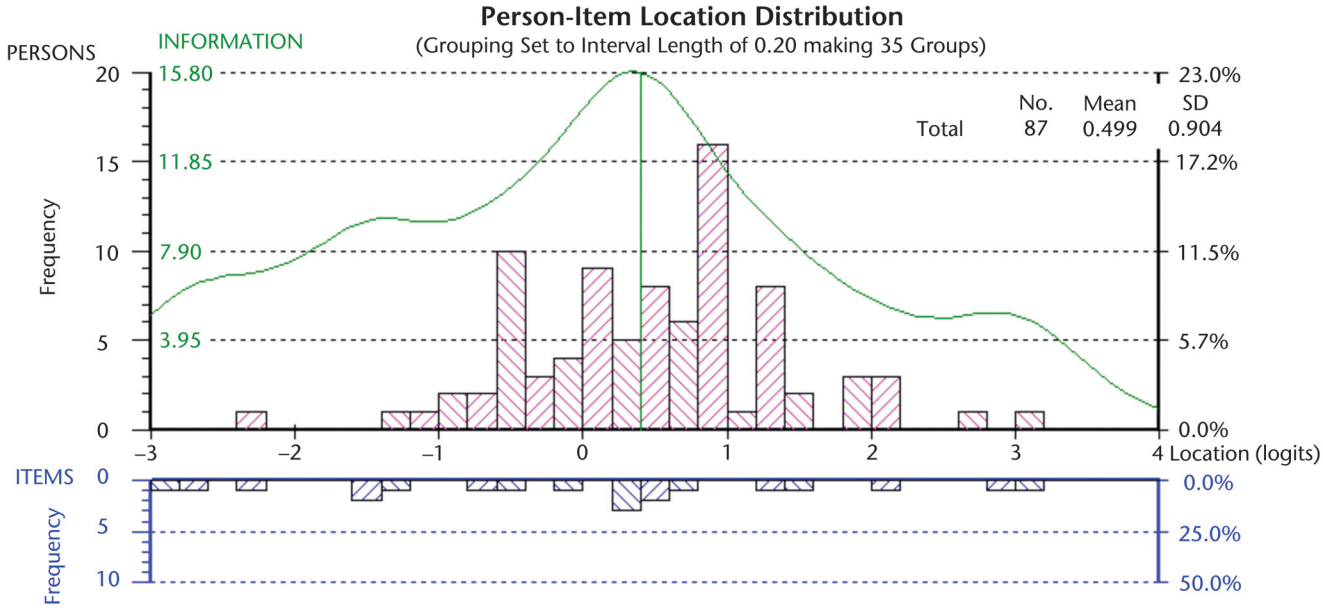

Figure 2.
Person-item threshold location distribution for the 20 activity items.

Statistical Analysis: Second Survey

The main results of the second survey are shown in Table 3. The 20 activity items showed a good fit to the Rasch model, with a chi-square value of 49.5 and a *P* value of .145 (Bonferroni adjusted: .05/20=.0025). All items had satisfactory fit residuals, in the range of −1.41 to 2.27, and the item difficulty levels ranged from −2.87 (most difficult items) to 3.12 (least difficult items). The person separation reliability index was high (.86).

The mean logit for the 87 participants was −0.19 (SD=1.24), indicating that the participants had an ability level slightly higher than the difficulty level of the scale. Despite the fact that 5 participants had extreme scores outside the limits of ±2.5, the calibrated participant ability level and item difficulty maps of the 20-item activity subscale revealed a good match between the difficulties of the items and the participants' levels of function (Fig. 2). The bars in the upper part of the graph in Figure 2 represent groups of participants and their ability levels, and the bars in the lower part of the graph represent the item locations and their distribution. Both participant ability and item difficulty levels are shown on the same linear scale (logit scale). A few items are located in the same place, in terms of difficulty, and this situation is represented as 1 block on top of another. Figure 2 shows that none of the participants had scores outside the range of measurement assessed by the scale. Therefore, there was no ceiling or floor effect, as shown by the distribution in Figure 2. The item "carry out sporting activities" had disordered thresholds. The threshold ordering for the 20 activity items was adequate after we collapsed the 2 highest scores on the "carry out sporting activities" item.

There was no uniform DIF for any of the subgroups, except for pregnancy and the items "dressing" and "push an object." When we adjusted for the uniform DIF by splitting the 2 items for pregnant and postpartum participants, the fit statistics were still excellent. There was a nonuniform DIF for pregnancy and the items "dressing," "walk less than 10 minutes," and "push with one foot," and there was a nonuniform DIF for level of education and the items "walk less than 10 minutes" and "carry out sporting activities."

The 2 subsets of items with positive and negative loadings on the first residual component were separately fitted to the Rasch model, and the person estimates were obtained. The difference between these 2 subsets was not statistically significant, supporting the unidimensionality of the final 20-item activity subscale.

The 5 symptom items showed a reasonably good fit to the Rasch model (chi-square value of 21.8; *P*=.03), with fit residuals in the range of −2.02 to 1.71 and item difficulty levels in the range of −2.32 to 0.53 (Tab. 3). The mean logit for the 87 participants was 0.66 (SD=1.19), with a fit residual of −0.21 (SD=0.91). The person separation

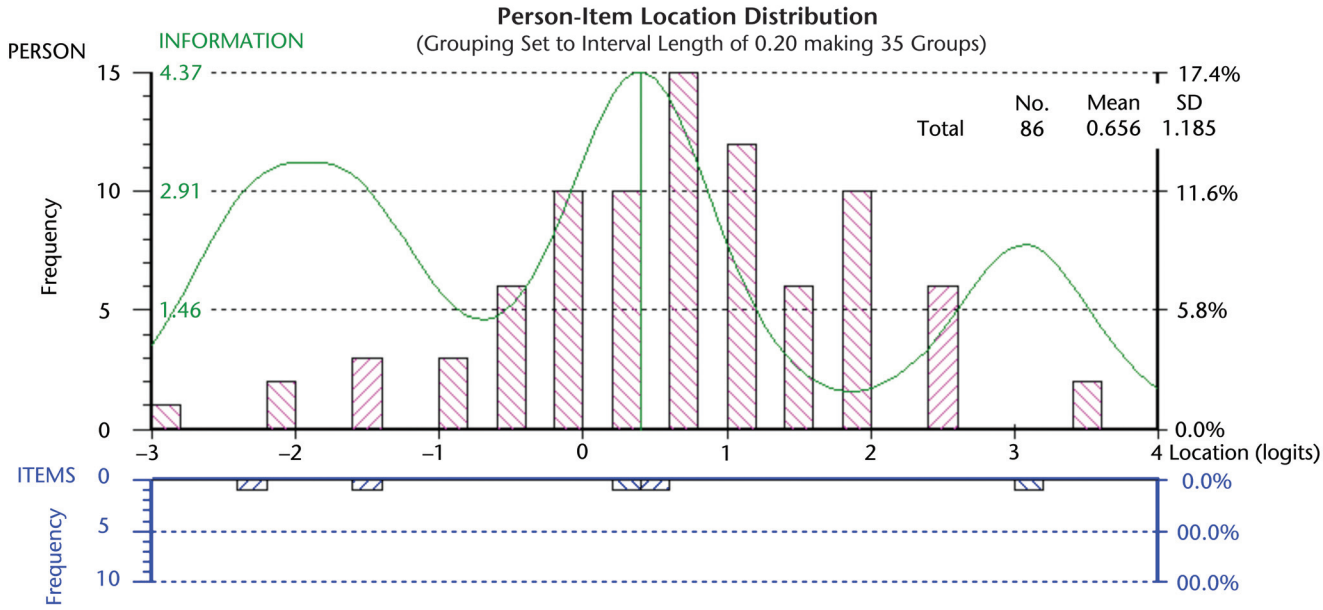

Figure 3.
Person-item threshold location distribution for the 5 symptom items.

reliability index for the 5 items was .64. None of the participants had extreme scores (with fit residuals beyond ±2.56). The calibrated participant ability and item difficulty maps of the 5-item symptom subscale showed no floor or ceiling effect (Fig. 3), and none of the 5 items had disordered thresholds. Furthermore, there was no uniform DIF for age group, pregnancy, work status, or level of education. There was 1 nonuniform DIF for pregnancy and the item "sense of leg(s) giving way." Testing of the dimensionality of the final 5-item symptom subscale confirmed that it was unidimensional.

Test-Retest Reliability

Test-retest reliability was based on data from 42 participants (21 pregnant and 21 not pregnant) who completed the PGQ within 1 week in the second survey. The mean differences between the 2 measurements were 1.17 (SD=7.37) for the 20-item activity subscale and 0.48 (SD=10.14) for the 5-item symptom subscale. The intraclass correlation coefficients for rescoring were .93 (95% confidence interval=0.86–0.96) for the 20-item activity subscale and .91 (95% confidence interval=0.84–0.95) for the 5-item symptom subscale. The standard errors of measurement between the test and retest scores were 5.21 for the 20-item activity subscale and 7.17 for the 5-item symptom subscale. Therefore, the repeatability values, including the 95% confidence intervals, were 14.43 for the 20-item activity subscale and 19.86 for the 5-item symptom subscale.

Discussion

The PGQ is a condition-specific, patient-reported outcome measure designed to assess aspects of quality of life for the evaluation of treatment outcomes that are important to patients with PGP. Evidence for the appropriateness of the PGQ for patients with PGP, in terms of acceptability, data quality, reliability, and validity, has been obtained. Moreover, the questionnaire is simple to administer, takes just 3 minutes to complete, and is feasible for use in clinical practice.

Treatment outcomes in clinical physical therapy research and evidence-based practice must be evaluated by measures that are reliable, valid, and responsive to changes in health. To our knowledge, the PGQ is the first condition-specific measure for PGP, and the present study provided evidence for its appropriateness. A major strength of the present study is that data were collected from people with clinically confirmed PGP. Studies have found differences in the perceptions of clinicians and patients regarding impact of disease, treatment outcome priorities, and values of various outcomes.[15,16] Traditional clinical measures are commonly based on clinical opinions and, therefore, lack content validity from the perspective of patients. The perspective of patients regarding how their conditions or diseases affect their lives must be considered in the development of measures that are designed to assess health status and outcomes of importance to recipients of health care. The content of the PGQ was based on information from a focus group and input from

experienced physical therapists. Moreover, the focus group comprised a representative sample of people with PGP with regard to age and duration of PGP. The severity of symptoms can influence the relevance of various outcomes[17]; thus, the participants had a wide range of symptom severity. The collaboration between physical therapists providing health care to women and researchers with experience in patient-reported outcome measures was a further strength of the present study.[18,19] In addition, the response rate was high, as in a previous study of PGP.[2]

Previous studies of PGP used questionnaires that were developed for patients with LBP and therefore may not be adequate for patients with PGP.[2,20,21] Pelvic girdle pain is considered to be a condition different from LBP[1]; thus, functional disabilities in patients with PGP may differ from those in patients with LBP. Items similar to those in the PGQ are found across LBP questionnaires, but all of the PGQ items are not found in any single one of these questionnaires. Hence, the PGQ has greater content validity from the perspective of patients with PGP. The PGQ will be compared with LBP questionnaires as part of a concurrent evaluation of measurement properties, including validity and responsiveness to change.

During the pilot stage of questionnaire development, we identified 3 functional activities that are not included in measures commonly used for LBP: "push a shopping cart," "legs/leg(s) giving way," and "push something with one foot." The "pulling a mat" test, in which a patient performs standing hip abduction and adduction, was found to be highly sensitive for anterior PGP in a recent study.[22] The Patient-Specific Functional Scale showed that 61% of participants (73% pregnant and 48% postpartum) reported problems with walking. Similarly, 81% of postpartum patients with PGP reported pain associated with walking in another study.[23]

Physical therapists commonly treat patients with PGP during pregnancy and in the postpartum period. Therefore, there is a need for an outcome measure appropriate for both stages.[6] Our analysis showed that participants reported difficulties with the same functional activities both during pregnancy and postpartum. The grading of difficulties, however, differed because participants who were pregnant presented with more severe problems. Thus, the PGQ is an appropriate outcome measure for patients with PGP during pregnancy and in the postpartum period. Our analysis showed that the PGQ comprises both activity and symptom subscales and includes items relating to functional disability, physical symptoms [such as "legs(s) giving way" and "do things more slowly"], pain, and sleep. Both morning pain and evening pain are included because of evidence that patients with PGP have markedly worse evening pain.[18,20] It is an advantage for a single measure to cover both function and pain in PGP for use in both research and clinical applications.

Compared with generic instruments, such as the Medical Outcomes Study 36-Item Health Survey Questionnaire, condition-specific instruments, such as the PGQ, have greater relevance to patients and clinicians and are usually more responsive to changes in health and quality of life after care for the health problem of interest.[24] However, the 2 approaches are complementary, and the PGQ is brief enough for use alongside generic instruments, which have a greater potential to measure side effects or unforeseen effects of care and are suitable for comparisons of outcomes for other patient populations.

The PGQ was developed and tested in Norwegian and has undergone a process of forward-backward translation in accordance with recommendations for questionnaire translation (Appendixes 1 and 2).[25] The questionnaire is recommended for application in different cultural settings after forward-backward translation when necessary and assessment of cross-cultural equivalence.

The PGQ is designed to evaluate treatment outcomes for patients with complaints associated with PGP. The questionnaire includes items relating to 2 scales (activity/participation and body functions [symptoms]) and is acceptable for use both during pregnancy and postpartum in patients with PGP. Evidence for the data quality, internal consistency, fit to the Rasch model and, internal validity of the PGQ has been obtained. Future research is needed to assess the responsiveness to change of the PGQ.

Dr Stuge and Dr Grotle provided concept/idea/research design. All authors provided writing and data analysis. Dr Stuge, Ms Krogstad Jenssen, and Dr Grotle provided data collection.

The Regional Committee for Medical Research Ethics in Norway approved the study protocol.

The authors thank The Norwegian Fund for postgraduate education in physiotherapy for grant support.

This article was submitted November 1, 2010, and was accepted February 24, 2011.

DOI: 10.2522/ptj.20100357

References

1 Vleeming A, Albert HB, Ostgaard HC, et al. European guidelines for the diagnosis and treatment of pelvic girdle pain. *Eur Spine J.* 2008;17:794–819.

2 Stuge B, Lærum E, Kirkesola G, Vøllestad N. The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized controlled trial. *Spine*. 2004; 29:351–359.

3 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;12:86–97.

4 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*. 2006;31:E149–E155.

5 Robinson HS, Mengshoel AM, Bjelland EK, Vollestad NK. Pelvic girdle pain, clinical tests and disability in late pregnancy. *Man Ther*. 2010;15:280–285.

6 Boissonnault JS. A review of self-report functional outcome measures in selected obstetric physical therapy interventions. *Journal of Women's Health Physical Therapy*. 2009;33:7–12.

7 van de Pol G, de Leeuw JR, van Brummen HJ, et al. The Pregnancy Mobility Index: a mobility scale during and after pregnancy. *Acta Obstet Gynecol Scand*. 2006;85:786–791.

8 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.

9 Roland M, Morris R. A study of the natural history of back pain, part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine*. 1983;8: 141–144.

10 Fairbank JCT, Davies JB, Couper J, O'Brien JP. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy*. 1980;66: 271–273.

11 Salèn BA, Spangfort EV, Nygren ÅL, Nordemar R. The Disability Rating Index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol*. 1994; 47:1423–1434.

12 *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: World Health Organization; 2001.

13 Stratford PW, Gill C, Westaway MD, Binkley JM. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiother Can*. 1995; 47:258–263.

14 Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60:34–42.

15 Hewlett S, Smith AP, Kirwan JR. Values for function in rheumatoid arthritis: patients, professionals, and public. *Ann Rheum Dis*. 2001;60:928–933.

16 Kvien TK, Heiberg T. Patient perspective in outcome assessments: perceptions or something more? *J Rheumatol*. 2003;30: 873–876.

17 Casarett D, Karlawish J, Sankar P, et al. Designing pain research from the patient's perspective: what trial end points are important to patients with chronic pain? *Pain Med*. 2001;2:309–316.

18 Stuge B, Veierød MB, Lærum E, Vøllestad N. The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a two-year follow-up of a randomized clinical trial. *Spine*. 2004;29:E197–E203.

19 Paulsen T, Grotle M, Garratt A, Kjeken I. Development and psychometric testing of the patient-reported measure of activity performance of the hand (MAP-Hand) in rheumatoid arthritis. *J Rehabil Med*. 2010; 42:636–644.

20 Nilsson-Wikmar L, Holm K, Oijerstedt R, Harms-Ringdahl K. Effect of three different physical therapy treatments on pain and activity in pregnant women with pelvic girdle pain: a randomized clinical trial with 3, 6, and 12 months follow-up postpartum. *Spine*. 2005;30:850–856.

21 Bastiaenen CH, de Bie RA, Wolters PM, et al. Effectiveness of a tailor-made intervention for pregnancy-related pelvic girdle and/or low back pain after delivery: short-term results of a randomized clinical trial [ISRCTN08477490]. *BMC Musculoskelet Disord*. 2006;7:1–13.

22 Fagevik Olsén M, Gutke A, Elden H, et al. Self-administered tests as a screening procedure for pregnancy-related pelvic girdle pain. *Eur Spine J*. 2009;18:1121–1129.

23 Mens JM, Vleeming A, Stoeckart R, et al. Understanding peripartum pelvic pain: implications of a patient survey. *Spine*. 1996;21:1363–1369.

24 Garratt A. Patient reported outcome measures in trials. *BMJ*. 2009;338:a2597.

25 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality-of-life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46:1417–1432.

Appendix 1.

Pelvic Girdle Questionnaire (English Version)[a]

To what extent do you find it problematic to carry out the activities listed below because of pelvic girdle pain? For each activity tick the box that best describes how you are today.

How problematic is it for you because of your pelvic girdle pain to:	Not at all (0)	To a small extent (1)	To some extent (2)	To a large extent (3)
1. Dress yourself				
2. Stand for less than 10 minutes				
3. Stand for more than 60 minutes				
4. Bend down				
5. Sit for less than 10 minutes				
6. Sit for more than 60 minutes				
7. Walk for less than 10 minutes				
8. Walk for more than 60 minutes				
9. Climb stairs				
10. Do housework				
11. Carry light objects				
12. Carry heavy objects				
13. Get up/sit down				
14. Push a shopping cart				
15. Run				
16. Carry out sporting activities*				
17. Lie down				
18. Roll over in bed				
19. Have a normal sex life*				
20. Push something with one foot				

* If not applicable, mark box to the right.

How much pain do you experience:	None (0)	Some (1)	Moderate (2)	Considerable (3)
21. In the morning				
22. In the evening				

To what extent because of pelvic girdle pain:	Not at all (0)	To a small extent (1)	To some extent (2)	To a large extent (3)
23. Has your leg/have your legs given way?				
24. Do you do things more slowly?				
25. Is your sleep interrupted?				

[a] Scoring procedure: the scores were summarized and recalculated to percentage scores from 0 (no problem at all) to 100 (to a large extent).

Appendix 2.

The Pelvic Girdle Questionnaire (Norwegian Version)[a]

I hvilken grad finner du det problematisk på grunn av plager fra bekkenet å utføre aktivitetene som er listet opp nedenfor. Sett et kryss for hver aktivitet som best beskriver hvordan du har det nå for tiden.

Hvor problematisk er det på grunn av bekkenet å:	Ikke i det hele tatt (0)	I liten grad (1)	I noen grad (2)	I stor grad (3)
1. Kle på deg selv				
2. Stå mindre enn 10 minutter				
3. Stå mer enn 60 minutter				
4. Bøye deg				
5. Sitte mindre enn 10 minutter				
6. Sitte mer enn 60 minutter				
7. Gå mindre enn 10 minutter				
8. Gå mer enn 60 minutter				
9. Gå trapper				
10. Husarbeid				
11. Bære lett				
12. Løfte tungt				
13. Reise/sette seg				
14. Skyve en handlevogn				
15. Løpe				
16. Utføre sportslige aktiviteter*				
17. Ligge				
18. Snu deg i sengen				
19. Ha et normalt seksualliv*				
20. Skyve noe med den ene foten				

* Hvis ikke aktuelt, kryss av i boksen til høyre.

Hvor sterke smerter har du:	Ingen (0)	Noe (1)	Moderate (2)	Svært mye (3)
21. Om morgenen				
22. Om kvelden				

I hvilken grad på grunn av plagene i bekkenet:	Ikke i det hele tatt (0)	I liten grad (1)	I noen grad (2)	I stor grad (3)
23. Svikter benet/bena under deg?				
24. Gjør du ting langsommere?				
25. Forstyrres nattesøvnen din?				

[a] Skårings prosedyre: Skårene summeres og kalkuleres til prosent fra 0 (ikke i det hele tatt) til 100 (i stor grad).