European Journal of Physiotherapy, 2015; Early Online: 1–10

informa
healthcare

ORIGINAL ARTICLE

Reliability and validity of the Body Awareness Rating Scale (BARS), an observational assessment tool of movement quality

LIV HELVIK SKJAERVEN[1], GUNVOR GARD[2], MARY ANNE SUNDAL[3], & LIV INGER STRAND[4]

[1]*Department of Occupational Therapy, Physiotherapy and Radiography, Faculty of Health and Social Sciences, Bergen University College, Bergen, Norway,* [2]*Department of Health Sciences, Lund University and Department of Health Sciences, Luleå University of Technology, Sweden,* [3]*Department of Occupational Therapy, Physiotherapy and Radiography, Faculty of Health and Social Sciences, Bergen University College, Bergen, Norway,* [4]*Department of Global Public Health and Primary Care, Physiotherapy Research Group, University of Bergen, Norway*

Abstract

Movement quality assessed by the Body Awareness Rating Scale (BARS) is used as an indicator of health and self-efficacy in patients with long-lasting musculoskeletal and mental health problems. The objective of the study was to examine reliability and construct validity of the movement quality scale. 25 patients and 25 healthy persons were included. Internal consistency was examined by Cronbach's α, reliability by intraclass correlation coefficient ($ICC_{agreement}$) and measurement error reported by standard error of measurement (SEM) and smallest detectable change (SDC). Construct validity was examined by testing hypotheses of moderate association between the observational scale and the self-report Short-Form Health Survey (SF-36) subscales and the General Perceived Self-Efficacy Scale (GPSES). A hypothesis about the difference in scores between groups being expected to differ in health states was tested. Internal consistency (α) was 0.92. Inter-tester reliability was ICC = 0.99 and SEM = 0.8. The test–retest reliability was ICC = 0.96 and SEM = 1.4, implying that improvement should be above 3.3 (SDC) to claim a treatment effect. BARS was moderately correlated ($0.30 \leq r_s < 0.60$) with most SF-36 subscales and GPSES. The patients demonstrated less movement quality than healthy persons. Evidence was provided of high internal consistency and reliability in qualified testers. Construct validity was indicated, as BARS reflected various aspects of health and self-efficacy.

Key words: *Assessment tool, Basic Body Awareness Therapy, measurement properties, movement analyses, movement awareness*

Introduction

Long-lasting musculoskeletal health problems are leading causes of disability in Western countries (1). Life experiences and individual coping are expressed in human movement and function. There is evidence that mental state and stressors in daily life influence several physiological processes in the body with impact on sensory-motor co-ordination and movement awareness (2). Decreased awareness is reflected in dysfunctional movements and compensatory movement strategies (3,4). Awareness can be defined as an attentive, relaxed and alert presence (5), and movements that are performed without awareness may have a mechanical appearance (6). Being aware is a relative phenomenon and means continually monitoring internal sensations and external environments (5). Thus, it is hypothesized that movement awareness to some degree is associated with perceived wellbeing, general health and self-efficacy (3,7,8).

The Body Awareness Rating Scale (BARS) was developed in the late 1980s (4,9,10), and is a physiotherapeutic assessment tool for patients suffering from long-lasting musculoskeletal disorders and mental health problems (4). It has roots in Basic Body Awareness Therapy (BBAT), a physiotherapy movement awareness modality, inspired by the French movement educator and psychotherapist Dropsy (3,11,12), brought into physiotherapy by Roxendal (13) and further developed (14,15).

Correspondence: Liv Helvik Skjaerven, Department of Occupational Therapy, Physiotherapy and Radiography, Faculty of Health and Social Sciences, Bergen University College, Inndalsveien 28, NO-5063 Bergen, Norway. E-mail: lhs@hib.no; liv.skjaerven@gmail.com

(Received 30 June 2014; accepted 22 November 2014)

ISSN 2167-9169 print/ISSN 2167-9177 online
DOI: 10.3109/21679169.2014.992470

BARS consists of two assessment parts, as supported by factor analysis (4): (i) observation and assessment of movement quality based on 12 movements (Figure 1) and (ii) interview with the patient about movement awareness immediately after exploring each movement (not part of present study). BARS, as an observational assessment tool, was developed to examine quality in general movement co-ordinations and movement habits, observing compensations and healthy movement resources. The patients' general movement quality is evaluated and scored according to the way the movements are performed, relating to space, time and energy (16–19). Movement quality is an umbrella term embracing physical, physiological, psycho-socio-cultural and existential perspectives on human movement, as supported by phenomenological research (16–19).

When observing movement quality using BARS, the physiotherapist directs attention to the whole moving person, more than separate parts of the body (4,17). Focusing the whole is an opportunity to observe how the dynamic interplay between breathing, postural balance and mental awareness influence human movements. The three are core elements of functional, free and economic movement co-ordinations (18). In addition, rhythm, flow and intention, expressed in the movements, are evaluated. In the examination situation, the therapist implements a standardized pedagogy to guide the patient in performing the movements (19) (Appendix 2 to be found online at http://informahealthcare.com/doi/abs/10.3109/21679169.2014.992470).

One of the primary reasons for using an assessment tool in clinical practice is to determine the therapeutic intervention as well as the effect of therapy. Repeated assessments are susceptible to several sources of error, including inconsistencies by the examiner, the instrument, the procedure applied and the patient's condition. Pilot studies of inter-tester reliability and validity were performed during early BARS development, indicating sufficient reliability for clinical use (4,9,10). A larger study was, however, needed to provide sufficient evidence of reliability and validity. The aim of the present study was therefore to examine internal consistency, inter-tester, test–retest reliability and construct validity of the BARS movement quality scale in patients with long-lasting musculoskeletal disorders and mental health problems, following recent international guidelines for examination of measurement properties.

Methods

Design

A cross-sectional design was used to examine internal consistency, inter-tester reliability and construct

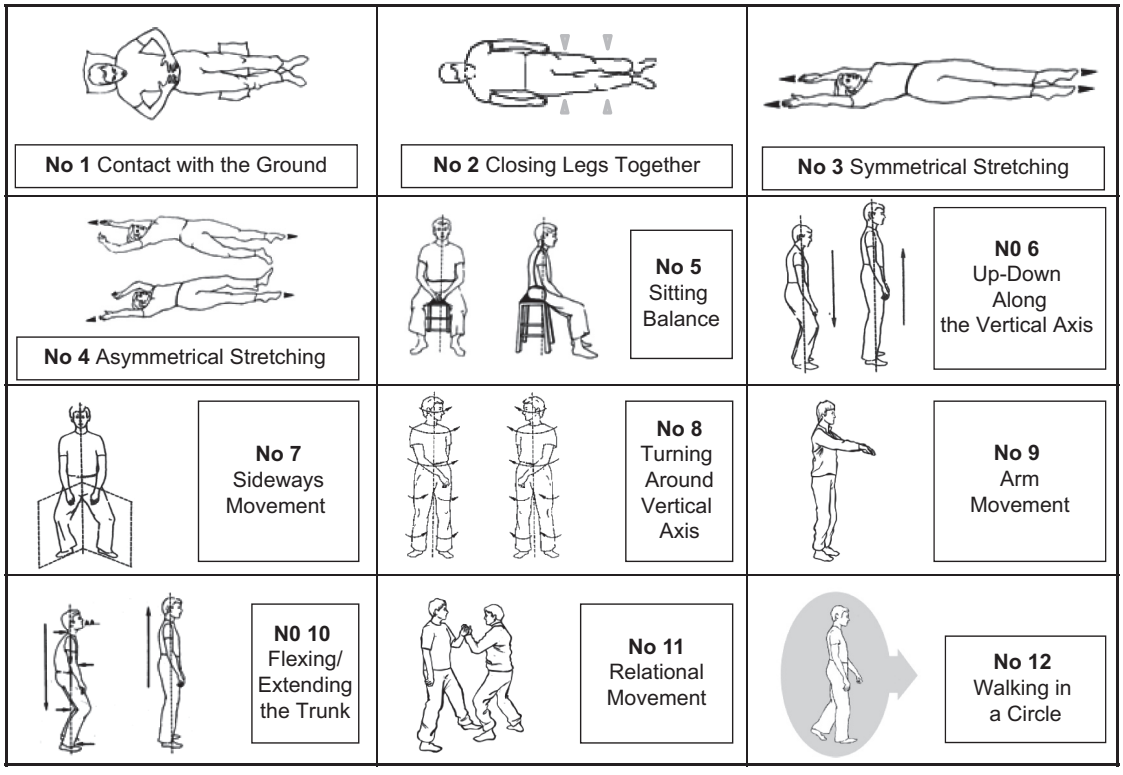

Figure 1. Movements in the Body Awareness Rating Scale [Nos 1–8, 10 and 12 (11); Nos 9 and 11: Centre for New Media, Bergen University College, 2013].

validity, and a longitudinal design to examine test–retest reliability.

Participants

A total of 50 persons above 18 years of age (25 patients and 25 healthy individuals) participated in the study to evaluate internal consistency, construct validity and discriminative ability of BARS. In the study of inter-tester and test–retest reliability, 30 of the 50 persons participated, 15 patients and 15 healthy individuals. These 30 were the first assessed among the 50 participants. The researcher invited them to come for a second assessment. All accepted the invitation.

The patients referred to physiotherapy from the medical doctor were recruited consecutively from the waiting list for physiotherapy of a community healthcare center over a period of 6 months. Inclusion criteria were long lasting (more than 3 months) musculoskeletal disorders and diagnose of depression and/or anxiety, combined with problems related to the musculoskeletal system. The informants were ambulatory without need of assistive device, and able to complete the test, implying movements for 40 min, lying, sitting, standing and walking. Exclusion criteria were acute problems related to the musculoskeletal system, serious psychiatric diagnosis not referred to physiotherapy, a recent surgical history, malignant basic diseases and acute traumas and infections. They should not be previously treated with BBAT. The recruitment followed the principle "first patient on the list". All patients agreed to participate and were tested; no one was excluded.

The group of healthy persons was a convenience sample recruited from a local organization close to the healthcare center. They were randomly selected from a pool of volunteers. They reported themselves to be well functioning and healthy, presenting no special disorders and not in need of physiotherapy. The inclusion followed the principle "first participants on the list". All agreed to participate and were tested; no one was excluded. Matching according to age and gender was not performed.

The study was approved by the Regional Committee for Medical Research Ethics in Western Norway and the Norwegian Data Protection Authority (181.06). All participants signed a written informed consent form. The Norwegian Fund for Postgraduate Education in Physiotherapy provided grants to support the study.

Assessment tools used in the study

Body Awareness Rating Scale (BARS). The physiotherapist guided each patient in the BARS movements (Appendices 1 and 2 to be found online at http://informahealthcare.com/doi/abs/10.3109/21679169.2014.992470). The physiotherapists assessed movement quality according to criteria described in the BARS-Movement Quality Scores (Table I). The patients performed five repetitions of each movement (3–4 min for each), and the most healthy, functional movement was scored. The BARS items are scored from 1 to 7. The scale includes half (0.5) scores to make the scale more sensitive to differences between individuals and sensitive to nuances of change, both over time and within each therapy session.

A score of 7 is defined as the most healthy, functional movement quality, described as balanced, free, centered, unified, rhythmic, and synchronous. A score of 1 is defined as the most pathological, dysfunctional movement quality, described as unstable, mechanical, stiff, un-rhythmical and with a lack of unity (18). The sum score of all items ranges from 12 to 84.

Short-Form Health Survey (SF-36). The SF-36 contains 36 items subdivided into eight dimensions of subjective health: Physical functioning (PF), Role Function – Physical aspects (RP), Bodily Pain (BP), General Health (GH), Vitality (V), Social Functioning (SF), Role Emotional (RE) and Mental Health (MH) (20–22). The score for each subscale ranges from 0 to 100, high scores indicating good health. The reliability has been reported to be satisfactory (23,24).

General Perceived Self-Efficacy Scale (GPSES). The GPSES is a 10-item questionnaire, used to capture the individual's general sense of self-efficacy, referring to people's optimistic beliefs of ability to cope with a variety of difficult demands in life (25,26). It consists of 10 items to be assessed on a 4-point rating scale ranging from "not at all true" scored as 1 to "exactly true" scored as 4. The total score ranges from 10 to 40 with high scores indicating high self-efficacy. The reliability has been reported to be satisfactory (27).

Testers and test procedure

Two physiotherapy specialists in mental health were recruited to examine reliability. Tester A and B, both women, were qualified in BARS. The testers knew each other professionally, but had separate working places. To calibrate their testing procedure before the start of the study, they tested one patient and one healthy person together, but scored BARS separately, and discussed the scores after each assessment. The testers had no relationship to the participants.

Table I. Body Awareness Rating Scale (BARS) – Movement quality scores.

MQ Score 7	*Very good functional movement quality*: The vertical axis is very well balanced, stable, firm and free. Movement characteristics: Very good functional form, flow, elasticity and rhythm; a very good intentional clarity and direction in the movements. The amount of energy expressed in the movement is very appropriate to the task. The movements originate very clearly from the centre in the trunk. The movements in the person as a whole are simultaneous; they are congruent and in accordance with each other, and are characterized by very good unity and integration. They express a very good movement harmony.
MQ Score 6	*Good functional movement quality*: The vertical axis is well balanced, stable, firm and free. Movement characteristics: Good functional form, flow, elasticity and rhythm; a good intentional clarity and direction in the movements. The amount of energy expressed in the movement is appropriate to the task. The movements originate clearly from the centre in the trunk. The movements in the person as a whole are characterized by good unity and integration. They express a good movement harmony.
MQ Score 5	*Moderate functional movement quality*: The vertical axis is moderately well balanced, stable, firm and free. Movement characteristics: Moderate functional form, flow, elasticity and rhythm; a moderate clarity in the intention and direction of the movements. The amount of energy expressed is moderately appropriately to the task. There are moderate signs of movement originating from the center in the trunk. The movements are characterized by a moderate and variable amount of unity and integration. The movements in the person as a whole are characterized by moderate unity and integration. They express moderate movement harmony.
MQ Score 4	*Some functional movement quality*: The vertical axis has some balance, stability, firmness and freedom. Movement characteristics: Some glimpses of functional form, flow, elasticity and rhythm; some glimpses of intention and direction of the movements. The amount of energy expressed in the movement is somewhat appropriate to the task. There are some signs of movement originating from the centre in the trunk. The movements in the person as a whole are characterized by some glimpses of unity and integration. They express some movement harmony.
MQ Score 3	*Weak functional movement quality*: The vertical axis has an uncertain balance, little stability, firmness and freedom. Movement characteristics: somewhat dysfunctional in form, somewhat mechanical, staccato, stiff, a-rhythmical and lifeless. The movements are characterized by some weakness in the intention and direction. The amount of energy in the movement is more discordant with the task, being smaller and more closed or larger and more open or having too much or too little energy. The movements originate more from the periphery than from the centre in the trunk. The movements are characterized by a weak unity and integration. They express weak movement harmony.
MQ Score 2	*Mostly dysfunctional movement quality*: The vertical axis is mostly lacking balance, stability, firmness and freedom. Movement characteristics: Mostly dysfunctional form, staccato, mechanical, stiff, a-rhythmical, lifeless, mostly lacking elasticity. The movements are characterized by a mostly lacking intention and direction. The amount of energy in the movements is mostly in discord with the task, either being far too small and closed or far too large and open or using far too much or far too little energy. The movements originate mostly from the periphery. There is mostly a lack of unity between upper and lower body. The movements are mostly lacking unity and integration. They express a lack of movement harmony.
MQ Score 1	*Dysfunctional movement quality*. The vertical axis is unstable and fragmented. Movement characteristics: Dysfunctional form, staccato, mechanical, stiff, a-rhythmical, lifeless, lacking elasticity. The movement is characterized by lacking intention and direction. The movements originate from the periphery and are disconnected to each other. The movements in the whole person are in discord, incongruent and counteract each other. They express movement disharmony.

The test procedure followed the standardized test protocol. The patients filled in demographic information of gender, age, body mass index, education, marital and work status, and the Short-Form Health Survey (20–22) and GPSES (25,26). BARS was assessed after filling in the demographic data.

Examination of reliability

Internal consistency. Internal consistency is defined as the degree of inter-relatedness among the items of a unidimensional scale (29). In this study, this meant inter-relatedness among the 12 BARS items.

Reliability and measurement error. Reliability is defined as "the degree to which measurement is free from measurement error" (29). Two therapists assessed the participants on the same occasion (inter-tester) and one therapist retested the same participants, 3 days later (test–retest). The participants explored each movement several times, guided by tester A, and the most optimal performance was scored. Tester B observed and scored the movement quality, independent from tester A. To avoid remembering previous scores, tester A repeated the test after 3 days. She tested three or four patients on the same day.

Examination of construct validity

Construct validity is defined as the degree to which the scores of a measurement instrument (here BARS sum scores) are consistent with hypotheses regarding relationships with scores of other instruments, or differences between relevant groups based on the assumption that the instrument validly measures the construct to be measured (29). This means that we

need evidence that we capture the phenomenon we intend to measure. This is addressed by examining the relationship between the particular measure, and a gold standard measure for the phenomena, if available, or with other measures that are expected to measure more or less the same phenomenon. We do this by testing hypothesis of degree of associations, as recommended by the consensus of the international COSMIN group (29). We generally expect a high correlation between measures that assess very similar phenomena and low correlations between measures that asses very different phenomena. As we lacked a gold standard for health and self-efficacy assessed by BARS, construct validity was examined by testing the hypotheses of relationships between BARS and two other measures.

Hypotheses 1–8. As movement quality assessed by BARS is seen to reflect physical, physiological, psycho-socio-cultural and existential aspects of a person's experience of health, we hypothesized a moderate correlation between movement quality assessed by BARS and the eight subscales of the Short-Form Health Survey (SF-36).

Hypothesis 9. We also hypothesized a moderate correlation between the BARS total scores and the General Perceived Self-Efficacy Scale (GPSES), which is a measure of ability to cope with a variety of difficult demands in life.

Hypothesis 10. We hypothesized that patients with musculoskeletal disorders displayed lower (more dysfunctional) BARS scores than persons who reported that they were healthy.

As BARS is an observational measure and SF-36 and GPSES are both self-report measures, we expected only a moderate correlation between them. We also hypothesized a statistical significant difference in scores between healthy persons and patients suffering from long-lasting musculoskeletal and mental disorders (Hypothesis 10).

Statistical analysis

Demographic data was examined by descriptive statistics. PASW 18 (SPSS Inc., Chicago, IL, USA) was used for statistical analyses.

Internal consistency was examined using Cronbach's α formula. Ideally, the α value should be between 0.70 and 0.90 (30). While inter-relatedness may be questioned in values lower than 0.70, a value higher than 0.90 may imply redundancy of items. The inter-item correlation should ideally be between 0.20 and 0.50 and not higher than 0.70 as the two items would then assess very much the same aspect. The correlation between each item and the total score (item-total correlation) should be above 0.3 to show discriminate ability. The contribution of each item to the total scale was examined by studying the alpha value if separate items were deleted.

A two-way mixed absolute agreement model of intraclass correlation ($ICC_{agreement}$) was used to examine inter-tester and test–retest reliability. In this model, the error variance consists of the residual variance plus variance due to systematic differences (30). Measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured (29). The standard error of measurement (SEM) is estimated by taking the square root of the within subject variance consisting of variance among the measures plus the residual variance. The smallest detectable change (SDC) was calculated as $[SEM \times \sqrt{2} \times 1.64]$ and is the test value that a patient must exceed to demonstrate an improvement above measurement error with 95% certainty (31).

Spearman's correlation was used to examine the association between BARS versus SF-36 subscales and GPSES (Hypotheses 1–9). Moderate correlation was defined as $0.30 \leq r_s < 0.60$ (32).

Difference in BARS sum scores between healthy people and patients with musculoskeletal and mental disorders was examined by independent samples *t*-test (Hypothesis 10), $p < 0.05$.

Result

Demographic data and test characteristics are described in Table II. More women were included in the patient group (92%) than in the healthy group (68%), while age and body mass index were similar in the two groups. Higher education was more frequent in the healthy group and they were more likely to be working than the patient group.

Internal consistency

Cronbach's α of the BARS items was 0.92. The correlation between each item and the sum score (corrected item-total correlation) varied between 0.57 and 0.75. No item increased the total α value if it was deleted from the scale (Table III).

Inter-tester reliability

High inter-tester reliability of the BARS sum scores was demonstrated between raters A and B (Figure 2). The ICC was 0.99 (95% confidence interval (CI), 0.97–0.99). The measurement error between the testers was very low, SEM = 0.8. The difference between

Table II. Demographic and test scores at baseline, comparing patients and healthy persons.

	Total group, $n=50$			Subgroup, $n=30$		
Variables	Patients, $n=25$	Healthy, $n=25$	Difference, p-value	Patients, $n=15$	Healthy, $n=15$	Difference, p-value
Gender, women: no. (%)	23 (92)	17 (68)	0.034[a]	13 (87)	9 (60)	0.099[a]
Age, years: mean (SD) min–max	42.1 (9.7) 18–63	47.5 (15.9) 20– 73	0.157[b]	42.5 (11.4) 18–63	49.2 (15.7) 21–73	0.190[b]
Body mass index: mean (SD) min–max	26.1 (6.4) 19.4–46.6	24.8 (3.7) 17.6–31.7	0.387[b]	25.1 (4.2) 19.4–35.4	25.3 (4.0) 17.6–31.7	0.860[b]
Education,[c] no. low, medium, high	8, 11, 6	2, 6, 17	0.006[a]	5, 6, 4	2, 4, 9	0.165[a]
Marital status, no. married, other	21, 4	15, 10	0.059[a]	14, 1	9, 6	0.031[a]
Work status, no. working, not working	6, 19	16, 9	0.004[a]	2, 13	9, 6	<0.001[a]
BARS sum score: mean (SD) min–max	46.3 (10.5) 23.5–63.5	55.4 (8.5) 38.0–75.5	0.001[b]	51.1 (6.5) 40.0–63.5	56.2 (7.6) 38.0–68.0	0.058[b]

[a]Chi square test. [b]Independent samples t-test. [c]Education: "low": completing elementary school, "medium": completing high school, "high": completing college/university.

a person's measurement and the true value would be expected to be less than 1.4 (1.96 × SEM) for 95% of observations.

Test–retest reliability

The test–retest reliability of the BARS sum score for rater A was also high, as shown in Figure 3. The ICC was 0.96 (95% CI 0.93–0.98). The measurement error according to the SEM between the two test points was 1.4, and the SDC ($1.4*\sqrt{2}*1.64$) was 3.3. An individual should accordingly improve more than 3.3 on the BARS sum scale (12–84) to demonstrate a change above measurement error.

Construct validity

Hypotheses 1–8 regarding moderate correlations between the BARS sum score and all SF-36 subscales were confirmed except for Role Emotional ($r_s = 0.24$): Physical Functioning: $r = 0.37$, Physical Role Functioning: $r = 0.36$, Bodily Pain: $r = 0.33$, General Health: $r_s = 0.45$, Vitality: $r_s = 0.38$, Social Functioning: $r_s = 0.46$ and Mental Health: $r_s = 0.36$. Hypothesis 9 regarding moderate correlation between BARS and GPSES was also confirmed: $r_s = 0.46$ (Table IV). The moderate correlations were all statistically significant ($p \leq 0.05$).

The last hypothesis that patients with musculoskeletal conditions would demonstrate lower (worse) scores on the BARS movement quality scale than healthy persons was also confirmed ($p = 0.001$). The mean score (± SD) for patients was 46.3 ± 10.5, ranging between 23.5 and 63.5, and for healthy persons 55.4 ± 8.5, ranging between 38.0 and 75.5. The distribution of scores is presented in box plots, showing some overlap of scores between groups (Figure 4).

Table III. Correlation between separate items and the Body Awareness Rating Scale (BARS) total movement quality score, and the alpha value if an item is deleted.

BARS items	Corrected item-total correlation	Cronbach's α if item deleted
1	0.567	0.922
2	0.590	0.921
3	0.701	0.918
4	0.691	0.917
5	0.717	0.916
6	0.750	0.914
7	0.637	0.919
8	0.695	0.917
9	0.734	0.915
10	0.659	0.918
11	0.741	0.915
12	0.740	0.915

Internal consistency (α) = 0.924.

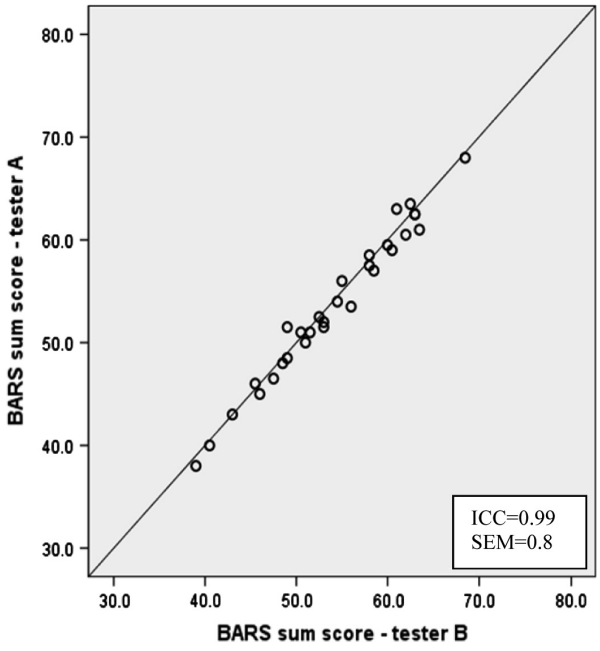

Figure 2. Inter-tester reliability of Body Awareness Rating Scale (BARS) movement quality scale.

Discussion

The BARS is used to assess movement quality in patients with long-lasting musculoskeletal disorders and mental health problems. BARS is an important tool because of its multi-perspective view of movement, for its broad scope of daily life movements, lying, sitting, standing, relational and walking, and its relation to health and self-efficacy. The physiotherapist takes the whole body into consideration when observing and assessing movement quality. The healthy movement aspects are particularly in the foreground. BARS is well structured regarding criteria for observation, movement guidance and communication with the patient. The scale is also used as a communication tool, awakening the patients' attention and adjustments of movement quality. In this study, indications of satisfactory internal consistency, inter-tester and test–retest reliability and construct validity was provided.

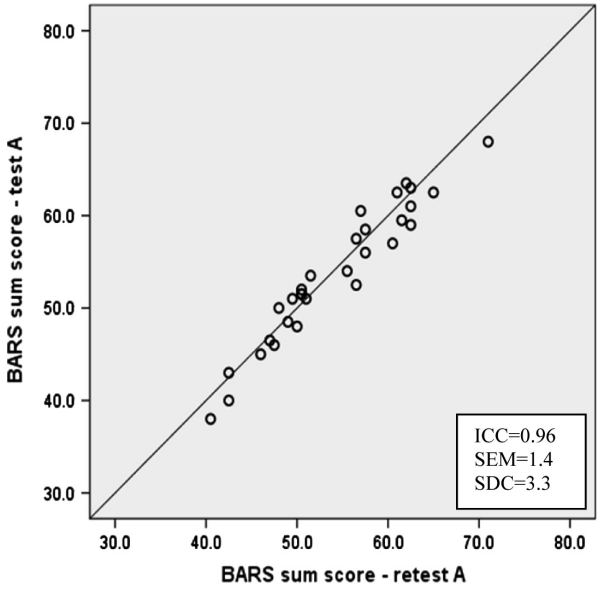

Figure 3. Test–retest reliability of Body Awareness Rating Scale (BARS) movement quality scale.

Internal consistency

Unidimensionality is a prerequisite for examining internal consistency of a scale (29). As unidimensionality was established for the movement quality scale in a previous study (4,9,10), it was not examined in the present study. Internal consistency was found highly satisfactory taking into account that BARS only contains 12 items. No item increased the total alpha value if it was deleted from the scale. Item-total correlations indicated that all the items contributed to measure a joint construct. As the alpha was found to be above 0.90, the scale might have been reduced to contain fewer items (30), making the scale more feasible for clinical use. It is, however, stated that many good questionnaires have higher alphas than 0.90, and a positive quality rating is given for a measure if Cronbach's α is between 0.70 and 0.95 (31).

Inter-tester reliability

An almost perfect inter-tester reliability (ICC = 0.99) was demonstrated in the study. Both testers were experienced in the use of BARS, working with persons suffering from musculoskeletal disorders and mental health problems. The almost perfect concordance in scores can partly be explained by that, but also to some extent by the procedure chosen for scoring. In this study, only one tester instructed, observed and scored the BARS movements while the other tester only observed and scored. This procedure might have violated the principle of independent administrations mentioned in the COSMIN checklist for methodological quality of studies on measurement properties (29). If the assessments had been performed on different occasions, but on the same day, reliability values would probably be somewhat lower, as variability in instruction and communication could cause differences in movement performances. The result of our study demonstrated nevertheless that two skilled physiotherapists assessed movement quality very similarly. Measurement error was also very low (SEM = 0.8), indicating that the

Table IV. Hypotheses tested of moderate correlation between the Body Awareness Rating Scale (BARS) movement quality scores, Short-Form Health Survey (SF-36) subgroups and General Perceived Self-Efficacy Scale (GPSES), examined by Spearman's correlations (r_s); $n = 50$.

	Measures		Hypothesis	
Hypothesis no.	SF-36 Subscales:	r_s	$0.30 \leq r_s < 0.60$	Hypothesis confirmed
1	Physical Functioning	0.37**		Yes
2	Physical Role Functioning	0.36*		Yes
3	Bodily Pain	0.33*		Yes
4	General Health	0.45**		Yes
5	Vitality	0.38**		Yes
6	Social Functioning	0.46**		Yes
7	Role Emotional	0.24		No
8	Mental Health	0.36*		Yes
9	GPSES	0.46**	$0.30 \leq r_s < 0.60$	Yes

*p, 0.05, **$p < 0.01$.

skilled therapists to a high degree agreed in their judgment of movement quality. Whether less skilled therapists can also obtain satisfactory inter-tester reliability remains to be seen.

Test–retest reliability

The stability in BARS scores from test to retest after 4 days examined by tester A was also highly satisfactory (ICC = 0.96). Measurement error in test–retest reliability can arise due to an inconsistency in scoring by the tester from one assessment to another, and can as well be a change in movement performance by the person observed due to day-to-day variability. One can also expect a systematic drift in scores in the direction of improvement from one assessment to another due to a learning effect. There was, however, no systematic drift in data in the present study. The data were rather symmetrically distributed around the diagonal representing close to perfect concordance in scores. The SDC value is useful in clinical practice, as it indicates how much variability that simply can be due to measurement error, taking error in the two assessments into consideration. To claim a treatment effect with 95% certainty, an improvement of more than 3.3 on the 12–84 point scale is needed. We conclude that movement quality assessed according to BARS in a test–retest situation can be performed with high reliability when assessment is performed by a skilled BARS's therapist. Whether the test–retest reliability values apply to therapists with less experience in using BARS needs to be examined in a broader group of therapists.

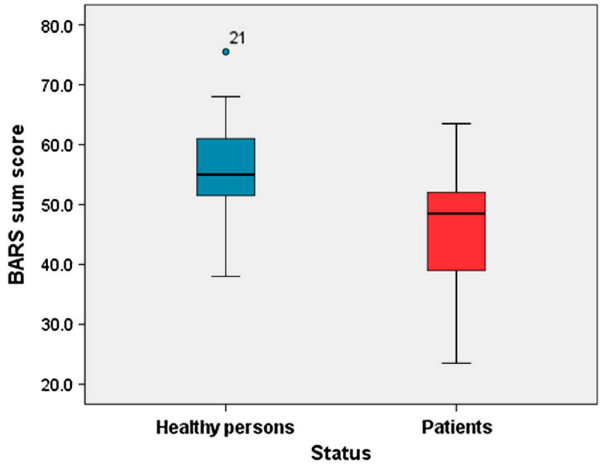

Figure 4. Box plots showing distribution of Body Awareness Rating Scale (BARS) movement quality scores in healthy persons ($n = 25$) and patients with musculoskeletal disorders ($n = 25$).

Construct validity

We hypothesized that movement quality assessed in BARS is a bodily expression of physical and mental health and self-efficacy, and evidence of construct validity was in fact provided in the study. Moderate correlations was demonstrated between BARS and seven SF-36 subscales, meaning that seven hypothesis were confirmed. We also found moderate association ($r_s = 0.46$) between BARS and GPSES assessing self-efficacy, or the ability to cope with a variety of difficult demands in life, meaning that Hypothesis 9 was confirmed. A similar association was also shown in a former study (33). Taking into account that both SF-36 and GPSES are self-report measures while BARS is an observational measure, the associations seem rather high.

BARS has been developed to be used with people suffering from musculoskeletal disorders and mental health problems. Consequently, we expected to find more dysfunctional movement quality in patients suffering from musculoskeletal disorders than in healthy persons. This hypothesis (no. 10) was also confirmed. Although the difference was statistically significant,

some overlap in scores was demonstrated, showing that persons who reported to be healthy might also demonstrate tension, restricted breathing and lack of contact with the body, even though they were working and not seeking help for their complaints.

Earlier studies have shown that the treatment counterpart of BARS, the BBAT fosters health resources, like ability to cope with symptoms of pain and fatigue, as well as ability to cope with a difficult situation (14,15,34,35). There are, accordingly, indications that BARS capture important health dimensions, by assessing how total movement co-ordinations are performed. More studies are, however, required to examine construct validity, including other self-report and performance measures of movement quality.

Acquiring competence in observing movement quality

Because of BARS movement analysis from a physical, psycho-socio-cultural and existential perspective, and assessment, in particular, of healthy movement aspects, skill training in using defined criteria is needed to obtain less subjective and more reliable and valid observations. A learning program in BARS of two 4-day courses with a 3-month intervening period of clinical implementation is offered to gain such competence.

In the assessment situation, each of the 12 BARS' movements is followed by an interview on "How did you experience this movement?" The type of ethical reflections, supporting the patient to stay present in the body and, at the same time, verbalize bodily sensations of own movement potentials, is necessary to uncover in patients suffering from long-lasting musculoskeletal disorders and mental health problems. The two roles of the physiotherapist, being observer and listener, are performed as separate and not mixed roles in BARS (19).

Limitations of the study

A sufficiently large sample size is recommended when examining measurement properties of questionnaires (29), and at least 50 participants are recommended for reliability and validity studies (31), although larger samples are preferred. The 50 participants included in the study of internal consistency and construct validity comply with these recommendations, while the number of participants examined for inter-tester and test–retest reliability should preferably have been higher than 30. Usually studies on measurement properties of physical tests include a lower number of participants than 50 as the task of obtaining a high number of data usually is more demanding than collecting data from self-report measures. Each BARS assessment implies, for instance an examination that lasts approximately 1 h. Due to this methodological challenge, the reliability estimates of physical tests may become less robust.

The BARS sum score ranges between 12 and 84, and no ceiling or floor effect was demonstrated in the present study with scores ranging between 24 and 76. The fact that we did not evaluate the entire scale, including scores at the upper and lower ends, is a limitation. To observe healthier aspects (higher scores) we could have invited athletes and dancers. To observe more pathological aspects (lower scores), patients with psychiatric diagnoses could have been invited. Movement quality according to BARS is considered a general norm, but it remains to be examined whether quality criteria should take age and gender into consideration. It was further a limitation that the group of patients and healthy persons were non-equivalent regarding demographic characteristics such as gender and education. There was a higher percentage of women in the patient group and more persons with higher education in the healthy group. For a later study, the demographic data of the study samples should be more similar.

Conclusion

The study showed high internal consistency, high inter-tester and test–retest reliability, and low measurement error when BARS was used by qualified testers. Evidence was provided that BARS reflects important aspects of health, functioning and coping abilities. Reliability and measurement error should be examined in a broad group of testers that are more or less experienced in the BARS assessment.

Acknowledgement

We thank Bergen University College, Norway, the University of Lund, Sweden, and the University of Bergen, Norway, for support.

Funding

This project was funded by the Norwegian Fund for Post-Graduate Training in Physiotherapy, Oslo and Bergen University College, Norway.

Declaration of interest

The authors report no conflicts of interest.

References

1. Hagberg M, Violante F, Bonfigliolo R, Descatha A, Gold J, Evanoff B, et al. Prevention of musculoskeletal disorders in workers: Classification and health surveillance – Statements

of the Scientific Committee on Musculoskeletal Disorders of the International Commission on Occupational Health. BMC Musculoskel Disorders. 2012;13:109.
2. Laisnè F, Lecomte C, Corbière M. Biopsychosocial predictors of prognosis in musculoskeletal disorders – A systematic review of the literature. Disabil Rehabil. 2011;34:355–82.
3. Dropsy J. Vivre dans son corps. Expression corporelle et relations humaines (Living in your body – Bodily expression and human contact). Paris: Epi SA; 1973.
4. Skatteboe UB. Basal kroppskjennskap og bevegelsesharmoni. Videreutvikling av undersøkelsesmetoden Body Awareness Rating Scale, BARS-Bevegelsesharmoni (Body awareness therapy and movement harmony. Further development of the assessment scale Body Awareness Rating Scale). Oslo: Oslo University College, Faculty of Health Sciences; 2000.
5. Brown W, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. J Pers Soc Psychol. 2003;84:822–48.
6. Alon R. Mindful spontaneity. Moving in tune with nature: Lessons in the Feldenkrais method. Chalmington, Dorset: Prism Press; 1990.
7. Brodal P. The central nervous system. Structure and function. Oxford: Oxford University Press; 2010.
8. Damasio A. Self comes to mind: Constructing the conscious brain. New York: Vintage Books; 2012.
9. Skatteboe UB, Friis S, Hope MK, Vaglum P. Body awareness group therapy for patients with personality disorders. 1. Description of the therapeutic method. Psychoter Psychosom. 1989;51:11–7.
10. Friis S, Skatteboe UB, Hope MK, Vaglum P. Body awareness group therapy for patients with personality disorders. 2. Evaluation of the Body Awareness Rating Scale. Psychother Psychosom. 1989;51:18–24.
11. Dropsy J. Le corps bien accordè – Un exercise invisible (The harmonic body – An invisible exercise). Paris: Desclée De Brouwer; 1984.
12. Dropsy J. Human expression – The coordination of mind and body. In: Skjaerven LH, editor. Quality of movement – The art and health. Lectures on philosophy, theory and practical implications to Basic Body Awareness Therapy. Bergen: Bergen University College; 1998. pp 8–20.
13. Roxendal G. Body awareness therapy and the Body Awareness Scale, treatment and evaluation in psychiatric physiotherapy [Doctoral dissertation]. Göteborg: University of Göteborg and Psychiatric Department II, Lillhagen Hospital, Hisings Backa; 1985.
14. Mattsson M. Body awareness applications in physiotherapy [Doctoral Dissertation]. Umeå: Umeå University; 1998.
15. Gyllensten AL. Basic Body Awareness Therapy [Doctoral Dissertation]. Lund: Lund University; 2001.
16. Skjærven LH, Gard G, Kristoffersen K. Basic elements and dimensions to quality of movement – A case study. J Bodyw Mov Ther. 2003;7:251–60.
17. Skjærven LH, Gard G, Kristoffersen K. Greek sculpture as a tool in understanding the phenomenon of movement quality. J Bodyw Mov Ther. 2004;8:227–36.
18. Skjaerven LH, Kristoffersen K, Gard G. An eye for movement quality: A phenomenological study of movement quality reflecting a group of physiotherapists' understanding of the phenomenon. Physiother Theory Pract. 2008;24:1:13–27.
19. Skjaerven LH, Kristoffersen K, Gard G. How can movement quality be promoted in clinical practice? A phenomenological study of physical therapy experts. Phys Ther. 2010;90: 1479–92.
20. Ware JE, Gandek B. The SF-36 Health Survey: Development and use in mental health research and the IQOLA project. Int J Ment Health. 1994;23:49–73.
21. Ware JE. SF-36 Health Survey update. Spine. 2000;25: 3130–9.
22. Ware JE, Kosinski M. SF-36 Physical and Mental Health Summary scales: A manual for users of version 1. Lincoln, RI: QualityMetric, Inc; 2001.
23. Ware JE, Kosinski M, Bayliss MS, McHorney CA, Roger WH, Raczek A. Comparison of methods for the scoring and statistician analysis of SF-36 health profile and summary measures: Summary of results from Medical Outcome study. Med Care. 1995;33:264–79.
24. Heyland DK, Hopman WH, Tranmer J, McColl MA, Coo H, Tranmer J, et al. Long-term health-related quality of life in survivors of sepsis. Short Form 36: A valid and reliable measure of health-related quality of life. Critic Care Medic. 2000; 28:3599–605.
25. Schwarzer R, Jerusalem M. General Perceived Self Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. Measures in health psychology: A user's portfolio. Windsor: NFER-NELSON; 1993. pp 33–37.
26. Schwarzer R, Jerusalem M. Generalized Self Efficacy Scale (GSES). Weinman J, Wright, S, Johnston, M, editors. Windsor: NFER-NELSON; 1995.
27. Leganger A, Kraft P, Røysand E. Perceived self-efficacy in health behavior research: Conceptualization, measurement and correlates. Psychol Health. 2000;1:51–70.
28. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale (GSES). Johnston M, ed. Windsor: NFER-NELSON; 1995.
29. Mokkink LB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–45.
30. de Vet H, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Practical guides to biostatistics and epidemiology. Cambridge: Cambridge University Press; 2011.
31. Terwee C, Bot SD, De Boer MR, van der Windt DA, Knol DA, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42.
32. Andresen E. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81(12suppl): 15–20.
33. Sundal MA. Bevegelseskvalitet – Et uttrykk for helse og velvære? (Movement quality – An expression of health and well-being?) [MSc]. Bergen: University of Bergen; 2007.
34. Mattsson M, Egberg K, Armelius K, Mattsson B, Nord J. Long-term effects of physiotherapeutic treatment in outpatient psychiatric care. Nord J Psychiat. 1995; 49:103–10.
35. Malmgren-Olsson EB, Brenholm IB. A comparison between three ohysiotherapy approaches with regard to health-related factors with non-specific musculoskeletal disorders. Disabil Rehabil. 2002;6:3008–316.

Supplementary material available online

Supplementary Appendix 1–2.