Published in final edited form as:
Pain. 2013 December ; 154(12): . doi:10.1016/j.pain.2013.08.024.

DEVELOPMENT AND VALIDATION OF A NEW SELF-REPORT MEASURE OF PAIN BEHAVIORS

Karon F. Cook[1], **Francis Keefe**[2], **Mark P. Jensen**[3], **Toni S. Roddey**[4], **Leigh F. Callahan**[5], **Dennis Revicki**[6], **Alyssa M. Bamer**[3], **Jiseon Kim**[3], **Hyewon Chung**[7], **Rana Salem**[3], and **Dagmar Amtmann**[3]

[1]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL

[2]Department of Psychiatry and Behavioral Science, Duke University Medical Center, Durham, North Carolina, USA

[3]Department of Rehabilitation Medicine, University of Washington, Seattle, WA

[4]School of Physical Therapy, Texas Woman's University Houston Center, Houston, TX

[5]Thurston Arthritis Research Center, Department of Medicine, University of North Carolina, Chapel Hill, NC

[6]Center for Health Outcomes Research, United BioSource Corporation, Bethesda, MD

[7]Department of Education, College of Education, Chungnam National University, Yuseong-gu Daejeon, South Korea

Abstract

Pain behaviors that are maintained beyond the acute stage post-injury can contribute to subsequent psychosocial and physical disability. Critical to the study of pain behaviors is the availability of psychometrically sound pain behavior measures. In this study we developed a self-report measure of pain behaviors, the Pain Behaviors Self Report (PaB-SR). PaB-SR scores were developed using item response theory and evaluated using a rigorous, multiple-witness approach to validity testing. Participants included: a) 661 survey participants with chronic pain and with multiple sclerosis (MS), back pain, or arthritis; b) 618 survey participants who were significant others of a chronic pain participant; and c) 86 participants in a videotaped pain behavior observation protocol. Scores on the PaB-SR were found to be measurement invariant with respect to clinical condition. PaB-SR scores, observer-reports, and the video-taped protocol yielded distinct, but convergent views of pain behavior, supporting the validity of the new measure. The PaB-SR is expected to be of substantial utility to researchers wishing to explore the relationship between pain behaviors and constructs such as pain intensity, pain interference, and disability.

Corresponding Author: Karon F. Cook, PhD, c/o Renetta Scurlock, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 625 N. Michigan Ave., Suite 2700, Chicago, IL. 60611, Phone: 713.291.3918 Fax 312-503-4800, karon.cook@northwestern.edu.

Keywords

Chronic pain; assessment; psychometrics

INTRODUCTION

In the past several decades, there have been substantial shifts in how pain is conceptualized. Once viewed strictly from a biomedical model, pain was defined as a response to disease or tissue damage. As evidence for psychosocial determinants of pain has accumulated, however, the biomedical model has proven inadequate for explaining the perception and impact of pain. For example, Keefe and colleagues found coping strategies to be more predictive of self-reported knee pain than X-ray evidence of disease. [13]

Fordyce was perhaps the first behaviorally-oriented researcher to explicate the important role of pain behaviors and their contribution to disability. [11] Pain behaviors are overt (i.e., observable and quantifiable) behaviors that can communicate pain to others, and may include, among other displays, resting, guarding, facial expressions, asking for help, and taking medication. Though often protective or effective for eliciting support and assistance, when maintained beyond the acute stage post-injury, such behaviors can contribute to subsequent psychosocial and physical disability. [18] Moreover, research suggests that pain behaviors are useful treatment targets for behavioral interventions; as pain behaviors decrease, so can subsequent pain intensity and pain-related impairment. [11]

Critical to the study of pain behaviors and their role in chronic pain is the availability of psychometrically sound pain behavior measures. Historically, pain behavior assessments obtained data using behavioral interviews, reports by significant others, daily diaries, or trained observers applying behavior sampling procedures. For example, one standardized approach to pain behavior sampling, developed in the mid-1980's, used direct observation of pain behaviors while patients moved through a series of standard behavioral tasks (i.e. sitting, standing, walking, reclining) likely to elicit pain. [12]

Another strategy is to measure pain behaviors using standardized self-report instruments. This approach assumes that the propensity to exhibit pain behaviors is a latent trait that can be estimated reliably based on persons' subjective reports of the relative frequencies with which they engage in pain behaviors. This measurement strategy is based on trait theory, and is conceptually different from the origins of pain behavior assessment with its roots in behavioral and cognitive-behavioral psychology and functional analysis. [12] Few studies have compared different strategies for measuring pain behaviors and, to our knowledge, none have compared self-report, significant-other reports, and behavior observation protocols in the same sample.

The first purpose of this study was to develop a self-report measure of pain behaviors based on item response theory (IRT), a methodologically rigorous and model-based approach to the development of measures, [5] and to link the resulting scores to a norm-referenced metric. The second purpose was to evaluate scores on the new measure using a "multiple witnesses" approach to validity testing, [28], i.e. comparing self-report, significant other reports, and behavioral observation. We refer to this new measure as the Pain Behaviors Self Report (PaB-SR) scale.

METHODS

The development and validation of the PaB-SR was an iterative process and included multiple study designs and data collection efforts. Validation information came from several

sources and was used to select and reject items for inclusion in the measure (based on explorations of item-level validity) and to evaluate the validity of scores on the finalized PaB-SR. For clarity, we begin with a description of the identification of items for the candidate PaB-SR. We then describe three data collection efforts detailing the participants, measures, and procedures for each data collection. Finally, we describe how the collected data were analyzed and evaluated in the development and testing of the PaB-SR.

DEVELOPMENT OF CANDIDATE ITEM POOL FOR THE PAB-SR

For the purposes of this study, pain behavior was defined as "behaviors that typically indicate to others that an individual is experiencing pain" [19] and include non-verbal displays, sighing, crying, resting, guarding, and facial expressions, as well as verbal reports (e.g. asking for help, verbal pain ratings). [19] In a series of successive meetings, the research team defined the range of content consistent with this definition, discussed the most relevant response categories for assessment, and reviewed the content of published scales of self-reported pain behavior. We used NIH's Patient Reported Outcome Measurement Information System [19] Pain Behaviors item bank (including items excluded by PROMIS investigators) as a starting point. [21] The team also developed new items that extended the content range of the candidate PaB-SR item pool. No additional items were found in other published measures.

Like the PROMIS item bank, the time frame for the candidate PaB-SR items is "in the past 7 days." However, we elected to use a different response option set than the one used by PROMIS investigators. The PaB-SR response options are: "never", "rarely", "sometimes", "often", and "always". The PROMIS Pain Behavior items have an additional response option — "had no pain". This additional option was included by PROMIS because PROMIS measures were developed to be appropriate both across chronic conditions and in the general population, including individuals who are not experiencing pain. In contrast, we developed the PaB-SR to be administered only to those experiencing pain. Another difference in the two measurement initiatives was that the PROMIS bank was developed to the specifications of an item bank. For item banks, the number of usable items is maximized and multiple items targeting the same range of the scale are allowed. [3] The PaB-SR, however, was designed as a static short form; thus, the focus was on retaining items that covered the range of the domain being measured but minimized item redundancy.

DATA COLLECTION

Three different studies were conducted in developing and validating the PaB-SR. Below, we describe the purpose of each study, the sample recruited for the study, the study measures, and the study procedures. Each sample included participants with one of three target diseases—arthritis, low back pain, or multiple sclerosis [27]. For reference and clarification, we have summarized this information in Table 1. All research procedures were reviewed and approved by the institutional Review Boards (IRBs) of the University of Washington, and the University of North Carolina, Chapel Hill. The IRBs applied research ethics standards for human participants.

Cross-sectional Survey Study—A cross-sectional study was completed to evaluate concordance between self-reports of pain behaviors to reports of an observer. Below we first describe the sample and then describe the survey contents and study procedures.

Sample: The sample for this study is referenced in Table 1 as Sample A. Participants were recruited from a variety of sources including web postings and research and clinical sites at the University of Washington, Seattle and the University of North Carolina, Chapel Hill. Pairs of participants with pain and their significant others (SOs) were recruited. SOs were

nominated by participants who had pain. We did not limit the choice to a particular relationship (e.g., spouse, caretaker). Only SOs who reported spending at least 7 hours per week with the participant with pain were included in the study.

Survey Measures: Separate surveys were developed for participants with pain and their SOs. Both surveys included demographic items. Participants with pain also answered items about their clinical status (e.g., time since MS diagnosis). Additional measures are described below by sample.

Survey of Participants with Pain: Participants with pain responded to all 46 candidate items for the PaB-SR. In addition they completed the six items of the PROMIS Pain Interference Short Form v1.0 (http://www.nihpromis.org/). Pain intensity was measured as "average pain" over the past 7 days reported on a scale from 0 (no pain) to 10 (pain as bad as could be). In addition, participants with pain completed the Pain Behaviors Checklist, [14] a 17-item scale in which participants report frequency of pain behaviors on a 0 to 6 scale where 0 = "never", and 6 = "very often".

SO Survey: SOs responded to 46 items that were exact parallels to the candidate items administered to participants with pain. For example, the parallel item for the item, "In the past 7 days, when I was in pain I took breaks" was, "In the past 7 days, when in pain my significant other took breaks". SOs also completed the items of the University of Alabama in Birmingham (UAB) Pain Behavior Scale. [22] Developers report good inter-rater reliability (0.95) and temporal stability (0.89) for scores on this measure. The10-item scale was designed to be administered by medical personnel but, in the current study, SOs completed the items. Its usage in the current study, therefore, is unique to this study. As such, results were interpreted cautiously and used to make relative comparisons among candidate items and provide a comparison with final PaB-SR scores.

Procedures: Participants with pain and their SOs were asked to complete their surveys independently but within a 48 hour time period of each other. To evaluate the validity of combining data from those pairs who responded within the requested timeframe and those who did not, associations and correspondence between responses by SOs and participants with pain were compared separately by time frame compliance. Based on the results, a decision was made whether to combine data for subsequent analyses.

If only one survey of a pair was received, reminder calls to the other member of the pair were completed immediately and 1–2 days after receipt of the first survey. If neither survey was returned, reminder calls were made (1–2 weeks and 2–3 weeks after surveys were sent). Participants with missing data were called or emailed up to three times to obtain missing responses. Participants with pain and SOs were paid $20 each for participation in the survey study. Sample A is the only study that included SOs

Sample B: Score Linking Study—The purpose of the Score Linking Study was to create a mathematical bridge between PaB-SR scores and PROMIS T-score metric. Because the PROMIS items have a different response scale than the PaB-SR items, it was necessary to collect additional data to link PaB-SR items to the PROMIS metric. Sample B served this purpose and consisted, predominantly, of a random sample of individuals with pain who had completed the study with a paired SO (Sample A). In addition, persons who had already participated in a videotaped pain behavior protocol (Sample C; described below).

Measures: A brief survey was constructed that included the final set of 20 items that eventually constituted the PaB-SR. At the time this sample was recruited, 25 items had been dropped from consideration based on results of psychometric analyses and content review.

(Note: Eventually, an additional item was dropped leaving 20 items included in the PaB-SR). Of the 21 items included in the survey, 14 were new items written by the PaB-SR team. Seven had the same item stem as a PROMIS Pain Behavior item, but had different response categories. The survey included the 14 new items with the 5-point PaB-SR response scale along with the seven PROMIS items on a six-point response scale. These data were used to anchor the PaB-SR item parameter estimates to the PROMIS published parameter estimates as described in the analytic methods.

Procedures: Individuals who expressed interest in the study and reported having pain were mailed paper surveys and asked to return them using a provided, self-addressed, posted envelope. Participants were paid $25 for completing the survey.

Sample C: Pain Observation Study—Sample C participants were recruited from the sample of individuals with pain who completed the study with a paired SO (Sample A). Clinician referrals, advertisements, flyers, and web postings augmented this sample. An invitation letter that included a contact number was sent to potential participants. Interested individuals who met eligibility criteria were scheduled for in-person sessions at the study office. Participants were videotaped doing a standard set of daily activities that included sitting, standing, and walking

Measures: Immediately prior and immediately after the videotaping sessions, participants reported their "current pain" on a 0–10 scale where '0' indicated "no pain" and '10' indicated "pain as bad as you could imagine". Note that the wording of the most severe anchor, "pain as bad as you could imagine" is slightly different from the wording of the parallel anchor for Sample A, "pain as bad as could be". This was an unintentional variation that occurred in the development of the forms. We judged the scores to be roughly comparable for the purposes of the current study. Either immediately before or after the videotaping sessions, participants completed a survey similar to that administered to study Sample A. A body pain map also was completed to indicate locations of participants' pain. The pain map is a graphic showing two outlines of a body, one facing forward and the other facing backward. [8] Rubbing behaviors were only counted if they were associated with an area on the pain map that participants' indicated as a location of their pain.

Procedures: Using the behavior sampling method developed by Keefe and Block, [12] participants in the pain observation study were videotaped doing a standard set of typical every day activities. The activities included sitting (1- and 2-minute intervals), standing (1- and 2-minute intervals), reclining (two 1-minute intervals), and walking (two 1-minute intervals). The order was randomized to minimize order effects. Each videotaped session took approximately 10 minutes. Participants were informed that the person videotaping them would not be interacting with them during the activities. Participants were paid $50 for completing the observation session and the survey.

A physical therapist on the research team (Roddey) was trained by one of the developers of the pain behavior observation protocol (Keefe) in a series of sessions in which both investigators coded videotapes archived from another pain behavior study. Training continued until agreement of 85% or greater was routinely reached. The pain behavior codes included *guarding, bracing, grimacing, rubbing* and *sighing*. Coding of each behavior was done using an interval recording method that used a 20 second observe, 10 second record interval. Thus coding of each pain behavior was dichotomous—it was coded as present or absent for each interval. Full descriptions of the pain behavior protocol have been published. [13; 16] Briefly, *guarding* can occur during sitting, standing, or reclining and includes stiffness, rigid movements, and use of canes or walkers during walking intervals; *bracing* can occur during sitting, standing, or reclining and is defined as at least 3 consecutive

seconds in which a limb is extended to support and maintain an abnormal distribution of weight; *grimacing* includes obvious facial expressions such as furrowed brow; and *sighing* is air exhalation that is obvious or exaggerated.

ANALYSES

Item Level Analyses

Evaluation of Item Response Theory Assumptions: We explored whether the collective set of items could be scaled using a unidimensional IRT model. An IRT unidimensional model does not require that there are no subdomains (e.g. guarding, rubbing). In fact, few if any datasets are strictly unidimensional. A finding that the response data were "essentially unidimensional", however, would indicate the possibility, but not the requirement, that a general pain behavior scale could be developed and meaningfully scored. In addition to essential unidimensionality, IRT models assume that the responses are locally independent. Unidimensionality is present when the predominant share of the variance in item scores is accounted for by the first factor. In other words, a single factor (the trait being measured) is driving how people respond to items. We evaluated unidimensionality by calculating hierarchical coefficient omega, a statistic that estimates the amount of variance accounted for by the general factor. [27] Coefficient omega values greater than 0.50 indicate that the general factor drives most of the variance in scores. [20] For completeness, we also evaluated the more commonly reported measure of internal consistency, coefficient alpha.

After assessing dimensionality of the data, we calibrated item responses to the graded response model, [23] an IRT model appropriate for data from items that have more than two response options. The calibration was accomplished using the software, IRTPRO. [4] This calibration allowed us also to evaluate a second assumption of IRT, local independence. The local independence assumption is that the trait measured (here, pain behaviors) accounts for the variance in scores. Local dependency occurs when, after the modeled variance is removed, there is remaining variance (residual variance) between scores of pairs of items. This indicates that the items have substantial residual variance in common (variance not accounted for by the trait being measured). We evaluated local dependency by calculating standardized LD X^2 values that are output in an IRTPRO calibration. Chen and Thissen suggest using a cut off of LD greater than ten as indicating likely local dependence. [6]

Selecting Final Items: Items were selected iteratively based on results from psychometric analyses. Because most items had good associations with other measures, the driving considerations for selection were content validity and local dependency. To minimize local dependency, items that shared substantial local dependency were grouped into item subsets. Investigators ranked items within subsets according to their preferences of which items to retain. The retained items were then recalibrated. This process continued for several iterations until we identified a subset of the candidate items that met IRT assumptions and fit the graded response model. [23]

Review for Content Validity: During the development of the candidate item pool, steps were taken to ensure that the universe of potential pain behaviors was properly sampled. This was accomplished by attending to the following questions: a) Do these items measure the construct of interest, pain behaviors? and, b) What content, if any, is missing; that is, are there subdomains of pain behaviors that are not represented in this item pool? The final set of items for the PaB-SR, however, only included 20 of the original 46 candidate items. To ensure that this smaller set of items retained content validity, we circulated the surviving 20 items along with the 46 excluded items. Study team investigators reviewed the items to determine if, in their opinion, the domain content was adequately represented in the retained items or if there were items they wished to nominate for re-introduction into the measure.

Linking PaB-SR Scores to the PROMIS T-Score Metric—When using an IRT model, scores from one measure can be linked to the metric of another if you have a sample in which there is overlap between items of each measure. The overlapping items, called "linking items", serve as a mathematical bridge between measures. There were seven items among the retained PaB-SR items that had the same item stem as a PROMIS Pain Behavior item. We tested these items for measurement invariance with respect to the PROMIS calibration sample and Sample A of the current study. Measurement invariance exists if items measure similarly across two or more subgroups. This is an important property for pain measures that are used to compare pain level across different clinical or other subgroups because it allows the distinction between: a) real differences in diagnostic groups and b) apparent differences that are due to something other than actual differences in persons' levels of pain behavior. We tested measurement invariance at the item response level by testing for differential item function (DIF). DIF occurs when subgroups of people, *who have the same level of the trait being measured*, respond differently to the items based on their subgroup classification. A subgroup of persons who have MS and a subgroup of persons with back pain, both having the *same level of pain behaviors*, should answer the pain behavior items the same. Note that this does not mean that different subgroups would have the same or even similar levels of the trait being measured, only that the items would function similarly in different groups. Another way to say this is, the level of pain behaviors should drive responses, not subgroup membership. In the current study, we tested for DIF using the software, LORDIF. [7] LORDIF uses an ordinal logistic framework. We used a criterion of beta change >0.1 as indicating meaningful DIF. Only items that did not exhibit meaningful DIF across samples were used for linking the items to a common metric.

We combined responses to the measurement invariant linking items and the new PaB-SR items in a single data set and calibrated them to the graded response model using IRTPRO. [4] In this calibration, the PROMIS item parameters were fixed (not estimated) at the published values for the PROMIS Pain Behavior item bank. [21] This put the new items on the same mathematical metric as the PROMIS measures; however, it did not accomplish the same for the PaB-SR items that had PROMIS stems, but only had five response categories. To calibrate these estimates, we did another IRT calibration using the original survey data. In this second calibration, the new PaB-SR item parameters were fixed at the values obtained in the first linking calibration and those of the other items were estimated. This completed the bridge with the PROMIS T-score metric. We then developed a concordance table that associated raw PaB-SR scores with their associated T-scores.

Validation of PaB-SR Scores

Associations with Observer- and Self-Reports: We considered the degree to which PaB-SR item and scale scores were associated with SO reports and scores on other pain measures. This was accomplished using responses of participants in the cross-sectional survey of participants with pain and their SOs (Sample A—see Table 1). In this sample Spearman correlations and deviation scores (pain participant score minus SO score) were calculated between item and scale scores and scores on other pain measures and between SO reports and pain participant reports. We also evaluated the association between self-reports and results of the videotape study.

Scores across Diagnostic Categories: Before using scores on a measure to compare levels of pain behaviors in different diagnostic categories, it is important to ascertain whether the items function similarly in different samples; i.e., scores do not exhibit DIF across diagnoses. Ideally, only level of pain behavior (and not other factors, such as demographics or clinical condition) drives responses to a pain behavior items. DIF occurs when subgroup membership also plays a role in how an individual responds to the item; i.e., the item is not

measurement invariant with respect to group membership. DIF is a problem in comparing scores of groups because, when there is DIF, it is unclear whether score differences between groups is due to actual difference in levels of the trait being measured or due to differences in the way members of different subgroups understand and respond to items. To evaluate DIF with respect to diagnostic category, we used ordinal logistic regression and the software, LORDIF, [7] the same approach we used in evaluating measurement invariance of linking items across samples (PROMIS vs. the current study). The same DIF criteria (if change in beta coefficient > 0.10, item is flagged as having DIF). If an item had DIF with respect to diagnostic category it was dropped from consideration for the PaB-SR. Thus scores on the PaB-SR are measurement invariant with respect to diagnostic criteria and, therefore, useful for comparing levels of pain of persons with arthritis, back pain, and MS. Differences in scores were evaluated using analysis of variance (ANOVA) with $p<0.05$ accepted as indicating that pain behaviors differed by diagnostic category.

RESULTS

DATA COLLECTION

Sample A: Cross-sectional Study—A total of 747 participants (arthritis=257, back pain=271, and MS=219) were eligible for the study and willing to participate. Table 2 provides demographics and diagnostic information for the eligible survey pain participants. Surveys were completed electronically (*n*=316), on paper (*n*=427), and by phone (*n*=4). All participants with pain were included in the calibration of PaB-SR responses to the graded response model. For validity analyses, data from participants with pain that did not have a matching SO survey were excluded. A total of 1236 participants completed the survey (618 participants with pain and 618 SOs). Of the participants with pain in the paired data, 203 had MS, 209 had back pain, and 206 had arthritis.

The mean time delay between completion by both members of a pair was 2.5 days (SD=7.3; range: 0–113 days). The association between scores and mean deviation scores (participant with pain's response minus SO's response) were compared for pairs reporting within 2 days of each other (n=486) and those reporting >2 days apart (n=132). In calculating association and item score deviations, only the 20 items included in the final PaB-SR were used. The correlation coefficients obtained in the subset of participants who responded within and outside the two day response window were 0.60 and 0.62, respectively. The mean deviations in item scores for those reporting within and outside the two day window were −0.07 and −0.06, respectively. Based on these results, we judged it appropriate to combine all paired responses for subsequent analyses.

Sample B: Score Linking Study—In total, 473 individuals participated in the score linking study in which participants completed both PROMIS and PaB-SR items. Of these, 154 had back pain, 158 had MS, and 161 had arthritis. Demographics for these are included in Table 2.

Sample C: Pain Observations Study—A total of 86 individuals participated in the video-taped, pain observation protocol. Of these 30 had back pain, 30 had MS, and 26 had arthritis. Demographics are included in Table 2.

ANALYSES

Item Level Evaluation of Candidate PaB-SR Items—The correlations between individual item scores and other pain measures as well as the relationships between pain participant and SO reports are reported in Table 3. This information was provided to the

research team for use in their selection of items from among the full item pool. Items that were excluded from the final PaB-SR measure are shaded gray in Table 3.

Item-level Associations with Observed Pain Behaviors: The column in Table 3 labeled, "Number Observed Pain Behaviors", includes results from the 86 individuals who participated in the video-taped observation study (Sample C). The median correlation between total pain behaviors observed and item level scores was 0.33. Values ranged from −0.04 ("I took a hot bath or shower") to 0.61 ("I used a cane or something else for support").

Item-level Associations with Self-Report Pain Measures: The median correlations between item scores and PROMIS Pain Interference and Average Pain intensity were 0.47 and 0.38, respectively. Median item correlation with the Pain Behaviors Checklist was 0.49.

Item-level Associations with Other-Report Pain Measures: The median correlation between item scores and the UAB pain behavior scale was 0.36. The median correlation between participants with pain and SOs item level responses was 0.41. These ranged from 0.27 ("I shifted my position") to 0.75 ("I used a cane or something else for support"). Deviation scores, defined as pain participant's item response minus SO's item response, are reported in the final column of Table 3. The 26 negative values indicate items for which, on average, SOs reported higher levels of pain behavior than did their matched pain participants. The median deviation score was −0.06 indicating a slight trend toward higher pain behavior reports by SOs than by pain participants. The greatest discrepancy was in responses to the item, "I talked about the pain." SO reports of this behavior were higher by 0.62 on a 5-point scale compared to pain participant reports.

EVALUATION OF ITEM RESPONSE THEORY ASSUMPTIONS

Evaluations of IRT assumptions were conducted iteratively using the pain participant cross-sectional data (n=661; Sample A). Here we report the results for the final selected set of 20 items that constitute the PaB-SR.

Unidimensionality: PaB-SR item responses had an omega hierarchical estimate of 0.79. This indicates that 79% of the variance in PaB-SR scores was attributable to a common, general factor. These results supported sufficient unidimensionality of the data, i.e., that item responses were driven by one primary dimension. Therefore we proceeded to model responses using the graded response model. Coefficient alpha value for the item responses was 0.92.

Local Dependency: Substantial local dependency was observed among candidate PaB-SR items. For example, the items "I moaned" and "I groaned" had standardized LD X^2 values >10. This finding indicates that, after controlling for level of pain behaviors, these items were associated beyond the level expected by chance. After several iterations of item exclusions based on research team input, however, the final set of items had standardized LD X^2 values that ranged from −1.2 to 7.7.

FINALIZED PAB-SR ITEMS

The research team considered the item validity information reported in Table 3 along with information obtained in calibrations of different item sets to the graded response model (e.g., local dependency, item fit). Twenty-one items were selected for inclusion in the measures but, one item was dropped because it met our criterion for DIF with respect to diagnoses (change in beta >0.10). The first 20 items in Table 3 were selected for inclusion in the PaB-SR. After this selection was made, the team reviewed the content of the items to evaluate

whether the content coverage of the full candidate item pool had been retained. The team judged these items to be a fully representative sample of what constituted a representative sample of the universe of self-reported pain behaviors.

LINKING PAB-SR SCORES TO THE PROMIS METRIC

Scores were linked based on six anchor items found to be measurement invariant in our sample compared to the PROMIS calibration sample (one potential anchor item was dropped because of DIF with respect to sample—PROMIS vs. current study data). Table 4 is the concordance table that associates raw PaB-SR scores with their associated T-scores. The T-scores, and not the raw scores, are the final PaB-SR scores. The T-scores were used in all validity analyses.

EVALUATION OF VALIDITY OF PAB-SR SCORES

The final row of Table 3 reports the correlations between individual PaB-SR scores (on T-score metric) and other pain measures. As anticipated, correlations for PaB-SR scores and these measures are higher than correlations for item level scores and other pain measures. This likely is due to the greater reliability of multi-item measures and the restriction of range when item level scores are correlated with scores on other measures.

Associations with Observer Reports: The correlation between total pain behaviors observed in the videotaped pain behavior protocol and PaB-SR scores was 0.57. This compares to a median correlation of 0.33 for item level correlations. The median correlation between PaB-SR and UAB scores completed by SOs, was 0.56.

PaB-SR Score Associations with Self-Report Pain Measures: The correlations between PaB-SR scores and PROMIS Pain Interference scores and with Self-Reported Average Pain were high—0.77 and 0.60, respectively. The correlation with Pain Behaviors Checklist scores also was high (0.75).

Comparing PaB-SR Scores across Diagnostic Categories: Twenty items were measurement invariant with respect to diagnostic categories. Based on these 20-items and using the correspondence table (Table 4), mean scores were calculated and compared by clinical condition in an ANOVA. Mean PaB-SR, PROMIS Pain Interference, and Average Pain intensity scores are reported by clinical condition in Table 5. PaB-SR scores were quite similar across groups. Recall that the PaB-SR scores are scored on a T-score metric in which SD=10. In SD units, the differences in mean scores by clinical condition ranged from 0.06 to 0.28 SD units. Though the differences in groups means are relatively small, the null hypothesis of no diagnostic group differences in means was rejected ($p<0.0001$). Post-hoc analyses showed that it was the differences between the scores for back pain and the scores for the other two categories that drove this result. There was no significant difference between MS and arthritis scores. Means of other pain reports by diagnostic group also were quite similar. Across all pain measures, those with back pain reported the highest scores.

DISCUSSION

To our knowledge, the work reported here is the most validation intensive study ever to develop a measure of self-reported pain behaviors. In addition to methodically tracking, evaluating, and re-evaluating the content validity of the items, PaB-SR scores were compared against a number of additional validity standards. This is consistent with current validity theory that posits instrument validation as an integrative and disciplined activity akin to the calling of multiple “witnesses” in a court of law. [28; 29] In the current work those validation witnesses included a strictly controlled pain behavior observation protocol,

comparisons to scores on other pain measures including pain intensity and interference, and observer reports from SOs who spent sufficient time with their pain participants to observe a range of pain behaviors. The results from these multiple perspectives supported the validity of PaB-SR scores. In addition, they indicate the disabling impact of pain and the association between pain-related disability and pain behavior.

We found that scores from self-reports, observer-reports, and a video-taped protocol, while significantly related, differed from each other. These differences may be due to method variance including differences in the focus and perspectives of the three measures. Though we did not expect perfect association among scores, we did expect substantial associations and this is what we found. This finding supported the construct validity of PaB-SR scores. Had there been low or no association between PaB-SR scores and scores on the other measures, we would have had reason to doubt that we were measuring the intended construct.

A particular strength of this study is the linking of PaB-SR scores to the PROMIS metric (mean =50; SD =10). Users of the PaB-SR not only can report differences in raw scores among groups, but can interpret pain behavior levels relative to a sample that is representative of the US population with respect to age, gender, and race/ethnicity. Some caution is warranted in interpreting scores relative to this metric, however, since there is some error inherent in linking scores. There is evidence, however, that comparisons of means of linked scores obtained from samples ≥150 are robust. [17]

As well as supporting the validity of PaB-SR scores, the findings from the current study are of substantive interest. SO reports of their pain participant pain behaviors were not, strictly speaking, proxy reports. However, the finding that, on average, SOs reported more pain behaviors than did their pain participants is consistent with other studies in which proxy quality of life ratings are lower than self-ratings. [2; 10; 25; 26] This is consistent with other studies as well. SO reports provide an alternative perspective on how individuals communicate their experiences of pain, and it is valuable to have their pain behavior reports, especially given the potential role that pain behaviors may play in relationships. Because we collected responses to analogous items of each of the PaB-SR items, in future work we plan to develop an SO version of the PaB-SR (PaB-SO). Scores on the PaB-SO could be used to indicate the differences between the perspectives of pain participants and those of their significant others. Alternatively, when the interest is in use of PaB-SR scores as a proxy measure, an alternative metric could be calibrated that allows one to “correct” for differences in SO and pain participant reports.

Another finding was that our low back sample reported more pain than either the sample of persons with arthritis or with MS. This is consistent with recent work by Davis and colleagues in estimating the incidence and impact of pain in comorbid conditions. [9] This study was a secondary analysis of 1,211,483 patients with one or more of 31 targeted pain conditions including low back pain, rheumatoid arthritis, and MS-associated pain. The most prevalent comorbid condition (defined as having a second of the 31 conditions) was low back pain (35%) followed by osteoarthritis pain (30%).

Our analyses indicated that the PaB-SR items were “unidimensional enough” for IRT calibration. We note, however, that this finding does not preclude the possibility or the usefulness of creating subscales from these items. In some contexts, it may be useful, for example, to distinguish guarding behaviors from other kinds of pain behaviors. Future work should evaluate the properties of the PaB-SR items disaggregated by categories of pain behaviors and scored as subscales.

Limitations

Though the sample sizes were large for this study, diagnoses were limited to individuals with MS, back pain, and arthritis. Future research should evaluate the function and validity of PaB-SR scores in other pain samples. Also, the PaB-SR was developed in a cross-sectional sample. Future work should evaluate the responsiveness of scores to intervention. A third limitation is that our sample was largely non-Hispanic white and well educated. The PaB-SR should be administered in a more heterogeneous sample to gauge the generalizability of our results to additional ethnic groups of individuals with pain. In addition, future research should evaluate the accuracy of the link between PaB-SR scores and the PROMIS metric in samples of different sizes and diagnostic categories. In addition, though our analyses focused on clinical condition as the most salient patient characteristic, future research should evaluate the results by other important pain discriminators such as pain site and number of pain sites.

An additional limitation is the use of a single coding system for behavior sampling. We used a method developed by Keefe and Block, [12] but there are a number of other published coding systems, including ones that focus specifically on facial expression. Notably, though the Keefe and Block includes defines the facial category, “grimacing” as including “an obvious facial expression of pain which includes furrowed brow, narrowed eyes, tightened lips, corners of mouth pulled back, and clenched teeth”, Keefe and Block [12] (p. 366), the inclusion of “clenched teeth” has been challenged both on empirical and clinical grounds. [15; 24]

Different sources of data about pain behavior (i.e., patients, spouse observers, and standardized behavioral observation) have inherent strengths and weaknesses. For example, patients have the most access to all of their (waking) behavior, and so may have the most opportunity to provide valid ratings of their behavior overall. However, patient characteristics (e.g., to be stoicism, tendency to express feelings more readily) unintentionally may bias their reporting. When observing pain behavior, partners and significant others may be less biased (at least by patient goals), and have many opportunities to observe patient pain behaviors than clinicians, yet their own characteristics (e.g. thoughts and feelings about the patient’s pain) could impact their ratings of pain behavior. Trained observers using behavior sampling procedures are likely to be the least biased, yield a sample of pain behavior during tasks that standardize the physical demands on patients, but will often only be able to provide a limited snapshot of a patient’s behavior. Given the strengths and weaknesses inherent in each data source, it is not likely that anyone can be viewed as the most useful and valid in all situations. Instead, the goal should be to develop the most useful measure(s) from each data source (patient, significant other, standardized procedure), and then select the measures that are most appropriate for addressing a given study or clinical question.

The current study provides support for what we believe may be a viable option for assessing pain behavior when a patient-reported measure is deemed appropriate. Finally, the item pool in this study was selected from a combination of items from existing measures and expert judgment. We then determined the content validity of the final items in the PaB-SR based on the judgment of the study investigators, all of whom have expertise in the development and evaluation of measures, and two of whom (FK and MPJ) have been working in the field as clinicians for decades. Using expert judgments to select items and determine content validity is common. [1] However, ethological descriptions of the pain behaviors most often observed and viewed as most important by patients and their significant others are lacking in this field. It would be useful to evaluate the content validity of the PaB-SR and other pain behavior measures using this approach (e.g., (Lin, 2011 #48)).

As described above, we evaluated whether we could scale a single, over-arching construct of pain behaviors using self-report items. The result of these analyses is the PaB-SR. However, we hasten to add that our results do not preclude scaling subdomains of pain behavior (e.g., rubbing, guarding). Though we think having a general measure of pain behaviors is valuable, we recognize the need for measures of pain behavior subdomains. Future research plans include the development of such scales.

Conclusions

In this study we developed and validated a 20-item instrument for measuring pain behaviors, the PaB-SR. This measure was developed using IRT and has excellent psychometric properties. The work described here both supports the validity of the PaB-SR scale and illuminates the relationships among different approaches to quantifying pain behavior. The PaB-SR and its future, analogue scales for SOs (PaB-SO) as well as measures of pain subdomains such as guarding could be of substantial utility to researchers wishing to explore these relationships in greater detail.

Acknowledgments

The project described was supported by Award Number RC1NR011804 from the National Institute of Nursing Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Nursing Research or the National Institutes of Health. The authors have no conflicts of interest to report related to this research study.

REFERENCE LIST

1. American Educational Research Association APA, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.
2. Beaupre P, Keefe FJ, Lester N, Affleck G, Frederickson B, Caldwell DS. A computer-assisted observational method for assessing spouses' ratings of osteoarthritis patients' pain. Psychol Health Med. 1997; 2(2):101–110.
3. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. Qual Life Res. 2007; 16 (Suppl 1):95–108. [PubMed: 17530450]
4. Cai, L.; Thissen, D.; du Toit, S. IRTPRO. Skokie, IL: Scientific Software International, Inc; 2011.
5. Cella D, Chang CH. A discussion of item response theory and its applications in health status assessment. Med Care. 2000; 38(9 Suppl):II66–72. [PubMed: 10982091]
6. Chen W-H, Thissen D. Local dependence indices for item pairs using item response theory. Journal of Educational and Behavioral Statistics. 1997; 22:265–289.
7. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. J Stat Softw. 2011; 39(8):1–30. [PubMed: 21572908]
8. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore. 1994; 23(2):129–138. [PubMed: 8080219]
9. Davis JA, Robinson RL, Le TK, Xie J. Incidence and impact of pain conditions and comorbid illnesses. J Pain Res. 2011; 4:331–345. [PubMed: 22090802]
10. Fleming A, Cook KF, Nelson ND, Lai EC. Proxy reports in Parkinson's disease: caregiver and patient self-reports of quality of life and physical activity. Mov Disord. 2005; 20(11):1462–1468. [PubMed: 16028212]
11. Fordyce WE, Brockway JA, Bergman JA, Spengler D. Acute back pain: a control-group comparison of behavioral vs traditional management methods. J Behav Med. 1986; 9(2):127–140. [PubMed: 2940370]
12. Keefe FJ, Block AR. Development of an observation method for assessing pain behavior in chronic low back pain patients. Behav Ther. 1982; 13:363–375.

13. Keefe FJ, Caldwell DS, Queen KT, Gil KM, Martinez S, Crisson JE, Ogden W, Nunley J. Pain coping strategies in osteoarthritis patients. J Consult Clin Psychol. 1987; 55(2):208–212. [PubMed: 3571674]
14. Kerns RD, Haythornthwaite J, Rosenberg R, Southwick S, Giller EL, Jacob MC. The Pain Behavior Check List (PBCL): factor structure and psychometric properties. J Behav Med. 1991; 14(2):155–167. [PubMed: 1880794]
15. McCrystal KN, Craig KD, Versloot J, Fashler SR, Jones DN. Perceiving pain in others: validation of a dual processing model. Pain. 2011; 152(5):1083–1089. [PubMed: 21388739]
16. McDaniel LK, Anderson KO, Bradley LA, Young LD, Turner RA, Agudelo CA, Keefe FJ. Development of an observation method for assessing pain behavior in rheumatoid arthritis patients. Pain. 1986; 24(2):165–184. [PubMed: 3960569]
17. Noonan VK, Cook KF, Bamer AM, Choi SW, Kim J, Amtmann D. Measuring fatigue in persons with multiple sclerosis: creating a crosswalk between the Modified Fatigue Impact Scale and the PROMIS Fatigue Short Form. Qual Life Res. 2011
18. Prkachin KM, Schultz IZ, Hughes E. Pain behavior and the development of pain-related disability: the importance of guarding. Clin J Pain. 2007; 23(3):270–277. [PubMed: 17314588]
19. PROMIS NIH. [accessed 5/1/2003] Domain Framework. 2012. available at http://www.nihpromis.org/measures/domainframework
20. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. J Pers Assess. 2010; 92(6):544–559. [PubMed: 20954056]
21. Revicki DA, Chen WH, Harnam N, Cook KF, Amtmann D, Callahan LF, Jensen MP, Keefe FJ. Development and psychometric analysis of the PROMIS pain behavior item bank. Pain. 2009; 146(1–2):158–169. [PubMed: 19683873]
22. Richards JS, Nepomuceno C, Riles M, Suer Z. Assessing pain behavior: the UAB Pain Behavior Scale. Pain. 1982; 14(4):393–398. [PubMed: 7162841]
23. Samejima, F. Psychometric Monograph No. 17. Richmond, VA: Psychometric Society; 1969. Estimation of Latent Ability Using a Response Pattern of Graded Scores.
24. Sheu E, Versloot J, Nader R, Kerr D, Craig KD. Pain in the elderly: validity of facial expression components of observational measures. Clin J Pain. 2011; 27(7):593–601. [PubMed: 21415714]
25. Sneeuw KC, Aaronson NK, Osoba D, Muller MJ, Hsu MA, Yung WK, Brada M, Newlands ES. The use of significant others as proxy raters of the quality of life of patients with brain cancer. Med Care. 1997; 35(5):490–506. [PubMed: 9140337]
26. Steel JL, Geller DA, Carr BI. Proxy ratings of health related quality of life in patients with hepatocellular carcinoma. Qual Life Res. 2005; 14(4):1025–1033. [PubMed: 16041898]
27. Williams, F. Reasoning With Statistics. New York: Holt, Rinehart and Winston; 1968.
28. Zumbo, B. Validity as contextualized and pragmatic explanation, and its implications for validation practice. In: Lissitz, R., editor. The Concept of Validity: Revisions, New Directions and Applications. Charlotte, NC: Information Age Publishing, Inc; 2009.
29. Zumbo, BD. Validity: Foundational Issues and Statistical Methodology. In: Rao, CR., editor. Handbook of Statistics: Psychometrics. Elsevier; 2007. p. 45-80.

Summary Statement

A new measure, the Pain Behaviors Self Report (PaB-SR), provides a valid tool with which to explore relationships between pain behaviors and pain correlates.

Table 1

Study Sample Descriptions

	Sample A: Cross-sectional Survey	Sample B: Score Linking Survey	Sample C: Pain observations Study
Recruitment	Participants from previous UW or UNC studies, patients at UW Pain Clinic, flyers at other Seattle clinics, and web posting on UW website	Random subset of PwP who participated in either Sample A or Sample C and indicated interest in participating in future research studies	Participants from Sample A, clinician referrals, advertisements, flyers, web postings
Inclusion/Exclusion Criteria			
PwP	Self-report of at least one of 3 target clinical conditions; ≥ 18 years of age; read/understand English; SO who spends ≥ 7 hours/week with them and willing to participate; average pain in past week ≥ 3 (0–10 scale), persisting ≥ 6 months; and able to walk (with or without assistive device). Individuals who used wheelchairs most or all of the time were excluded because use of a wheelchair could impact how and what pain behaviors are displayed.	Same as for Sample A except no SO reporting was required.	Same as for Sample A except that self-reported average 7 day pain score required to be ≥ 3 (0–10 scale), and pain reported to affect walking, sitting or reclining. Participants were not required to have a SO.
SOs	Nominated by PwP and willing to participate.	n/a	n/a
Number of PwP and SOs by Clnical Condition			
Arthritis	N= 226 PwP, N=206 SOs	N=161 PWP	N=26 PwP
Back Pain	N=225 PwP, N=209 SOs	N=154 PWP	N=30 PwP
Multiple Sclerosis	N=210 PwP, N=203 SOs	N=158 PWP	N=30 PwP
Data collected			
PwP	Demographics, clinical status, candidate PaB-SR items, other pain measures	Subset of PaB-SR items and PROMIS items that had the same item content but different response scales.	Demographics, clinical status, current pain (0–10), candidate PaB-SR items, other pain measures
SOs	Demographics, parallel items to candidate PaB-SR items, UAB Pain Behavior Scale	n/a	n/a
How data were used	Item-level validity, score level validity	Align PaB-SR scores to PROMIS metric	Item-level validity, score level validity

UW = University of Washington, Seattle; UNC = University of North Carolina, Chapel Hill; PwP = participants with pain; SOs = significant others; PaB-SR = Pain Behaviors Self Report

Pain. Author manuscript; available in PMC 2014 December 01.

Table 2

Characteristics of Participants in the Administration of Candidate Items for the Pain Behaviors Self Report (PaB-SR) scale

	Cross-Sectional Study Pain Participants n=618		Cross-Sectional Study Significant Others n=618		Observation Study Participants n=86	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age (years)	57.4	14.1	54.8	15.7	57.1	14.6
Years pain has been a problem	11.9	10.2			11.7	9.8
Hours spent with pain participant per week			51.2	31.8		
Average pain intensity at screening (0–10)	6.0	2.0			6.0	1.9
	N	**Percent**	***N***	**Percent**	***N***	**Percent**
Gender						
Female	467	75.6%	274	44.3%	59	68.6%
Male	151	24.4%	344	55.7%	27	31.4%
Race/Ethnicity						
Non-Hispanic White	520	84.1%	494	80.2%	69	80.2%
Non-Hispanic Black	55	8.9%	63	10.2%	7	8.1%
Hispanic	10	1.6%	23	3.7%	3	3.5%
More than 1 race	26	4.2%	17	2.8%	7	8.1%
Other	7	1.1%	19	3.1%	0	0.0%
Education						
Some High School	34	5.5%	43	7.0%	2	2.3%
High School Grad/GED	105	17.0%	119	19.3%	8	9.3%
Some College/Technical Degree/AA	255	41.3%	222	35.9%	39	45.3%
College Degree	126	20.4%	140	22.7%	22	25.6%
Advanced Degree	98	15.9%	94	15.2%	15	17.4%
Employment Status[1]						
Homemaker	99	16.0%	37	6.0%	11	12.8%
Unemployed	59	9.5%	49	8.0%	9	10.5%
Retired	207	33.5%	193	31.4%	26	30.2%
On disability	194	31.4%	55	8.9%	26	30.2%
On leave of absence	9	1.5%	2	0.3%	0	0.0%

	Cross-Sectional Study Pain Participants n=618		Cross-Sectional Study Significant Others n=618		Observation Study Participants n=86	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Full-time employed	97	15.7%	247	40.2%	19	22.1%
Part-time employed	68	11.0%	78	12.7%	9	10.5%
Full-time student	8	1.3%	23	3.7%	2	2.3%
Relationship with Pain Participant						
Spouse			363	58.7%		
Family member			117	18.9%		
Partner			41	6.6%		
Friend			52	8.4%		
Co-worker			5	0.8%		
Caregiver			5	0.8%		
Roommate			3	0.5%		
Other			6	1.0%		
More than 1 relationship listed			26	4.2%		
Mode of Administration						
Paper survey	347	56.1%	317	51.3%	86	100.0%
PROMIS Assessment Center	269	43.5%	301	48.7%	0	0.0%
Phone survey	2	0.3%	0	0.0%	0	0.0%

[1] Percentages total more than 100% since participants may choose more than one response

Table 3

Associations and Comparisons among PaB-SR Item and Full Scale Scores and Related Measures (Sample size for all comparisons is 618 pain participants and/or SOs except for PBNEW19 (N=268). For this item only, participants were allowed to answer “not applicable”)

Item Name	Item Content	Spearman Correlations Coefficients						
		# Observed Pain Behaviors	PROMIS Pain Interference	Average pain (0–10)	PBCL	UAB	Pain Pt/SO	Pain Pt minus SO item responses(mean ± SD)
PAINBE44	I bit or pursed my lips.	0.46	0.42	0.33	0.48	0.30	0.27	0.09 ± 1.35
PBNEW2	It showed on my face.	0.43	0.58	0.45	0.59	0.45	0.39	−0.49 ± 1.16
PBNEW24	I asked for someone to help me.	0.41	0.47	0.39	0.48	0.42	0.38	−0.09 ± 1.13
PAINBE3	I grimaced.	0.40	0.47	0.36	0.54	0.34	0.31	−0.13 ± 1.25
PAINBE51	I avoided physical contact with others.	0.37	0.61	0.41	0.53	0.35	0.41	−0.05 ± 1.33
PAINBE35	I groaned.	0.36	0.46	0.44	0.59	0.47	0.49	−0.25 ± 1.17
PBNEW21	I told people I couldn’t do things with them.	0.35	0.69	0.46	0.6	0.42	0.47	−0.15 ± 1.19
PBNEW18	I told people I couldn’t do my usual chores.	0.33	0.63	0.47	0.58	0.47	0.46	−0.36 ± 1.20
PAINBE38	I drew my knees up.	0.32	0.29	0.26	0.35	0.22	0.34	0.15 ± 1.19
PBNEW31	I stayed very still.	0.31	0.47	0.32	0.48	0.3	0.31	−0.11 ± 1.21
PBNEW4	My muscles tensed up.	0.31	0.47	0.38	0.45	0.28	0.38	0.23 ± 1.23
PAINBE4	I took medication for the pain.	0.23	0.42	0.38	0.39	0.33	0.55	−0.11 ± 1.07
PBNEW12	I used pillows or other objects to get more comfortable.	0.23	0.45	0.31	0.37	0.3	0.45	−0.04 ± 1.27
PBNEW13	I changed how I breathe.	0.23	0.49	0.4	0.49	0.32	0.42	0.12 ± 1.20
PAINBE28	I squirmed.	0.16	0.43	0.38	0.44	0.3	0.34	0.15 ± 1.25
PBNEW32	I lay down.	0.15	0.49	0.36	0.44	0.37	0.52	−0.16 ± 1.08
PBNEW14	I took breaks.	0.14	0.53	0.41	0.44	0.39	0.41	0.07 ± 1.13
PAINBE7	I rubbed the site of the pain.	0.12	0.36	0.35	0.41	0.31	0.4	0.20 ± 1.21
PAINBE5	I talked about the pain.	0.11	0.36	0.36	0.4	0.34	0.39	−0.62 ± 1.14
PBNEW7	I changed my posture.	0.04	0.46	0.33	0.46	0.34	0.33	0.23 ± 1.17
PAINBE29	I used a cane or something else for support.	0.61	0.39	0.28	0.51	0.47	0.75	0.00 ± 0.95
PBNEW16	I asked for help when walking.	0.53	0.41	0.33	0.5	0.38	0.47	−0.22 ± 0.97
PBNEW9	You could hear it in my voice.	0.50	0.6	0.48	0.62	0.45	0.41	−0.46 ± 1.24

Pain. Author manuscript; available in PMC 2014 December 01.

Item Name	Item Content	Spearman Correlations Coefficients						
		# Observed Pain Behaviors	PROMIS Pain Interference	Average pain (0–10)	PBCL	UAB	Pain Pt/SO	Pain Pt minus SO item responses(mean ± SD)
PAINBE17	I gasped.	0.48	0.47	0.44	0.53	0.43	0.37	−0.07 ± 1.23
PBNEW20	I told people I couldn't do the things that I usually enjoy doing.	0.46	0.67	0.47	0.59	0.44	0.44	−0.25 ± 1.25
PBNEW23	I asked people to leave me alone.	0.46	0.5	0.38	0.51	0.28	0.41	−0.11 ± 1.14
PBNEW27	I became withdrawn.	0.46	0.62	0.34	0.58	0.33	0.41	0.01 ± 1.26
PBNEW25	I limped.	0.43	0.39	0.33	0.5	0.38	0.52	0.20 ± 1.28
PAINBE37	I isolated myself from others.	0.41	0.58	0.36	0.54	0.3	0.41	0.03 ± 1.27
PBNEW6	I walked slowly	0.40	0.58	0.4	0.62	0.44	0.5	−0.12 ± 1.15
PAINBE43	I walked carefully.	0.36	0.49	0.34	0.53	0.37	0.44	−0.05 ± 1.21
PBNEW26	I became quiet.	0.36	0.49	0.31	0.43	0.24	0.33	0.12 ± 1.22
PBNEW3	I moaned.	0.34	0.45	0.44	0.57	0.45	0.48	−0.26 ± 1.17
PAINBE18	I asked for help doing things that needed to be done.	0.33	0.52	0.42	0.47	0.42	0.41	−0.22 ± 1.17
PBNEW29	I moved my limbs carefully.	0.33	0.52	0.4	0.49	0.33	0.33	−0.07 ± 1.21
PBNEW30	I sat down.	0.33	0.47	0.35	0.47	0.31	0.37	−0.06 ± 1.10
PBNEW5	I changed the way I walked.	0.33	0.54	0.42	0.58	0.44	0.42	0.08 ± 1.21
PBNEW19	I told people I couldn't go to work or school.	0.32	0.49	0.35	0.43	0.38	0.56	−0.33 ± 0.93
PBNEW17	I asked for help with changing positions.	0.29	0.4	0.41	0.46	0.36	0.4	−0.20 ± 0.98
PBNEW28	I was careful about the way I moved from one position to another.	0.27	0.53	0.41	0.51	0.38	0.32	0.00 ± 1.22
PAINBE24	I moved stiffly.	0.26	0.51	0.39	0.58	0.39	0.38	0.08 ± 1.28
PBNEW10	I reclined.	0.20	0.44	0.32	0.4	0.29	0.38	−0.06 ± 1.17
PBNEW22	It caused me to lean or bend while walking.	0.18	0.48	0.38	0.61	0.46	0.48	0.35 ± 1.30
PBNEW1	I touched or held the part of my body that hurt.	0.16	0.34	0.3	0.39	0.28	0.38	0.15 ± 1.12
PBNEW8	I shifted my position.	0.16	0.45	0.3	0.38	0.29	0.27	0.26 ± 1.20
PBNEW11	I took a hot bath or shower.	−0.04	0.18	0.2	0.19	0.16	0.53	0.06 ± 1.13
PaB-SR Score (based on final items)		**0.57**	**0.77**	**0.6**	**0.75**	**0.56**	**n/a**	

Pt = Pain Participant

SO = Significant Other

PBCL = Pain Behavior Check List (pain participant report)

UAB = University of Alabama and Birmingham Pain Behavior Scale

SD = Standard Deviation

Table 4

Cross-walk between Summed Scores (sum of all item scores) and PaB-SR Scores

Summed Score	PaB-SR Score
1	38.0
2	40.6
3	42.4
4	43.8
5	45.0
6	46.0
7	46.9
8	47.7
9	48.4
10	49.0
11	49.6
12	50.2
13	50.7
14	51.2
15	51.7
16	52.1
17	52.6
18	53.0
19	53.4
20	53.8
21	54.2
22	54.6
23	55.0
24	55.3
25	55.7
26	56.0
27	56.4
28	56.7
29	57.1
30	57.4
31	57.8
32	58.1
33	58.5
34	58.8
35	59.1
36	59.5
37	59.8

Summed Score	PaB-SR Score
38	60.1
39	60.5
40	60.8
41	61.1
42	61.5
43	61.8
44	62.2
45	62.5
46	62.9
47	63.2
48	63.5
49	63.9
50	64.3
51	64.6
52	65.0
53	65.3
54	65.7
55	66.0
56	66.4
57	66.8
58	67.2
59	67.5
60	67.9
61	68.3
62	68.7
63	69.1
64	69.6
65	70.0
66	70.5
67	70.9
68	71.4
69	71.9
70	72.5
71	73.0
72	73.7
73	74.3
74	75.1
75	76.0
76	76.9

Summed Score	PaB-SR Score
77	78.1
78	79.5
79	81.2
80	83.7

Table 5

Mean and Standard Deviations (SD) of Pain Reports by Clinical Condition

	Multiple Sclerosis	Arthritis	Back Pain	F	p
Pain Behavior Self-Report (PaB-SR)	58.4 (4.6)	59.0 (6.1)	61.2 (5.4)	15.6	<0.0001
NIH-PROMIS Pain Interference (T-score)	61.4 (6.6)	60.5 (7.2)	64.6 (6.4)	22.0	<0.0001
Pain on average in last 7 days (0–10)	5.0 (1.8)	5.2 (2.4)	5.9 (2.0)	12.4	<0.0001

NIH-PROMIS: National Institutes of Health Patient Reported Outcome Measurement Information System