PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE

FINAL RESEARCH REPORT

Developing Measures of Pain Appraisal and Pain-Related Self-Efficacy for People Living with Chronic Pain

Dagmar Amtmann, PhD; Mark Jensen, PhD; Dennis Turk, PhD; Danielle Lavallee, PhD, PhR, MD; Kendra Liljenquist, PhD; Alyssa Bamer, MPH; Rana Salem, MA

AFFILIATION:

University of Washington, Seattle

Original Project Title: Extending PROMIS Pain Item Banks: Pain Self-Efficacy and Pain Catastrophizing
PCORI ID: ME-1403-12550
HSRProj ID: 20152315

To cite this document, please use: Amtmann D, Jensen M, Turk D, et al. (2019). *Developing Measures of Pain Appraisal and Pain-Related Self-Efficacy for People Living with Chronic Pain.* Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/7.2019.ME.140312550

TABLE OF CONTENTS

ABSTRACT

Background: The Patient-Reported Outcomes Measurement Information System (PROMIS®) is a system of reliable and flexible measures of patient-reported health status that include several instruments for measuring different aspects of pain. However, 2 pain-related constructs are missing: pain catastrophizing (PC) and pain-related self-efficacy (PRSE). Pain catastrophizing is a form of ‘‘cognitive distortion” that can lead to negative affect and can amplify symptoms and distress. PRSE refers to a person’s confidence in living well with chronic pain and successfully managing the impact of pain on their lives. Understanding and measuring these aspects of chronic pain is important for designing treatments that can improve quality of life.

Objectives: The purpose of this study was to develop brief, flexible, psychometrically sound, patient-centered, and clinically meaningful measures of PC and PRSE.

Methods: Development of both item banks followed the PROMIS methodology. Patient advisors worked with investigators on every aspect of the study. A panel of pain researchers defined the PC construct. Focus groups and cognitive interviews with people living with chronic pain reviewed the definitions and provided feedback on meaningfulness and clarity of all items. Individuals with chronic pain recruited from various sources (online recruitment companies, previous studies, research registry, pain clinic) completed the candidate items via online or paper surveys. Pain Catastrophizing Scale and Pain Self-Efficacy Questionnaire were used as legacy measures (ie, existing gold standard self-report measures of PC and PRSE). Test–retest stability data were collected from a subset of respondents 40 to 80 hours after initial administration of the candidate items. Reliability was evaluated using intraclass correlation (ICC), and items were examined for unidimensionality, local dependence (redundancy), and differential item functioning (DIF). Item response theory (IRT) was used to calibrate the items.

Results: Final banks included 24 PC and 29 PRSE items calibrated on a demographically (eg, gender, age, race) and clinically (ie, various chronic pain conditions) heterogeneous sample of people with chronic pain (n = 795). The PC instrument was renamed “Pain Appraisal Scale” (PAS) after feedback from patient groups indicated “catastrophizing” was perceived as stigmatizing. Items for short forms were selected by considering item parameters and content coverage. Six items that captured the trait continuum were selected for each short form. Correlations between the 6-item short forms and the full item bank scores (PAS $r = 0.99$; PRSE $r = 0.85$) and test–retest reliability (PAS ICC = 0.93; PRSE ICC = 0.90) were excellent. Correlations of short form scores with legacy measures of related constructs were high (>0.8).

Conclusions: Six-item short forms measure well across the PC and PRSE continuums and the scores are highly correlated with the full item bank scores. The results support the validity and reliability of the PAS and PRSE short forms. Short form scores are on the same metric and directly comparable to full item bank scores, are brief, and are well suited for research and clinical practice. The short form and the full item bank will be available publicly and free of charge at uwcorr.washington.edu.

Limitations and Subpopulation Considerations: If people with additional chronic pain conditions were included, they may have provided different perspectives on PRSE and PC. Analyses suggest scores are primarily driven by the level of PRSE and PC, rather than demographic characteristics (eg, age, gender).

BACKGROUND

Chronic pain puts significant burdens on the individual experiencing pain, their significant others, the health care system, and the nation's economy. Chronic pain affects the lives of approximately 100 million adults, exacting a tremendous toll in rehabilitation resources, health care costs, and lost worker productivity.[1] Initially, pain signals the need to heal and protect tissue damage, but when it persists beyond recovery and recuperation, it can become maladaptive.[2] The consequences of chronic pain are extensive and are known to affect psychological, physical, and social well-being.[3,4] As a result, chronic pain can lead to short- and long-term disability and unemployment, with substantial costs to society and employers.[5,6] The negative economic impact of chronic pain has been estimated at $560 billion to $635 billion annually in the United States.[7]

Over the past 2 decades, there have been considerable shifts in how pain is conceptualized. In addition to neurobiological factors, evidence for psychosocial determinants of pain, disability, and treatment response have been accumulating. In a large prospective study, Jarvik et al[8] found that psychosocial factors (eg, catastrophizing) were more predictive of onset of back pain and disability than baseline physical or biomedical measures. Similarly, Carragee et al[9] observed that psychosocial factors (eg, depression) were more predictive of persistent back pain and disability than objective pathological measures of diagnostic imaging. Among others, Burns et al[10] found that cognitive factors and mainly changes in perceived self-efficacy were the best predictors of response to pain rehabilitation treatment. Consequently, the biomedical model has proved inadequate for explaining the impact and experience of pain, and psychosocial factors need to be considered in treatment plans aimed at improving pain outcomes.[11]

Chronic pain is associated with various prevalent conditions (eg, arthritis) and disorders (eg, low back pain). To treat chronic pain, appropriate ways to assess patient experiences and impacts of pain are needed. A crucial finding of the Institute of Medicine's 2011 report *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research* is the need for better pain assessment.[7] A substantial body of evidence supports the

efficacy of self-management interventions that address self-efficacy and catastrophizing in improving the mental health, pain, and health-related quality of life of people with chronic pain.[11,12] Understanding and measuring the belief that people with chronic pain can successfully self-manage and live effectively with pain is important for designing treatment regimens that can lead to better quality of life. Critical to this is the need for improved pain assessment,[7] including availability of brief, precise, and patient-centered outcome measures that assess important psychosocial determinants such as pain catastrophizing and pain-related self-efficacy.

Pain-related self-efficacy (PRSE) beliefs have been used to describe a range of behaviors and aspects of pain experience and treatment response. PRSE beliefs for people with chronic pain incorporate not only the expectation that they can perform a specific activity but also their confidence in their ability to perform the activity despite pain. Thus, higher patient confidence in the ability to tolerate pain can predict actual tolerance and psychosocial factors and has been shown to be more predictive of back pain and disability than physical tests or diagnostic imaging.[8,9,13,14]

Pain catastrophizing (PC) refers to the exaggerated negative thoughts and feelings about pain that some people have when they are in pain (eg, “Because of my pain, my life is over”). PC is a form of cognitive distortion that is hypothesized to contribute to the development and maintenance of depression and negative affect. Chronic pain and depression are commonly overlapping syndromes—30% to 60% of patients with chronic pain present significant depressive symptoms.[15] The catastrophizing that accompanies depression can negatively influence patient perceptions, memories, expectations, and, consequently, their experiences.[16,17] As a further consequence, they may develop passive coping styles (eg, lack of control, helplessness, rumination) that further exacerbate their plight.[18] In a study of 2618 adults with chronic pain, catastrophic thinking and depression were significantly correlated with pain-related disability.[19] In fact, catastrophizing has emerged among the most important and consistent psychosocial predictors of nearly every key pain-related outcome, such as pain intensity, psychological functioning, and disability.[20-29]

Three scales to measure PRSE[30-32] and 2 versions of a PC scale[33,34] are currently available. All of them are several decades old and have been developed using classical test theory, rather than item response theory (IRT). One of the strengths of IRT lies in its ability to address important psychometric and substantive issues that are difficult to evaluate in the classical test theory framework (eg, reliability at every level of the trait).[35] The Pain Self-Efficacy Questionnaire (PSEQ)[30] is a 22-item instrument that incorporates the role of pain in rating self-efficacy beliefs.[36] Although people with chronic pain were asked to review items in an interview setting, no focus groups were conducted with people with chronic pain to ensure that all important facets of catastrophizing are included, and items were not analyzed using IRT. The Chronic Pain Self-Efficacy Scale[31] is a 22-item questionnaire designed to measure chronic pain patients' perceived self-efficacy to cope with the consequences of chronic pain. Finally, the Arthritis Self-Efficacy Scale[32] provides a classical test theory–based summary score and the subscale scores. The Coping Strategy Questionnaire (CSQ)[33] is a 44-item questionnaire developed for people with rheumatoid arthritis pain that contains a 6-item catastrophizing scale. The Pain Catastrophizing Scale (PCS)—an extension of the CSQ—was developed by Sullivan et al[37] and includes 13 items that can be totaled for an overall score or grouped into 3 subscales: rumination, magnification, and helplessness. The scale's internal consistency and validity have been demonstrated in clinical undergraduate and community-based samples.[34,38]

In the time since these scales were developed, psychometric advances (eg, IRT) have made shorter or dynamically administered instruments possible. In particular, the development of IRT-based item banks has facilitated flexible administration options that use item combinations personalized to the respondent while maintaining comparability of the scores across studies and populations.[39-41] The Patient-Reported Outcomes Measurement Information System (PROMIS®) is a system of precise, reliable, and flexible measures of patient-reported health status developed through the NIH cross-institute roadmap initiative. PROMIS includes several specific pain-related sets of questions (item banks) that measure pain intensity, pain interference, pain quality, and pain behavior.[42] The PROMIS initiative was tasked with developing measures for several health domains (eg, depression, physical function, fatigue, social function) in addition to pain.[42] Therefore, the initial PROMIS initiative could only allocate

enough resources to develop measures for a limited number of pain constructs; PC and PRSE were not part of the PROMIS portfolio of item banks. New and sophisticated patient-centered measures of PRSE and PC developed using IRT-based approaches will facilitate pain research and clinical practice while reducing patient-respondent burden. Adding measures of PC and PRSE would greatly enhance the utility of PROMIS pain instruments for conducting comparative effectiveness research in chronic pain populations. Thus, the item banks we propose to develop to measure PC and PRSE would fill an important void in the currently available PROMIS pain domain instruments. These banks would allow researchers and clinicians to take advantage of applications of IRT and of the scientifically rigorous instrument development process adopted by the PROMIS initiative.

Currently, researchers who are targeting improved understanding of the subjective experience of pain and related disability and who want to use IRT-based PROMIS tools are limited by the absence of PROMIS instruments for measuring PRSE and PC. The specific aims of this study were to (1) develop an item pool for assessing PC and PRSE, (2) conduct IRT analyses to calibrate the item banks, and (3) examine the psychometric properties of the new PC and PRSE scores.

PARTICIPATION OF PATIENTS AND OTHER STAKEHOLDERS IN THE DESIGN AND CONDUCT OF RESEARCH AND DISSEMINATION OF FINDINGS

Types and Number of Stakeholders Involved, and How the Balance of Stakeholder Perspectives Was Conceived and Achieved

The main stakeholders for the measurement instruments developed in this study were (1) people living with chronic pain, because the instruments must be meaningful and understandable by people with different types of chronic pain; and (2) researchers and clinicians who treat and study chronic pain. This project involved a diverse group of stakeholders both as part of the research team and as consultants at key points in the research process. Chronic pain experts participated in every aspect of this study: They participated in expert panels that guided instrument development, provided guidance on definitions of the constructs, provided feedback on the PC and PRSE items, and edited those items as needed.

This study also included 2 patient partners (Penny Cowan, a founder and executive director of the American Chronic Pain Association, and Mary Scott; both persons live with chronic pain) who represented perspectives of patients with chronic pain. In addition, consistent with the rigorous PROMIS methodology, individuals with chronic pain were extensively involved in all stages of instrument development to ensure that the items are meaningful and measure all important facets of the PC and PRSE.

Methods Used to Identify and Recruit Stakeholder Partners

We worked with patient partners through the Comparative Effectiveness Research Translation Network (CERTAIN), a PCORI-funded project called the University of Washington Patient Voices Network. Dr Danielle Lavallee and Sarah Lawrence participated in teleconferences with patient partners and helped the study team create best strategies for meaningful engagement of patient advisors.

Methods, Modes, and Intensity of Engagement

Together with the principal investigator and co-investigators, a panel composed of 5 pain experts and clinicians participated in panels in February, May, and July 2015, to guide study design, help develop definitions of the constructs, and place the constructs in a larger conceptual model of pain. Expert panel members discussed (1) what conceptual frameworks could be used to generate items for both instruments (ie, framework[s] that outline the relevant subdomains of the constructs); (2) the definition of both constructs, including subdomains; (3) how the constructs are related to other variables that could be used for validation purposes; and (4) what potential items should look like (eg, instructions, time frame, item stem, response options). Once the items were developed, the expert panel members reviewed them and provided feedback.

Patient partners also made an important contribution to the quality of the products developed by the study. Ms Cowan and Ms Scott reviewed all recruitment materials and provided feedback on the recruitment flyer and focus group materials, such as questions asked at the focus groups. This resulted in more patient-friendly recruitment and study materials. Patient partners provided feedback during regular teleconferences as well as during in-person meetings and in written form. For instance, Ms Scott provided detailed edits to the recruitment flyer, such as simplifying the content for clarity and editing the sentences to make them more meaningful to people with pain. This resulted in more comprehensive and user-friendly recruitment material and questionnaires.

People living with chronic pain who reviewed our proposal for PCORI and the focus group participants expressed concerns about how clinicians might interpret patient answers to the PC items. We extensively discussed the issue of the potentially stigmatizing effect of the term “pain catastrophizing” and decided come up with a more neutral term, settling on the name “Pain Appraisal Scale” (PAS). Therefore, a brief guidance statement was developed to help clinicians interpret PAS scores to avoid stigmatizing patients with chronic pain. We interviewed people with chronic pain (n = 8) and asked them to review and provide feedback on the final PAS items as well as the guidance statement. Most participants did not have any

concerns about the questions being asked in the PAS, although some also expressed concern about the interpretation of some responses (eg, worried that some items may be interpreted as suicidal ideation). Most participants agreed with the guidance statement and felt it would help in avoiding stigmatization of patients with chronic pain. Several participants provided suggestions for the guidance statement, and we edited it to address concerns raised in interviews (eg, indicate referral to other specialists in addition to mental health professionals; emphasize that responses to the PAS do not indicate that a person is seeking drugs; explain that external factors, both related to a patient's pain and unrelated, could affect PAS responses). This resulted in specific guidance for clinicians on how to interpret the scores and substantially addressed the patient concerns. In addition, modified instructions will help assure patients that their responses will not be used to stigmatize them. In conjunction with the modified instructions, our patient advisor (Penny Cowan) suggested we include a few "positive" items to reduce the impact of negativity. We followed her suggestion and developed a PC short form that includes 2 items (prse23 and prse21) from the PRSE bank (which are positive in nature) at the end of the scale; the 2 items do not contribute to the PAS score.

Perceived or Measured Impact of Engagement

1. **Relevance of the research question.** The expert panel felt strongly that the research questions were highly relevant to both clinical practice and research. Patient partners (and research participants), on the other hand, felt that the construct of PC was too negative and they recommended focusing on more positive constructs, such as PRSE. However, the negative aspect of PC is exactly what makes it a useful therapeutic target. As a result, while we continued to develop the measure of PC, we took several steps to address patients' concerns.

2. **Study design, processes, and outcomes.** Clinicians and researchers who participated in the expert panels guided the study, both to develop a framework that would be useful to further the knowledge of chronic pain and to define constructs and review research methods.

3. **Study rigor and quality.** Stakeholder engagement shaped the project's processes and outcomes in important ways, including changing the name of the measure for PC, developing guidance for clinicians on how to interpret the score, creating items that are

easily understood and meaningful to patients, and developing instructions that provide useful guidance to respondents.

4. **Transparency of the research process.** Involvement of the patient partners also made the processes more transparent because the procedures at different stages had to be described in lay language and materials summarized at each step. While the statistical methods used were too technical for the patient partners to truly understand, the patient partners made important contributions to the final measures. For instance, when statistical analyses indicated the need to remove items similar to other items, the patient partners worked with researchers to select which items to keep and which to delete.

5. **Adoption of research evidence into practice.** Because patient partners viewed PC as such a negative construct, we recommend including the 2 items from the PRSE 2-item short form when administering the PAS 6-item short form.

METHODS

To ensure the new instruments for measuring PC and PRSE maximized their utility, we followed the recommended qualitative and quantitative methodologies set forth by the PROMIS network.[43,44] The University of Washington Human Subjects Division approved all procedures before study implementation.

Study Design

Expert panel members and patient advisors guided construct and item development. A diverse sample of individuals with chronic pain was recruited to participate in focus groups, cognitive interviews, and a prospective large-scale cross-sectional survey. Data from the focus groups were used to verify the definition of the constructs, to identify gaps in subdomains of PRSE and PC, and to examine the language people with chronic pain use to describe the constructs of PRSE and PC. Data provided the information needed to ensure that the instruments used language meaningful to people living with chronic pain. We used data from cognitive interviews to examine whether the items were meaningful and understandable to the target audiences. We used data from the cross-sectional survey to calibrate and validate the new PC and PRSE measures.

STUDY PARTICIPANTS

Individuals in a national research registry who had previously participated in pain-related studies at the University of Washington were invited to participate in focus groups and cognitive interviews. Study participants for the cross-sectional study were recruited from multiple sources: (1) two online panel companies (Toluna and Op4G) that maintain databases of people available to participate in online surveys, (2) participants from previously completed research studies who were interested in future research, (3) the University of Washington's registry of individuals with various chronic health conditions and disabilities who are interested in future research, and (4) the University of Washington Center for Pain Relief.

To be eligible for any part in this study (focus group, cognitive interview, or cross-sectional survey), participants were required to (1) be 18 years of age or older; (2) be able to read, speak, and understand English; (3) have had pain for at least half the days of the 6 months before the study; (4) have a pain intensity level of 3 or higher on the 0 (no pain) to 10 (worst imaginable pain) numerical rating scale; (5) allow the discussion to be audio-recorded (focus groups and cognitive interviews only); and (6) have self-reported low back pain, osteoarthritis of the knee, painful diabetic neuropathy, multiple sclerosis, a spinal cord injury, and/or a lower limb amputation at or below the hip and at or above the ankle. We chose these diagnosis groups because they are commonly associated with pain and were populations that were accessible to the researchers. However, participants could have had multiple sources or types of pain related or unrelated to their diagnosis. The conditions were used only for recruitment and in no context were the participants asked to attribute or rate their pain with respect to their condition. For all study phases, if the research staff believed that a participant's inability to understand questions or follow research procedures would interfere with discussions and question completion, then that participant was excluded from the study. Additionally, those who participated in the focus groups were not eligible to participate in the cognitive interviews.

Measures

All participants responded to a survey that included demographic and clinical information, as well as pain-related and well-being–related measures.

Demographic and Clinical Information

Participants completed a questionnaire to collect demographic (eg, gender, age, ethnicity, race, marital status, income, education, employment status) and clinical information (eg, which conditions cause them pain, which of those causes the most pain).

Validity Measures

We administered the PROMIS 29 profile v.2,[45] Pain Interference short form 6b,[46] and the PROMIS Global[47] to all participants. We scored measures according to published instructions using look-up tables. We did not generate profile and global scores for records with missing items; we imputed the pain interference score for records missing 1 or 2 items and did not score if missing more than 2 items. Subscale scores from the profile used in validity analyses included depression and anxiety scores. All PROMIS scales are scored on a T-score metric with a mean of 50 for the US general population, with higher scores indicating more of the trait being measured. In addition to the PROMIS measures, we asked participants to rate their average pain over the previous 7 days on a scale from 0 to 10. Using this rating, participants were grouped into mild (scores 1-4), moderate (scores 5-6), or severe (scores 7-10) pain categories[48]; participants reporting 0 were ineligible. We used these groups because they are the most commonly recommended cut points for noncancer pain patients reported in the literature.

Legacy Measures

We administered the Pain Catastrophizing Scale developed by Sullivan et al[37] as a legacy measure of PC. The PCS is a 13-item measure with 3 subscales—rumination, helplessness, and magnification—and an overall score. Scale developers did not provide instructions about how to score records with missing data, so we did not generate the overall and subscale scores for records missing any PCS item. An overall PCS score of 30 or greater may represent a clinically significant level of PC.[49] The internal consistency and validity of the scale has been demonstrated in clinical undergraduate and community-based samples and the test–retest correlation was high (r = 0.75) across a 6-week period.[34,38]

The Pain Self-Efficacy Questionnaire[30] is a 10-item instrument that incorporates the role of pain in rating self-efficacy beliefs.[50] The scale developers did not provide instructions about how to score records with missing data, so we did not generate a score for records missing any PSEQ item. Scores of <17 to <20 have indicated levels of low pain self-efficacy in other studies.[51,52] The PSEQ has excellent internal reliability (Cronbach α = .92) and a good test–retest correlation (0.73) over a 3-month period.[30]

Measures of Neuropathic Pain

These measures included the self-report version of the Leeds Assessment of Neuropathic Symptoms and Signs pain scale (S-LANSS)[53] and the PainDETECT Questionnaire[54] (focus groups and cognitive interviews only), and the PROMIS Pain Quality-Neuropathic Pain (PROMIS-PQ-NP)[55] Short Form (all phases). The S-LANSS is a sensitive (74%-78% depending on the cutoff used) measure of neuropathic pain, identifying 75% to 80% of pain types depending on administration method.[53] PainDETECT is a screening measure of neuropathic pain that has high sensitivity (84%-85%), specificity (80%-84%), and positive predictive accuracy (83%-84%).[54] Last, the 5-item PROMIS-PQ-NP measure is a brief and reliable measure that can be used to identify patients more likely to have neuropathic pain and to distinguish between levels of neuropathic pain.[55] The PainDETECT does not provide instructions on how to score records with missing data. Instructions for the S-LANSS indicate that it should not be scored if 1 or more items are missing. Instructions for the PROMIS-PQ-NP state that the look-up scoring table should not be used for records with missing data. Therefore, for each of the 3 neuropathic pain measures, we did not generate a score if a record was missing any items in that measure.

Data Collection

Expert Panel

An expert panel comprised of 5 pain researchers—3 of whom are also clinicians who treat patients with chronic pain and 2 are pain outcome measurement experts—was convened by telephone to help define the constructs of PC and PRSE, including subdomains, and to select

item format (eg, question/statement, response options, instructions, time frame). Each meeting was summarized and a report was compiled with all the key decisions.

Patient Partners

Two patient partners living with chronic pain provided input throughout the development process. Advisors reviewed the summary of the expert panel discussions, including proposed construct definitions and focus group guiding questions; they also helped revise candidate items.

Focus Groups of People Living With Chronic Pain

We conducted focus groups in person, scheduling them to last no more than 2 hours. We planned a minimum of 3 focus groups with 5 to 10 participants each. Focus groups included (1) no more than 2 people with a mild pain score (3 or 4) across all focus groups; (2) at least 2 participants from ethnic minority backgrounds across all focus groups; (3) at least 3 participants each with low back pain, osteoarthritis of the knee, and other conditions (eg, painful diabetic neuropathy, multiple sclerosis, spinal cord injury, amputation) across all focus groups; (4) at least 2 male participants in each focus group; (5) at least 2 people with neuropathic pain in each focus group; and (6) at least 1 person with low back pain and 1 with osteoarthritis of the knee in each focus group. An experienced moderator facilitated the focus groups, which included discussion of both PC and PRSE. All focus groups were audio-recorded and transcribed by a Realtime Captioner.

Cognitive Interviews

Study investigators, with input from the expert panel, focus groups, and patient advisors, developed PRSE and PC items consistent with the proposed definitions. We tested the items in one-on-one cognitive interviews with people living with chronic pain. The research staff conducted the cognitive interviews in person or on the telephone using a semistructured interview format that lasted approximately 1 hour. They asked open-ended questions to guide

the interview (eg, “What were you thinking of when you chose your answer?” or “How did the response options work for you?”). The staff audio-recorded all interviews and took notes.

At least 5 individuals reviewed each item.[44] The sample of cognitive interview participants was required to include at least 2 men, 2 people with painful diabetic neuropathy, 1 person with low back pain, 1 with knee osteoarthritis, no more than 2 people with mild pain (a score of 3 or 4 on the 0-10 numerical rating scale of pain intensity), and at least 1 person from a minority ethnic background. We used participants’ feedback to modify, add, or delete items. We tested substantially revised items in a second round of cognitive interviews with at least 3 participants and eliminated items that did not function well after being revised.

Large-Scale Administration/Field Testing

To ensure adequate representation among this sample of people with chronic pain, we targeted enrollment in the cross-sectional survey to people with different demographic characteristics (eg, younger [<45 years] and older [≥75 years] age, males, less than high school education, and Hispanic and African American race/ethnicity). We needed responses from a minimum of 500 individuals, as this sample size is considered sufficient for reliable estimates of item parameters using a graded-response model with a small to moderate number of items.[56] Panel companies sent study invitations by email, via mobile text, or through an application, with a direct link to the online survey administered through the Research Electronic Data Capture (REDCap)[57] platform.

Individuals recruited from other sources were either sent a link to the online survey or mailed a paper survey with a return envelope, depending on their preference. Individuals recruited from panel companies responded to the eligibility screening questions online before starting the survey and were required to provide their contact information to participate. Study staff contacted all participants who submitted completed survey responses and screened them again via telephone; study staff screened individuals recruited from other sources once via phone before sending the survey or link to the survey. Staff reviewed all survey responses for missing data and consistency; they also double-entered paper surveys to control for data entry

errors. Staff contacted participants to resolve missing data and/or minor inconsistent responses. If participants could not be contacted to resolve missingness, we designated the item(s) as missing and included the record in analyses. In instrument development, missing data are much less of a threat to validity than in other contexts, such as comparisons of group means. The biggest problem with missing data is too few data points for extreme categories. Consistent with best practices, we collected data until we received a minimum of 10 responses in each response category (eg, at least 10 responses each in never, rarely, sometimes, often, always) for each PRSE and PC candidate item.

Reliability Testing

Participants who completed the survey online were invited to retake the PC and PRSE items 40 to 80 hours after the original administration; the link to the online retest survey was sent to individuals about 40 hours after submitting the first survey, with instructions to complete it within the next 24 hours. We continued to invite participants to complete the retest survey until at least 200 were completed within the 40- to 80-hour window.

We selected 40 hours as the minimum retest period to mitigate the potential for recollection bias. We selected 80 hours as the maximum duration between tests to minimize natural changes in PC and PRSE.

After completing study procedures, participants either received a small incentive from the panel company or were offered $25 (for completing the focus group, cognitive interview, or 1 cross-sectional survey) or $35 (for completing 2 surveys in the cross-sectional study), if recruited from other sources. Participants who were interviewed about the guidance statement received $20. The University of Washington Human Subjects Institutional Review Board approved all procedures.

Data Analysis

Qualitative Data

Following the best practices for instrument development, 2 doctoral-level researchers independently analyzed and coded focus group transcripts. A third researcher reconciled discrepancies when necessary. After each focus group, we constructed a saturation matrix to guide the number of focus groups needed. Saturation was achieved once no new information was identified. We used cognitive interview notes and recordings to compile a detailed summary that we used to decide whether items needed modification and whether we should delete them.

Large-Scale Cross-sectional Data

Item bank confirmation. We completed analyses to (1) confirm construct unidimensionality, (2) evaluate local dependence, (3) calibrate items to a graded-response IRT-model,[58] (4) evaluate item fit, (5) examine differential item functioning (DIF), (6) examine test–retest reliability, and (7) evaluate scale reliability and construct validity of the bank and short forms. Before fitting items to an IRT model, we confirmed unidimensionality of the bank. This was completed using confirmatory factor analyses (CFA), using the mean- and variance-adjusted weighted least squares estimator in M-plus 7.2.[59] A comparative fit index value of 0.90 or greater is evidence of acceptable model fit, and 0.95 or higher indicates excellent model fit.[60] We also examined local independence violations, using the matrix of residual correlations from the CFA. We used residual correlations greater than 0.2 (Kim et al[62]) to identify item pairs with local dependence (eg, items highly correlated with one another). Once the assumption of unidimensionality was met, we calibrated items to a graded-response IRT model using IRTPRO.[62] IRTPRO uses all available data and only uses listwise deletion to calculate the coefficient α. We evaluated item fit using Orlando and Thissen's[63] S-χ^2 calculated by IRTPRO. We judged items with an S-χ^2 $P < .01$ as misfitting and removed them from the item bank. We used an iterative process to identify problematic items that did not meet the local independence assumption or had poor item fit. We dropped 1 item from each item pair with local dependence

based on item preference or lack of fit to the IRT model, and reran the model until no items displayed local dependence or had significant misfit. Once an item set was identified that met these requirements and had CFA statistics that supported unidimensionality, we completed additional analyses to evaluate DIF using the lordif[64] program in R.[65] Consistent with analysis plans used by major measurement initiatives,[44] we evaluated DIF by sex (male versus female), age (<46, 46-65, ≥66 years), and education (less than college degree versus college or professional degree). We considered items with a change in pseudo-R^2 statistic <0.13 (Zumbo et al[66]) or <5% change in β coefficients[67] to have significant DIF. We further examined any item that met the criteria for DIF by calculating the impact of DIF on participants' scores. We did this by calculating full bank T-scores for participants using subgroup-specific parameter estimates and comparing those scores to T-scores calculated using overall sample parameter estimates. If score differences were less than 1 point on the T-score metric, we considered DIF negligible and kept the item in the bank. We dropped from the bank items with DIF that we considered nonnegligible. The final item bank, therefore, consisted of items that had adequate fit to the IRT model and showed no significant local dependence or DIF.

The graded-response 2-parameter IRT model generates a difficulty and discrimination parameter for each item. Items with higher discrimination parameters are better at differentiating among respondents with similar scores. The difficulty parameter represents the level of the construct (ie, PC or PRSE) at which a person is most likely to choose a given response. For example, answering yes to the question "Are you able to walk a block?" requires less physical ability (and therefore is an easier item) than saying yes to "Are you able to run a mile?," which would have a higher difficulty parameter.

Short form selection. We selected 6 items from the final item bank to comprise a fixed-length short form for administration in cases where CAT or computer administration is not feasible. We chose short form items by considering items' IRT parameters, expert panel item preference, and content coverage. We also evaluated the appropriateness and function of the 5-category response option set by visually inspecting the category response curves generated by IRT analyses. Ideally, category response curves have distinct peaks, indicating that

at a certain level of the train each person has the highest probability of choosing only 1 response.[68] In addition to the 6-item short form, we created a 2-item short form to use in situations where extremely short scales are needed. We chose the 2 items by balancing item difficulty and expert panel item preference. We examined representativeness of the short forms by evaluating the correlation between each short form and PRSE and PAS scores based on the final item bank.

Reliability. We evaluated test–retest reliability for full bank scores and the short forms using the 2-way random-effects (2,1) intraclass correlation coefficient (ICC)[69] using Stata14.[70] Only individuals who completed the second survey within the 40- to 80-hour window were included in these analyses. For a scale, an ICC of 0.7 is thought to indicate acceptable test–retest reliability.[69] We also evaluated reliability of the final item bank and each short form by examining the IRT-based test information function graphs. Reliability of the scale and short forms across the trait spectrum can be determined by evaluating the trait levels for which the information function is above 5 or 10 points. Information function is the IRT equivalent of the concept of reliability in classical test theory. When information is large, item parameters can be estimated with precision (ie, the parameter estimates are reliable). It is an indication of item quality and the item's ability to differentiate among respondents. Reliability of 0.8, which is considered sufficient for group comparisons, is equivalent to scale information of 5. Similarly, information of 10 is equivalent to reliability of 0.9, which is considered sufficient for individual comparisons.[71] We also calculated the percentage of the sample within the effective measurement range for both the full bank and the 6-item short form.

Validity. We evaluated the item bank's construct validity by examining Pearson correlations between PAS or PRSE scores and the legacy (eg, PCS and PSEQ) total and subscale scores, PROMIS anxiety, PROMIS depression, and PROMIS pain interference. We hypothesized that PAS and PRSE scores would be highly positively correlated (about $r > 0.7$) with legacy scores of the PCS and PSEQ. In addition, based on previous literature, we hypothesized the PRSE would be moderately negatively (about $r < –0.5$) correlated with anxiety, depression, and pain interference.[72,73] Conversely, we hypothesized that PAS would be moderately positively

(about $r < 0.5$) correlated with anxiety, depression,[74] and pain interference.[38,75] We assessed known groups validity by comparing the mean PAS or PRSE scores of participants with different levels of pain intensity (mild, moderate, severe) using a 1-way analysis of variance with Bonferroni correction for multiple comparisons, and by gender using the 2-sample *t* test.

Based on results seen using the PCS,[38] we hypothesized that average scores of people with higher pain intensity will have statistically significantly higher scores of PC and that women will report higher levels of PC. Similarly, we hypothesized that those with higher pain intensity would report significantly lower scores of PRSE.[72]

Handling missing data. Instrument development using classical test theory will omit records that have missing data; thus, missing data used to be a significant issue when developing instruments. Instrument development and calibration using IRT uses all available data for each item, and as such responses from participants with partial missing data can still be utilized in IRT analyses.[62]

RESULTS

Refer to Tables 1 and 2 for information on participant flow and details of the study phases. No adverse events or deaths occurred among study participants.

Qualitative Study Results

Expert Panel

The expert panel consisted of pain researchers and clinicians who treat people who have chronic pain. The following definitions of PRSE and PC were agreed on by the expert panel and reviewed by the patient advisors and focus group participants.

- **PRSE** is a person's confidence in his or her ability to minimize the impact of pain on physical and psychological functioning (eg, sleep, fatigue, mood), activities (eg, leisure activities, self-care) and participation (eg, work responsibilities, social interactions, relationships). Identified subdomains include control/tolerance of/cope with symptoms, ability to manage the impact of pain on mood and psychological functioning and interpersonal relationships, and confidence to accomplish goals despite pain.

- **Pain catastrophizing** cognitions are extremely negative appraisals (thoughts) about pain and its impact on one's life now and in the future. This term includes magnification of pain and its impact, helplessness, rumination, and beliefs about the worst-case scenarios.

Table 1. Participant Flow

	Study phase, n		
	Focus groups	**Cognitive interviews**	**Large-scale administration**
Started	**28**	**33**	**1146**
Completed	**19**	**22**	**795**
Not completed	**9**	**11**	**351**
Adverse event	0	0	0
Death	0	0	0
Lack of efficacy	0	0	0
Lost to follow-up	3	7	200
Physician/Investigator decision	0	0	62
Pregnancy	0	0	0
Protocol violation	0	0	0
Withdrawal by subject	6	4	11
Incomplete	0	0	78

Table 2. Details of Study Phases

	Study phase			
	Focus groups	**Cognitive interviews**	**Large-scale administration**	
Aim or goal	Develop an item pool for PC and PRSE that covers all subdomains	Evaluate the items from the patients' perspective using cognitive interviewing	Conduct IRT analyses to calibrate the item banks	Examine psychometric properties of the new PC and PRSE scores (construct validity, known group validity, test–retest reliability)
Description	In-person focus groups convened to discuss PC and PRSE and to identify potential gaps in the subdomains. New items will be developed as needed to adequately cover all subdomains and the entire range of the construct continuum.	This is a 1-time interview. Cognitive interviews help identify problematic items. Participants are asked to paraphrase the item, explain their understanding of important terms, and report their level of confidence in the answers they provide to the questions.	Participants respond to candidate PC and PRSE items.	Participants respond to legacy gold standard measures (ie, PSEQ and PCS); PROMIS Profile 29 to measure pain interference, pain intensity, physical function, depression, anxiety, anger, and social function; and PROMIS Global to provide global physical and mental quality of life scores.
Time frame	A single 1- to 2-h focus group	A single 1-h interview	1 survey or 2 surveys 40-80 h apart	
No. of participants analyzed	19	22	795	

Abbreviations: IRT, item response theory; PC, pain catastrophizing; PCS, Pain Catastrophizing Scale; PROMIS, Patient-Reported Outcomes Measurement Information System; PRSE, pain-related self-efficacy; PSEQ, Pain Self-Efficacy Questionnaire.

In addition to the definitions, the expert panel recommended the context of "How confident are you that. . . ?" and 5-point response options (not at all, a little bit, somewhat, quite a bit, very much) for all PRSE items. The panel preferred that no time frame be used for PRSE (eg, in the past 7 days). For PC, the expert panel similarly preferred a 5-point response scale (never, rarely, sometimes, often, always) and proposed the question prompt of, "In the past 7 days, how often did you have the following thought when you were in pain?"

Last, the expert panel discussed whether or how to address the issue of pain medication when asking about PRSE, because a person who is confident they can manage pain without medication likely has higher PRSE than a person who manages pain with medication. The expert panel decided to include 2 questions that specifically asked about PRSE strategies that did not involve medication: "How confident are you that. . . " (1) ". . . you can do things other than taking medication to limit the effects of pain?" and (2) ". . . you can cope with your pain without medication?", but they agreed not to specify with or without medication in most items.

However, because we recognize that the use of pain medication is 1 of many strategies individuals use to manage their pain, we did not specifically ask people during focus groups and cognitive interviews to distinguish their responses to discussions and items pertaining to PRSE regarding pain medication use.

Focus Groups

Nineteen individuals participated in 3 focus groups and discussed both PRSE and PC. Participants included 10 women and 9 men; 58% were White. Medical conditions reported included low back pain, osteoarthritis of the knee, painful diabetic neuropathy, and amputation. Additional participant demographic and clinical characteristics are summarized in Table 3. Clinically relevant levels of PC on the PCS were reported by 16% of the sample and 11% reported PSEQ scores that were <20, indicating low levels of pain self-efficacy.

Pain-related self-efficacy. Participants reviewed the proposed definition of PRSE and identified no new PRSE subdomains. The same subdomains identified by the expert panel were used to code the transcripts. No new codes were identified from the coding of transcripts.

Table 3. Focus Group and Cognitive Interview Sample Demographics and Clinical Characteristics

	Focus groups (n = 19)	Cognitive interviews (n = 22)	Overall (N = 41)
	n (%)	n (%)	n (%)
Sex			
Female	10 (52.6)	11 (50.0)	21 (51.2)
Male	9 (47.4)	11 (50.0)	20 (48.8)
Race			
American Indian or Alaska Native	1 (5.3)	1 (4.5)	2 (4.9)
Asian	0 (0)	1 (4.5)	1 (2.4)
Black or African American	2 (10.5)	2 (9.1)	4 (9.8)
White	11 (57.9)	15 (68.2)	26 (63.4)
>1 race	5 (26.3)	1 (4.5)	6 (14.6)
Unknown or not reported	0 (0)	2 (9.1)	2 (4.9)
Ethnicity			
Not Hispanic or Latino	17 (89.5)	20 (90.9)	37 (90.2)
Unknown or not reported	2 (10.5)	2 (9.1)	4 (9.8)
Marital status			
Married/Living with significant other	7 (36.8)	6 (27.3)	13 (31.7)
Other	12 (63.2)	14 (63.6)	26 (63.4)
Not reported	0 (0)	2 (9.1)	2 (4.9)
Education			
Some college or less	13 (68.4)	10 (45.5)	23 (56.1)
College degree or higher	6 (31.6)	10 (45.5)	16 (39.0)
Not reported	0 (0)	2 (9.1)	2 (4.9)
Age, by category			
≤18 y	0 (0)	0 (0)	0 (0)
18-65 y	12 (63.2)	16 (72.7)	28 (68.3)
≥65 y	7 (36.8)	4 (18.2)	11 (26.8)
Age, mean (SD), y	59.1 (13.3)	59.1 (8.6)	59.5 (11.0)
Age, median, y	59.9	60.0	59.9
Annual household income, median (min-max), $	13 000 (721-135 900)	24 000 (1700-250 000)	16 918 (721-250 000)

	Focus groups (n = 19)	Cognitive interviews (n = 22)	Overall (N = 41)
	n (%)	n (%)	n (%)
Diagnosis based on survey information	**n (%)**	**n (%)**	**n (%)**
Low back pain	13 (68.4)	13 (59.1)	26 (63.4)
Osteoarthritis of the knee	6 (31.6)	7 (31.8)	13 (31.7)
Diabetes-related neuropathic pain	5 (26.3)	4 (18.2)	9 (22.0)
Multiple sclerosis	0 (0)	2 (9.1)	2 (4.9)
Spinal cord injury	4 (21.1)	6 (27.3)	10 (24.4)
Lower limb amputation	5 (26.3)	8 (36.4)	13 (31.7)
Not reported	0 (0)	2 (9.1)	2 (4.9)
Most painful condition			
Low back pain	6 (31.6)	7 (31.8)	13 (31.7)
Osteoarthritis of the knee	5 (26.3)	0 (0)	5 (12.2)
Diabetes-related neuropathic pain	2 (10.5)	1 (4.5)	3 (7.3)
Multiple sclerosis	0 (0)	2 (9.1)	2 (4.9)
Spinal cord injury	4 (21.1)	5 (22.7)	9 (22.0)
Lower limb amputation	2 (10.5)	5 (22.7)	7 (17.1)
Not reported	0 (0)	2 (9.1)	2 (4.9)
Regular use of pain medication	14 (73.7)	13 (59.1)	27 (65.9)
Neuropathic pain			
S-LANSS score ≥12, suggesting neuropathic pain	9 (47.4)	11 (50.0)	20 (48.8)
Pain detect (neuropathic pain likely)	5 (26.3)	7 (31.8)	12 (29.3)
PROMIS neuropathic pain quality T-score, mean (SD)	53 (8.1)	55.6 (7.7)	54.7 (7.9)
T-score ≥50, suggesting neuropathic pain, No. (%)	13 (68.4)	15 (68.2)	28 (68.3)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**
Years since diagnosis	18.9 (9.3)	16.5 (12.2)	17.7 (10.8)
Average pain intensity at interview (0-10)	6.4 (1.8)	6.9 (1.4)	6.7 (1.6)
PSEQ	34.8 (14.1)	28.4 (12.4)	31.6 (13.5)
PCS			
Total	18.6 (12.8)	22.4 (11.6)	20.5 (12.2)

	Focus groups (n = 19)	Cognitive interviews (n = 22)	Overall (N = 41)
	n (%)	n (%)	n (%)
Rumination subscale	6.3 (4.8)	8.5 (3.7)	7.4 (4.3)
Magnification subscale	4.2 (3.4)	4.8 (3.1)	4.5 (3.2)
Helplessness subscale	8.1 (5.5)	9.2 (5.6)	8.6 (5.5)
PROMIS T-scores			
Global mental health	43.6 (9.0)	41.1 (7.6)	42.3 (8.3)
Global physical health	38.8 (8.6)	32.8 (5.4)	35.8 (7.7)
Physical function	36.3 (7.5)	31.5 (5.7)	33.8 (7.0)
Anxiety	57.2 (9.5)	59.2 (8.6)	58.2 (9.0)
Depression	55.1 (9.9)	56.2 (10.0)	55.7 (9.9)
Fatigue	52.8 (9.6)	61.7 (8.5)	57.3 (10.0)
Sleep disturbance	51.2 (5.9)	60.0 (10.3)	55.7 (9.5)

Abbreviations: PCS, Pain Catastrophizing Scale; PROMIS, Patient-Reported Outcomes Measurement Information System; PSEQ, Pain Self-Efficacy Questionnaire; S-LANSS, Leeds Assessment of Neuropathic Symptoms and Signs pain scale.

Participants shared many feelings and experiences concerning their PRSE but primarily talked about their strategies for managing pain. One participant with low back pain shared that he could *control and cope with symptoms* by accepting that pain was now a part of his life: "I had to make friends with my pain a long time ago. It was going to be there for the rest of my life and it was going to stop me from doing things." The same participant shared he had *confidence to accomplish goals despite pain* owing to his ability to focus on all the things he can do: "[I] . . . acknowledge what I can do, and I just keep doing what I can do, and when I have a little loss, I notice it's a little loss because I still have so much I can do. And I'm not going to stop." Another participant with amputation shared that because of previous experiences with pain, he would be able to handle whatever pain-related difficulties may arise: "I used to get hurt as a young man and I learned to put one foot in front of the other." A participant with low back pain shared that she was able to *manage the impact of pain on mood, psychological functioning, and interpersonal relationships* by generating lists that reflected positive attributes about herself. She would list all she had been able to accomplish as a positive reminder that despite the pain, she can still accomplish what she wants to. Last, another man with low back pain and

osteoarthritis of the knee shared that he was able to *manage the impact of pain* by just getting up and going: "No matter what, when I'm feeling I don't want to do it, that's when I get up and go more. That's when I go do it, because I know I'm feeling like, 'oh, I'm looking for excuses for some reason not to do it.'" Overall, participants talked about their PRSE in a positive manner, with many participants sharing the psychological strategies they utilized to help themselves feel self-efficacious about their ability to manage their pain.

Pain catastrophizing. Participants reviewed the proposed definition and suggested no new components or changes. We coded transcripts using the subdomains of the proposed definition: *magnification of pain and its impact, helplessness, rumination, and beliefs about the worst-case scenarios.* No new subdomains were identified based on the transcripts. Contrary to PRSE, it was difficult for participants to talk about catastrophizing because they felt catastrophizing was unhelpful and just made it more difficult to live with chronic pain. Those who did speak about PC shared that when they catastrophize, they often think about *the worst-case scenarios*, such as that they may not be able to work or that their relationships will fall apart because of pain. As one man with knee osteoarthritis and low back pain shared, when he was in pain he would be mean to family and friends and, as a result, he worried about losing them: "That's probably the main thing I worry about, is losing everybody around you that cares a little bit." Others talked about sometimes *magnifying pain and its impact*. A man with knee osteoarthritis shared: "[I] then think, well, I'm going to go here or do this and I'm going to be miserable the whole time, so I stopped going out of my house as much, and it wasn't, I don't think it was based on reality." Another participant with osteoarthritis of the knee shared: "I think we do blow it up bigger than it should be. I wish I could be a positive person all the time, but it's hard when you have all that pain. There's a lot of negative there." Several participants expressed *helplessness* in dealing with pain. "I've been through a lot of stuff in my life and this pain is just killing me," was a sentiment several participants shared. Another person with low back pain opened up about feeling angry when he feels helpless: "But most of the time, I'm just angry that I am feeling it, angry that I can't do the things that my peers do or that I feel like an adult that's in society should be able to do, and I feel like less of an adult if other people have to help me out with things." *Rumination*, a tendency to think repetitively about pain, was

discussed by only 1 participant with a diagnosis of knee osteoarthritis and concerned her wish to not be depressed as a result of her pain: “and then I can’t stop from getting depressed now. If I could stop it, maybe I could feel better about myself and wouldn’t think about the pain all the time.”

Finally, in addition to sharing their experiences with PC, participants almost uniformly reported that the term “pain catastrophizing” has a very negative and stigmatizing connotation: “It sounds like you’re making something up. Like you’re trying to blow it out of proportion and it doesn’t need to be.” They were worried that the catastrophizing scores would just support health care providers’ belief that their pain wasn’t real or be used to deny treatments. As a result, the name of the PC scale was changed to the Pain Appraisal Scale.

Overall, no new information was identified in the third focus group and saturation was achieved for both PRSE and PC.

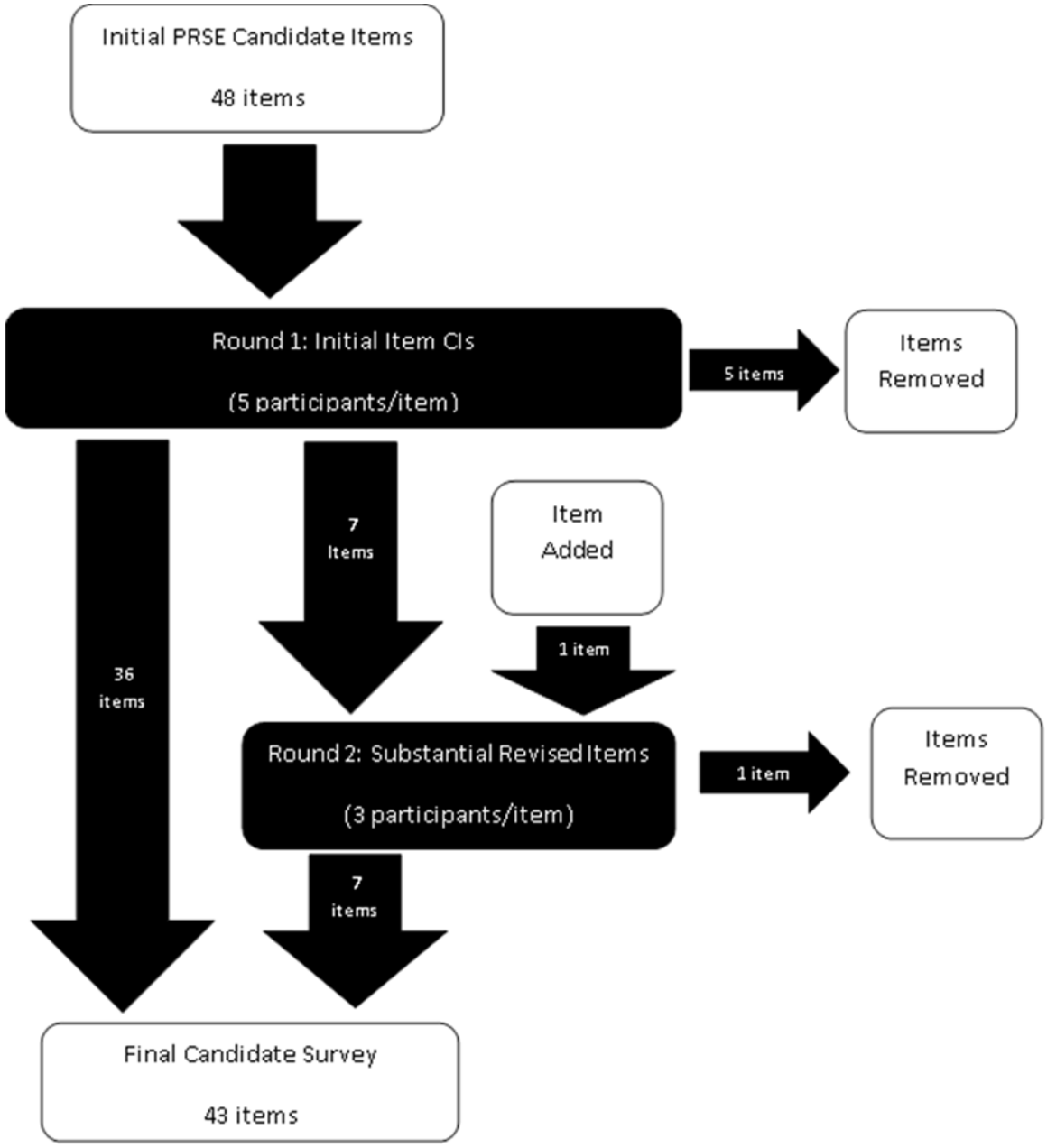
Cognitive Interviews

Twenty-two participants reviewed both the candidate PRSE and PC items. Eleven women and 11 men with numerous conditions (eg, low back pain, amputation, spinal cord injury) participated in the cognitive interviews. Their average age was 59.1 years and 77% were White (Table 3). Clinically relevant levels of PC on the PCS were reported by 25% of the sample, and 21% reported PSEQ scores that were <20, indicating low levels of pain self-efficacy.

Pain-related self-efficacy. We tested 48 PRSE items in cognitive interviews with people living with chronic pain. Of the initial 48 items, 36 (75%) functioned as intended and required no or minor revisions. In the first round of cognitive interviews, 5 items did not function well and were dropped. We substantially revised 7 items and tested them in a second round of cognitive interviews. We divided 1 item into 2 items, making a total of 8 items reviewed during the second round of cognitive interviews. Of these, 7 items worked well, while 1 still had issues and was deleted. The final PRSE set contained 43 items. Figure 1 illustrates this process. The response options worked well for all items and required no changes.

Item clarity. For 4 items, participants observed that their response would depend on their current level of pain. Therefore, we revised these items to indicate if participants should think about *typical* pain or *worst* pain, as appropriate. In addition, we modified 2 items for clarity and relevance. One item asked about maintaining relationships with a spouse or significant other. Participants who did not have a spouse or significant other said they could not answer. We revised that item to ask about maintaining relationships with people who are important to the individual. Another item asked about being able to live a normal lifestyle despite the pain. Many participants were not sure what "normal" meant; to address this, we revised the item to "How confident are you that you can do the things you most want to do in spite of your pain?"

Figure 1. Schematic Depicting Cognitive Interview Process for PRSE Item Bank



Abbreviation: PRSE, pain-related self-efficacy.

Dropped items. Eliminated items asked about the impact of pain on cognitive function (eg, learn new things, remember things, multitask) and social function. Participants reported that they could not meaningfully attribute problems with cognitive function to pain. Socializing with family was also problematic because some participants did not have family or did not care

to socialize with their family, so we dropped this item. Additionally, we eliminated the item "How confident are you that you can keep your pain from being the center of your life?" because participants reported that pain was the center of their life when they were in pain, but that it was not when they were pain free. Table 4 shows a full list of dropped PRSE items.

Table 4. Dropped PRSE and Pain Appraisal Candidate Items

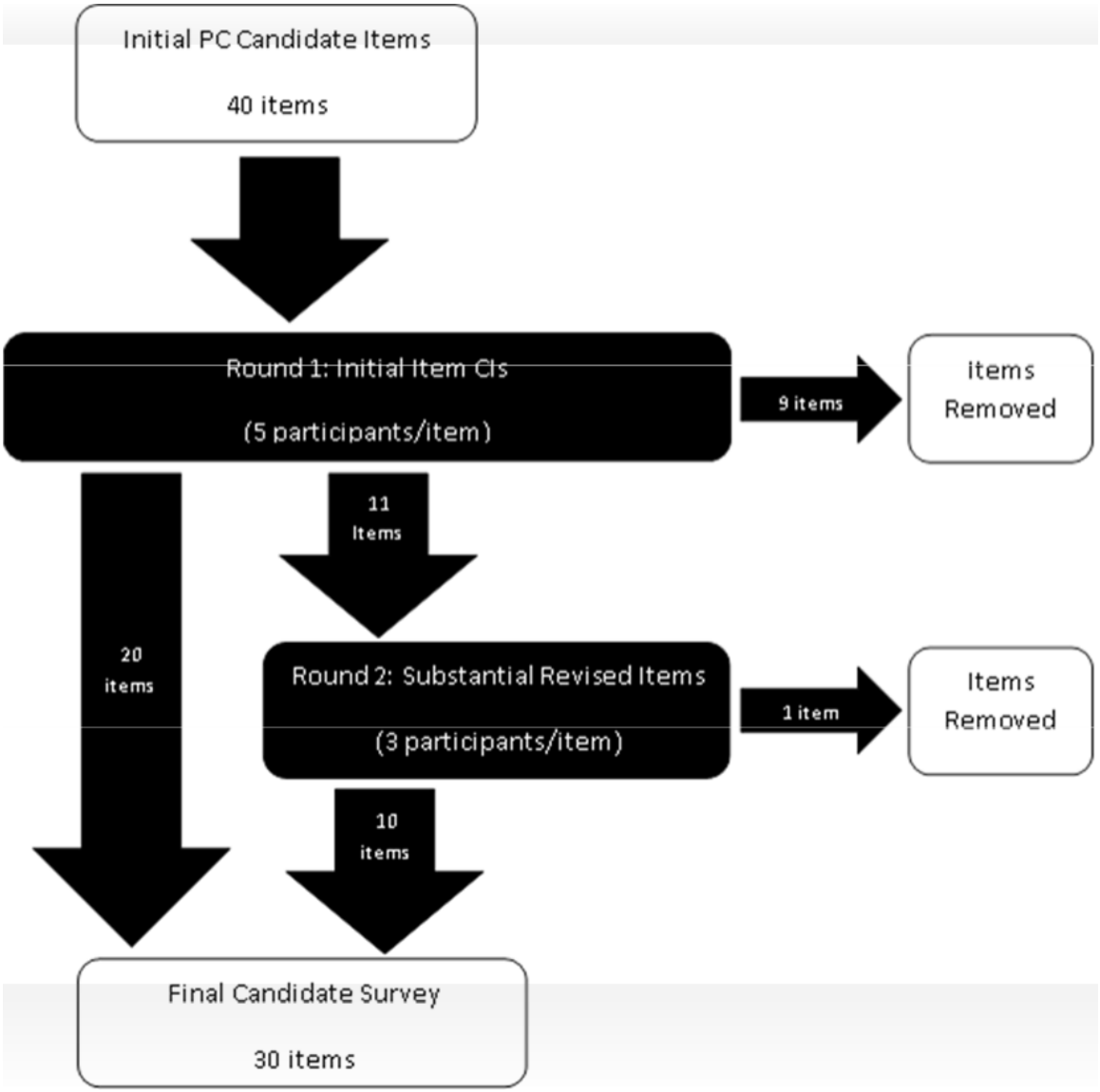
PRSE item bank	**Pain appraisal item bank**
How confident are you that…	*In the past 7 days, how often have you thought …?*
You can keep the pain from being the center of your life?	I can no longer do anything.
You can socialize with family in spite of your pain?	I see no end to my pain.
You can do several tasks at once when you are in pain?	I will forever depend on my family or friends.
You can learn new skills in spite of your pain?	I will never get a good night's sleep.
You can remember important things in spite of your pain?	I will be depressed for the rest of my life. This pain is awful. No one will ever want to be around me. Other people have to do everything for me. There is nothing I can do about my pain.

Abbreviation: PRSE, pain-related self-efficacy.

Pain catastrophizing. Participants gave feedback on the context ("How often did you have the following thought when you were in pain?"), response options (never to always), and wording of the candidate items. Of the initial 40 items, 38 (95%) were problematic and required modifications. Of the 38 items, we deleted 9 because they could not be reworded to function as intended, 18 required minor revisions, and 11 required substantial modifications.

The revised 11 items underwent a second round of cognitive interviews. Of the 11 items, 10 worked well and 1 was deleted. The final set of PC items for large-scale administration contained 30 items. Figure 2 demonstrates this process. Participants felt that 5 response options were appropriate and sufficient.

Figure 2. Schematic Depicting Cognitive Interview Process for Pain Appraisal Item Bank



Abbreviation: PC, pain catastrophizing.

Item clarity. Participants had difficulty separating the impact of pain on functioning from the impact of other health issues on functioning. Some items initially asked participants to rate how often they had a particular thought when they were in pain (eg, "I can no longer do anything, I am a burden on my family and friends"). However, participants often answered thinking about other health issues, not just pain. For example, for the item, "In the past 7 days, how often did you have the following thought when you were in pain? 'I will never get a good

night's sleep,'" participants were actually answering the question, "Do you have trouble sleeping?" rather than the intended question. To avoid this issue, we added "Because of my pain" to items.

Dropped items. Participants could not meaningfully answer some items. For instance, the item, "I will forever depend on my family or friends," was intended to describe a worst-case scenario. Instead, some participants felt they would always rely on their friends and family to a certain degree, regardless of pain, and others truly did need to have others do things for them because of other chronic conditions so the statement was literally true and not an endorsement of a worst-case scenario. Table 4 shows a full list of deleted items.

Large-Scale Cross-sectional Study Results

Participant Characteristics

We invited a total of 4982 individuals to participate and screened 3592 for eligibility. Of these, we presented 1146 eligible and interested participants with the survey, and 982 completed it. After attempting to rescreen panel company participants and reviewing completed surveys, we dropped 187 records (eg, unable to contact participant for rescreen, self-reported most or all the eligible health conditions, inconsistent responses to identical screening and survey questions). The final calibration sample included 795 individuals with chronic pain, with n = 206 completing the candidate item banks again 40 to 80 hours later (see Figure 3 for flow diagram). The sample was mostly female (64.4%), White non-Hispanic (79.9%), and an average age of 55.1 years old (SD, 16.8 years). A total of 18.6% participants (n = 148) completed the paper version of the survey and the remaining completed the online survey. Online survey participants were younger, tended to be female, were married or living with their significant other, and had a college degree or higher (all $P < .05$).

Table 5 shows additional participant characteristics and mean scores on legacy and validity measures. Twenty-two percent of the sample reported clinically relevant levels of PC on the PCS and 19% of the sample reported PSEQ scores that were <20, indicating low levels of pain self-efficacy.

Figure 3. Patient Flow Diagram

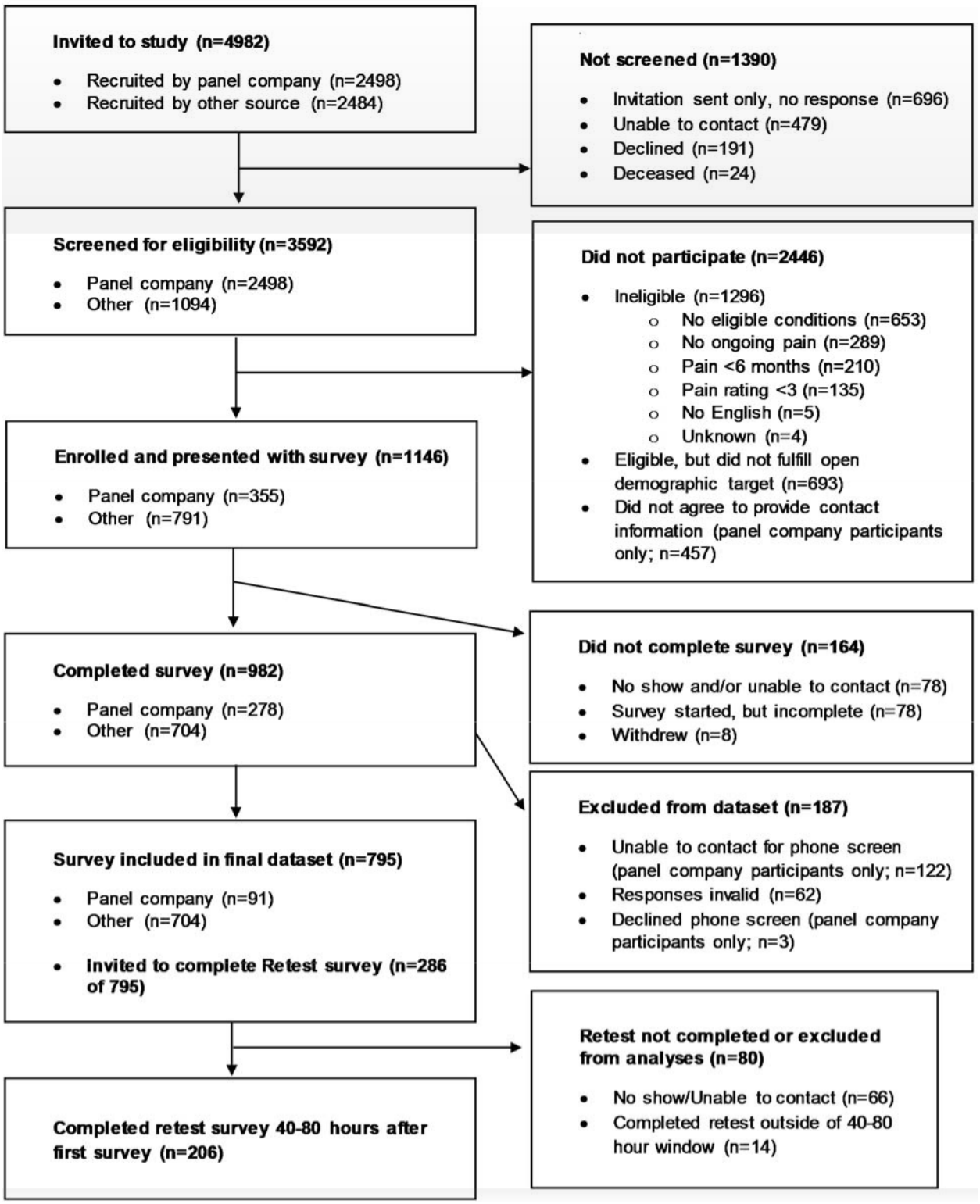


Table 5. Calibration Sample Demographics and Clinical Characteristics (N = 795)

Variable	Values
Age (n = 791), mean (SD), range, y	55.1 (16.8), 19-99
Age, median, y	57.6
Age, category, No. (%)	
≤18 y	0 (0%)
18-65 y	555 (70.1%)
≥65 y	236 (29.8%)
Sex, No. (%)	
Male	283 (35.6%)
Female	512 (64.4%)
Race (n = 778), No. (%)	
Caucasian	668 (85.9%)
African American	66 (8.5%)
Native American/Alaskan Native	4 (0.5%)
Asian	4 (0.5%)
Native Hawaiian/Pacific Islander	4 (0.5%)
>1 race	32 (4.1%)
Ethnicity (n =784), No. (%)	
Hispanic	55 (7.0%)
Non-Hispanic	729 (93.0%)
Education (n = 787), No. (%)	
Some high school	42 (5.3%)
High school graduate or GED	134 (17.0%)
Some college/AA	296 (37.6%)
College degree (BS/BA)	207 (26.3%)
Advanced degree	108 (13.7%)
Pain-related diagnosis (n = 788)	
Low back pain, No. (%)	363 (46.1%)
Osteoarthritis of the knee, No. (%)	60 (7.6%)

Variable	Values
Diabetes-related neuropathic pain, No. (%)	37 (4.7%)
Multiple sclerosis, No. (%)	172 (21.8%)
Spinal cord injury, No. (%)	114 (14.5%)
Lower limb amputation, No. (%)	42 (5.3%)
Average pain intensity at interview (n = 790; possible range: 0-10), mean (SD), range	5.8 (1.9), 1-10
Pain Self-Efficacy Questionnaire (n = 786; 0-60), mean (SD), range	32.2 (14.7), 0-60
Pain Catastrophizing Scale (n = 789)	
Total (0-52), mean (SD), range	19.4 (12.8), 0-52
Variable	
Rumination subscale (0-16)	7.4 (4.4), 0-16
Magnification subscale (0-12)	3.8 (3.1), 0-12
Helplessness subscale (0-24)	8.1 (6.1), 0-24
PROMIS T-scores (range n = 786-789), mean (SD), range	
Global mental health (21.2-67.6)	43.9 (9.9), 21.2-67.6
Global physical health (16.2-67.7)	37.2 (7.1), 16.2-61.9
Physical function (22.9-56.9)	35.4 (6.5), 22.9-56.9
Anxiety (40.3-81.6)	55.3 (10.0), 40.3-81.6
Depression (41.0-79.4)	54.7 (10.6), 41.0-79.4
Fatigue (33.7-75.8)	59.1 (9.1), 33.7-75.8
Sleep disturbance (32.0-73.3)	55.8 (8.7), 32.0-73.3
Ability to participate in social roles and activities (27.5-64.2)	43.9 (7.8), 27.5-64.2
Pain interference (41.0-78.3)	61.2 (6.7), 41.0-78.3
Neuropathic pain quality (37.0-74.0)	52.9 (8.5), 37.0-74.0

Abbreviations: AA, Associate of Arts; BA, Bachelor of Arts; BS, Bachelor of Science; GED, general educational development.

Missing Data

For the PRSE candidate bank, the level of missing data ranged from 1 missing response (ie, n = 794 participants responded to an item instead of 795) to 24 missing responses (for prse13, n = 771 responded to "You can/could have a pleasurable sex life in spite of your pain?"); most questions in the bank (n = 31 items) contained only 2 missing responses. Two other items contained a higher number of records with missing data (prse22 with 21 missing responses: "You can/could have a satisfying sexual relationship in spite of your pain?" and prse42 with 20 missing responses: "You can/could participate in sexual activity in spite of your pain?"). This pattern of missingness is typical, because items that relate to specific topics that do not apply to all individuals (eg, people who are not sexually active cannot meaningfully answer questions related to this topic, people who don't work outside the home can't answer questions related to employment, people who don't have children cannot provide meaningful responses to questions about children) always have more missing data than items that apply to all. Following best practices, the "not applicable" option was not offered and respondents were instructed to skip the item if they could not provide a meaningful answer. Providing a not applicable response resulted in missing data and the distinction was usually irrelevant to instrument development. For the PC candidate bank, the level of missing data ranged from 2 missing responses (n = 793 for 4 items) to 10 missing responses (n = 785 for 6 items); almost half the items in the candidate bank (n = 14 items) were missing 9 responses (n = 786 responded to the items). Seven participants accounted for most of the missing data, as they were mistakenly sent a paper survey with an old version of the PC item bank.

Pain-Related Self-efficacy

Final Item Bank Confirmation

The final item bank contained 29 items. Initial CFA and IRT analyses identified 14 items that needed to be removed because of local dependence, model misfit, or both. Once all 14 items were removed, CFA indexes supported a unidimensional model (CFI = 0.92). No items had residual correlations greater than 0.2. The remaining 29 items were fit to a graded-response IRT model with no items displaying significant misfit (all $S\text{-}\chi^2$ $P > .01$). DIF analyses found that no

items displayed DIF by age, gender, or education using either the R^2 statistic <0.13 or 5% change in β criterion. Final item bank parameter estimate slopes ranged from 1.4 to 3.1 and thresholds ranged from –3.34 to 2.40 (see Table 6). We generated participant final scores using item parameters from the final 29-item bank. We then converted the scores to a T-score metric so that the sample mean is 50 with an SD of 10.

Short Form Selection

To create a static short form for paper-and-pencil administration, we identified 6 items from the final item bank. Table 7 shows the 6 items chosen for the short form. We chose items that represented at least 1 of the identified domain areas of PRSE: control/tolerance of/cope with symptoms, ability to manage the impact of pain on mood and psychological functioning and interpersonal relationships, and confidence to accomplish goals despite pain. The correlation between the 6-item short form and the full item bank was 0.97 (see Appendix A for sum score to T-score conversions). We also generated a 2-item short form that includes 1 item pertaining to participation in daily activities (prse21) and another pertaining to management of pain during daily activities (prse23). The correlation between the 2-item short form and the full item bank was 0.90.

Table 6. PRSE Scale Items and Corresponding Item Response Theory–Based Item Parameters[a]

Item ID	Item stem	Discrimination	Difficulty			
			b1	b2	b3	b4
prse03	You can maintain your personal hygiene in spite of your pain?	1.6	–3.34	–1.97	–0.92	0.45
prse06	You can get necessary work done in spite of your TYPICAL pain (if you don't work outside of home consider household work or unpaid work)?,	2.37	–1.8	–0.73	0.45	1.64
prse23	**You can manage your pain during your daily activities?[b]**	2.5	–2.06	–0.85	0.3	1.8
prse07	You can do something to help yourself feel better when you are in pain?	1.69	–2.19	–0.84	0.47	2
prse20	You can take part in relaxing social activities (such as eating with others or visiting over coffee) in spite of your TYPICAL pain?	2.61	–2.01	–1.01	–0.13	1.02
prse11	**You can keep your pain from interfering with your social life?**	3.16	–1.38	–0.6	0.37	1.37
prse12	You can do many of the things you enjoy doing, such as hobbies or leisure activities, in spite of your pain?	2.91	–1.44	–0.47	0.49	1.63
prse15	You can keep your pain from interfering with the things you want to do?	2.75	–1.31	–0.4	0.66	1.97
prse17	You can keep your pain from interfering with family relationships?	2.5	–1.89	–0.93	–0.02	1.18
prse09	You can do some form of work in spite of your pain(work includes housework, paid and unpaid work)?	2.17	–1.79	–0.48	0.44	1.7
prse21	**You can do most of your daily activities in spite of your pain?[b]**	3.02	–1.83	–0.58	0.32	1.42
prse49	You can get necessary work done in spite of your WORST pain (if you don't work outside of home, consider household work or unpaid work)?	2.04	–0.95	–0.04	0.97	2.15
prse24	You can accomplish most of your goals in life in spite of your pain?	2.93	–1.21	–0.44	0.41	1.62
prse26	You can go shopping for groceries or clothes in spite of your pain?	2.15	–1.88	–0.77	0.23	1.33

Item ID	Item stem	Discrimination	Difficulty			
			b1	b2	b3	b4
prse29	You can cope with your pain in most situations?	2.43	–2.35	–1.16	–0.04	1.37
prse31	You can exercise daily in spite of your pain?	1.98	–1.14	0.03	1.06	2.05
prse32	You can keep your pain from interfering with managing financial affairs?	1.7	–2.05	–1.1	–0.3	0.84
prse33	You can maintain an active lifestyle in spite of your pain?	2.78	–1.04	–0.19	0.7	1.76
prse34	**You can be in a good mood in spite of your pain?**	2.02	–2.43	–1.25	–0.01	1.36
prse35	You can successfully plan activities or events when you're in pain?	2.6	–1.47	–0.63	0.37	1.56
prse36	**You can get a good night's sleep in spite of your pain?**	1.6	–1.02	–0.02	1.11	2.37
prse37	**You can do the things you most want to do in spite of your pain?**	2.95	–1.24	–0.36	0.62	1.79
prse39	You can maintain your physical appearance in spite of your pain?	2.14	–2.12	–1.11	–0.02	1.06
prse40	You can socialize with friends in spite of your pain?	3.02	–1.86	–0.82	0.05	1.03
prse44	You can minimize the effects of your pain on what you want to do?	2.54	–1.53	–0.52	0.62	1.88
prse45	You can cope with your pain without medication?	1.2	–0.88	0.05	1.12	2.5
prse46	You can maintain your oral health in spite of your pain?	1.43	–3.3	–1.84	–0.9	0.28
prse47	You can have a fulfilling life in spite of your pain?	2.68	–1.67	–0.92	–0.02	1.01
prse48	You can deal with the irritability your pain may cause you?	2.08	–2	–0.84	0.36	1.71

Abbreviations: ID, identifier; PRSE/prse, pain-related self-efficacy.

[a]Items in **bold** are included on the 6-item short form. All items have the following context: "How confident are you that. . . ?" All items have the response set: 1 = not at all, 2 = a little bit, 3 = somewhat, 4 = quite a bit, 5 = very much. The discrimination parameter and the difficulty parameter thresholds are provided by IRT analyses and used for administration by computerized adaptive testing, development of custom short forms, or scoring of custom short forms using IRT software.

[b]Items included on the 2-item short form. The 2-item short form is also recommended for inclusion at the end of the PAS 6-item short form to address issues of stigma relating to the PAS.

Table 7. PAS Items and Corresponding Item Response Theory–Based Item Parameters[a]

Item ID	Item stem	Discrimination	Difficulty			
			b1	b2	b3	b4
pas01	Because of my pain, I am a burden on my family or friends.	1.74	–1.02	–0.07	1.11	2.24
pas04	Because of my pain, I can't go on.	2.29	0.08	0.91	1.79	2.96
pas05	I can't stand my pain anymore.	2.24	–0.61	0.13	1	1.94
pas11	I will never be able to take care of my most basic needs because of my pain.	2.02	–0.26	0.62	1.67	2.63
pas12	**Because of my pain, I will never be happy again.**	3.36	–0.04	0.61	1.4	2.13
pas14	I will never be able to do many of the things I enjoy because of my pain.	2.38	–1.03	–0.19	0.89	1.78
pas16	Because of my pain, I will be in a bad mood for the rest of my life.	2.67	–0.08	0.65	1.6	2.36
pas17	Because of my pain, I will be unhappy for the rest of my life.	3.26	–0.05	0.58	1.34	2.17
pas18	My pain is terrible.	2.17	–1.61	–0.7	0.37	1.67
pas19	My pain overwhelms me.	2.61	–1.07	–0.22	0.8	1.9
pas21	Because of my pain, my life is over.	4.09	0.42	1.02	1.64	2.47
pas22	**Because of my pain, my life is terrible.**	3.87	–0.07	0.56	1.28	2.09
pas23	**My life will only get worse because of my pain.**	3.39	–0.3	0.37	1.14	2.03
pas24	**My pain is more than I can manage.[b]**	3.1	–0.48	0.22	1.15	2.02
pas26	My pain will get worse.	2.44	–1.16	–0.31	0.72	1.65
pas30	Because of my pain, something really bad is going to happen to me.	2.28	0.04	0.78	1.69	2.52
pas31	My pain means something is seriously wrong with me.	1.91	–0.69	0.12	1.09	1.93
pas32	My pain will become even more intense and hurtful in the coming years.	1.98	–1.4	–0.5	0.62	1.71
pas34	My pain will never end.	2.03	–1.19	–0.41	0.45	1.27
pas36	**Did you keep thinking about how much it hurts?[b]**	1.65	–2.29	–0.8	0.77	2.02

Item ID	Item stem	Discrimination	Difficulty			
			b1	b2	b3	b4
pas38	**Did you have trouble thinking of anything other than your pain?**	1.73	–1.54	–0.14	1.29	2.55
pas40	Could you only focus on how bad your pain feels?	1.98	–1.35	–0.04	1.25	2.51
pas44	Did your pain completely fill up your mind?	2.25	–0.86	0.11	1.2	2.33
pas45	Did you desperately want your pain to go away?	1.47	–2.29	–1.18	–0.06	0.87

Abbreviations: ID, identifier; PAS/pas, Pain Appraisal Scale.
[a]Items in **bold** represent 6-item short form recommendations. Items pas1 to pas34 have the following item context: "In the past 7 days, how often did you have the following thought when you were in pain?" Items pas36 to pas45 have the context: "In the past 7 days, how often. . . ?" All items have the response set: 1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always. The discrimination parameter and the difficulty parameter thresholds are provided by IRT analyses and used for administration by computerized adaptive testing, development of custom short forms, or scoring of custom short forms using IRT software.
[b]Items with * represent recommendations for the 2-item short form. The 2-item PRSE short form is also recommended for inclusion at the end of the PAS 6-item short form to address issues of stigma.

Reliability

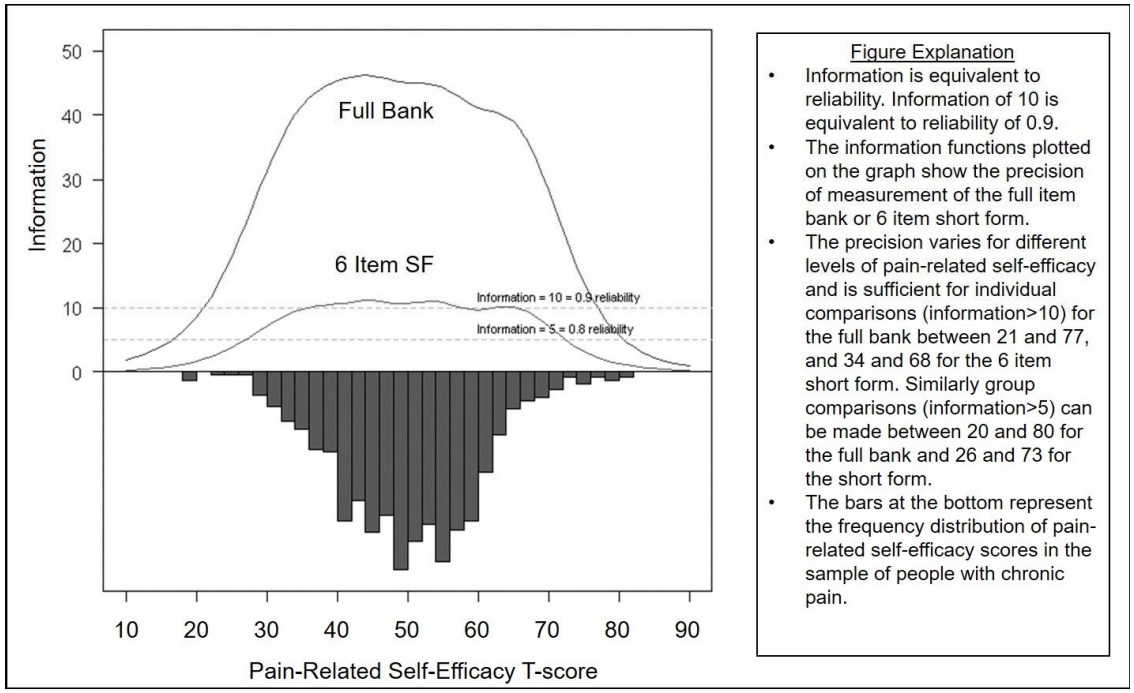
Within the 40- to 80-hour period, 206 participants completed the second survey for test–retest reliability analyses. Test–retest reliability was high for the full item bank (ICC[2,1] = 0.95 [95% CI, 0.93-0.96]) and the 6-item short form (ICC[2,1] = 0.92 [95% CI, 0.90-0.94]). The 2-item short form (ICC[2,1] = 0.86 [95% CI, 0.81-0.87]) demonstrated acceptable test–retest reliability. Reliability was also evaluated by examining the range of scores for which the scales information function was above either 5 or 10 points. For the full bank, reliability was high (information >10) between 21 and 77 and acceptable for group comparisons (information >5) between 20 and 80. Similarly, for the 6-item short form, reliability was high (information >10) between 34 and 68 (see Figures 4 and 5) and acceptable (information >5) for group comparisons between 26 and 73 on the T-score metric (see Figure 5). Reliability of the 2-item short form never reached acceptable levels for individual comparisons; group comparisons (information >5) were reliable between 50 and 54 on the T-score metric. The percentage of participants measured with reliability greater than 0.9 (information >10) was 99.0% for the full bank and 93.3% for the 6-item short form. Similarly, the percentage of those measured with reliability greater than 0.8 (information >5) was 100% and 98% for the full bank and short form,

respectively. The percentage of those measured with reliability >0.8 (information >5) was 18% for the 2-item short form.

Validity

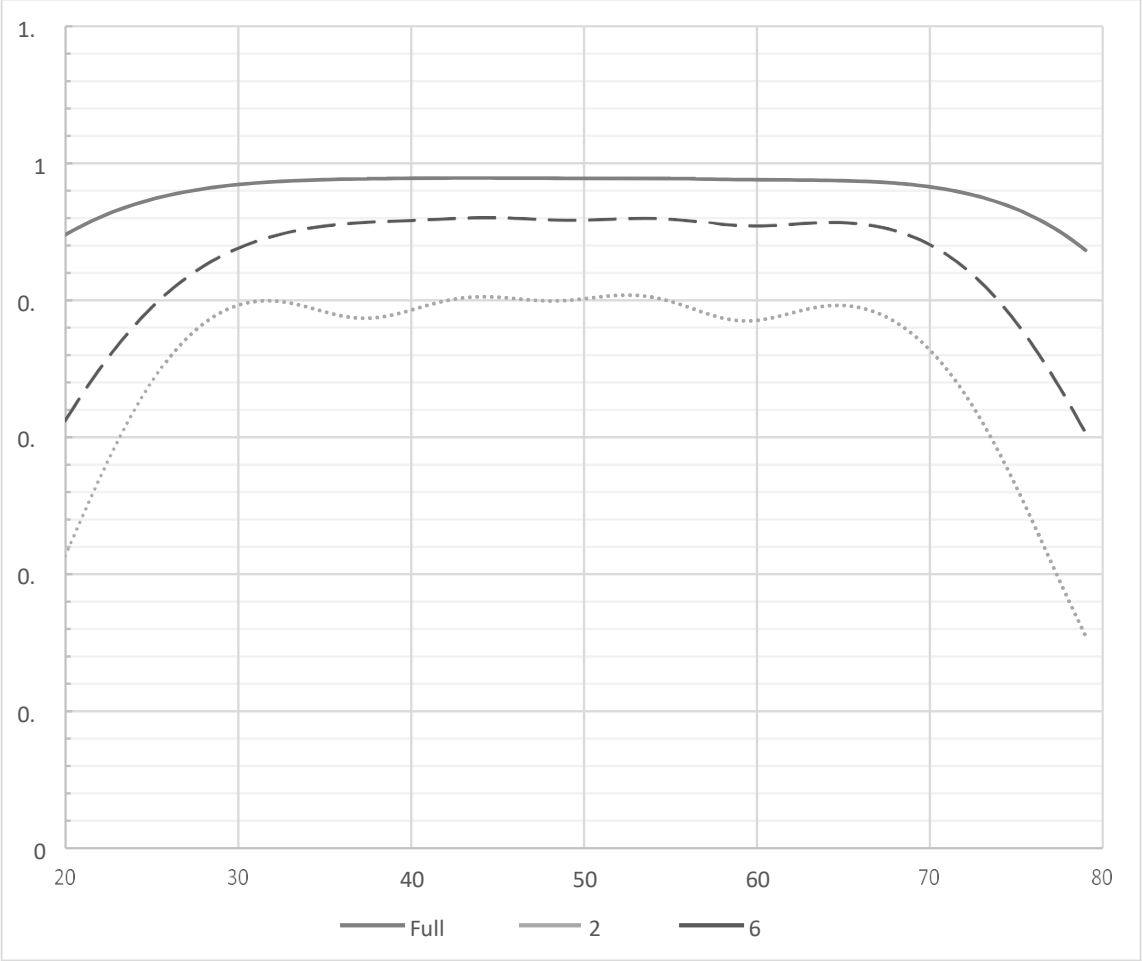
All correlations with identified measures were in the hypothesized direction and magnitude supporting construct validity of the PRSE. PRSE correlation with the legacy PSEQ scale was 0.87. Correlations between the PRSE and PROMIS measures included depression (–0.58), anxiety (–0.55), and pain interference (–0.77). Known groups validity was also supported, as scores between individuals with mild, moderate, and severe pain differed significantly at the $P < .05$ level, ($F[2786] = 93.4$; $P < .0001$). Using a Bonferroni multiple-comparisons post hoc test, all group scores differed significantly ($P < .01$) from one another; mild (mean, 55.7; SD, 8.2), moderate (mean, 51.0; SD, 8.4), and severe (mean, 45.0; SD, 9.9). However, PRSE scores did not differ significantly at the $P < .05$ level between men (mean, 50.6; SD, 9.9) and women (mean, 49.7; SD, 9.8; $t[793] = 1.36$; $P = .18$).

Figure 4. Information, Reliability, and Histogram of Participant PRSE Scores on T Scale Metric



Abbreviations: PRSE, pain-related self-efficacy; SF, short form.
The histogram shows the frequency of participant PRSE scores on the T-score metric. The full item bank and 6-item short form lines show score reliability across the continuum of PRSE T-scores for both classical and IRT frameworks.

Figure 5. Reliability of Full Bank, 6-Item Short Form, and 2-Item Short Form PRSE Scores Along the T-Scale Metric



Abbreviation: PRSE, pain-related self-efficacy.
Scores above 0.9 are considered reliable enough for individual comparisons, as are those above 0.8 for group comparisons. A T-score of 50 is the mean of the calibration sample. For the full bank, reliability was 0.9 or higher between 21 and 77 and 0.8 or higher between 20 and 80. Similarly, for the 6-item short form, reliability was 0.9 or higher between 34 and 68 and 0.8 or higher between 26 and 73 on the T-score metric. Reliability of the 2-item short form was 0.8 or greater between 50 and 54 on the T-score metric, and it did not reach 0.9 for any score.

Pain Appraisal Scale

Final Item Bank Confirmation

After completing the initial CFA and IRT analyses, we removed 6 items owing to problems with local dependence, misfit to the IRT model, or both. These items focused on a variety of subtopics, including helplessness, belief in worst case, magnification of impact, and rumination, with no single subtopic being eliminated from the bank. After we eliminated the 6

items, CFA indexes supported unidimensionality of the model (CFI = 0.94) and no items displayed residual correlations greater than 0.2. The graded-response IRT model was fit on the remaining 24 items and no items displayed significant misfit (all S-χ^2 $P > .01$). Subsequent DIF analyses on the 24 items found that 1 item (pas11: "I will never be able to take care of my most basic needs because of my pain") displayed DIF by age when using the 5% change in β coefficients[62] criterion. No other items displayed DIF by age, gender, or education using either the R^2 statistic <0.13 or 5% change in β criterion. Subsequent analyses indicated that the impact of the DIF from item pas11 on PAS score was very small, with mean score differences of less than 0.005 (SD, 0.005) on the T-score metric when comparing scores adjusted for DIF and not. Thus, we retained item pas11 despite the statistically significant DIF observed. Final item bank parameter estimates had slopes that ranged from 1.5 to 4.1 and thresholds that ranged from –2.29 to 2.96 (see Table 7). Final scores for participants were generated using item parameters from the final 24-item bank and were converted to the T-score metric (see Appendix B for sum score to T-score conversions).

Short Form Selection

Once we generated final item parameters, we chose 6 items to create a fixed-length short form. We chose at least 1 item to relate to each of the 4 components of magnification of pain and its impact, helplessness, rumination, and beliefs about the worst-case scenarios. Table 7 delineates the 6 items we chose for the short form. We also identified a 2-item short form; it includes 1 item each that relates to helplessness (pas24) and rumination (pas36). Correlations between PAS scores based on the full item bank and the 6-item and 2-item short forms were 0.997 and 0.714, respectively.

Reliability

The test–retest reliability was high for the full bank (ICC[2,1] = 0.93 [95% CI, 0.89-0.96]), the 6-item short form (ICC[2,1] = 0.91 [95% CI, 0.87-0.93]), and the 2-item short form (ICC[2,1] = 0.85 [95% CI, 0.79-0.89]). For the full bank, reliability was high (information >10) between 31 and 80 on the T-score metric and acceptable for group comparisons (information >5) between

25 and 80. Similarly, for the 6-item short form, reliability was high between 44 and 74 and acceptable for group comparisons between 40 and 78 on the T-score metric (see Figures 6 and 7). Reliability of the 2-item short form never reached acceptable levels for either group or individual comparisons (see Figure 7). The percentage of participants measured with reliability greater than 0.9 was 97.5% for the full bank and 70.3% for the 6-item short form. Similarly, the percentage of those measured with reliability greater than 0.8 was 99.2% and 85.6% for the full bank and short form, respectively.

Validity

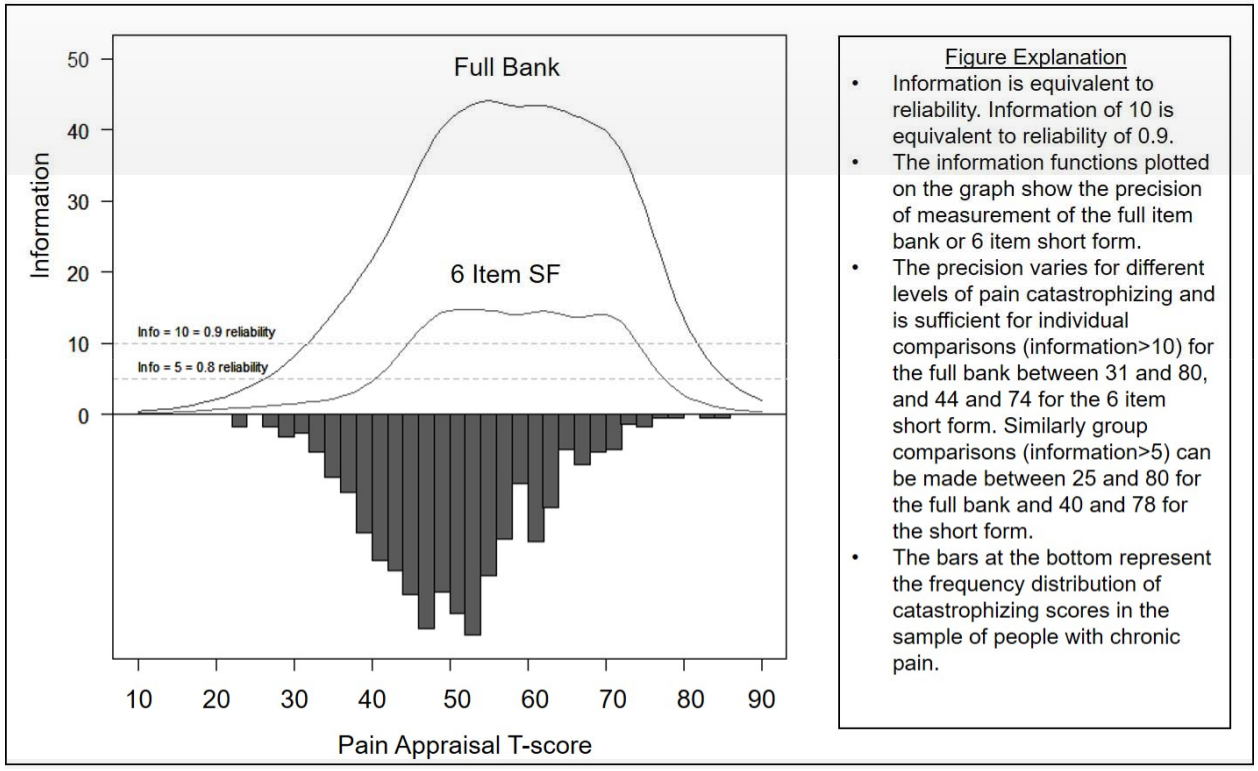
Construct validity of the scale was supported as all correlations with identified measures were in the magnitude and direction hypothesized. Correlations between full bank PAS scores and the legacy PCS scale and subscales were 0.83 for the PCS total score, 0.73 for PCS rumination, 0.75 for PCS magnification, and 0.83 for PCS helplessness. Correlations between full bank PAS scores and PROMIS depression, anxiety, and pain interference scores were 0.74, 0.72, and 0.72, respectively. Known groups validity was also supported, as full bank PAS scores between individuals with mild, moderate, or severe pain intensity levels were significantly different at the $P < .05$ level ($F[2,786] = 139.3$; $P < .0001$). Post hoc comparisons using the Bonferroni multiple-comparisons test indicated the mild group (mean, 43.3; SD, 7.4) was significantly different ($P < .01$) than both the moderate (mean, 49.1; SD, 8.0) and severe (mean, 55.9; SD, 9.5) groups, and the moderate and severe groups were also significantly different ($P < .01$). However, comparisons by gender did not follow the hypothesized pattern, as mean PAS scores between men (mean, 49.96; SD, 9.6) and women (mean, 50.01; SD, 9.9) were not significantly different ($t[792] = –0.07$; $P = .94$).

DISCUSSION

Study Results in Context

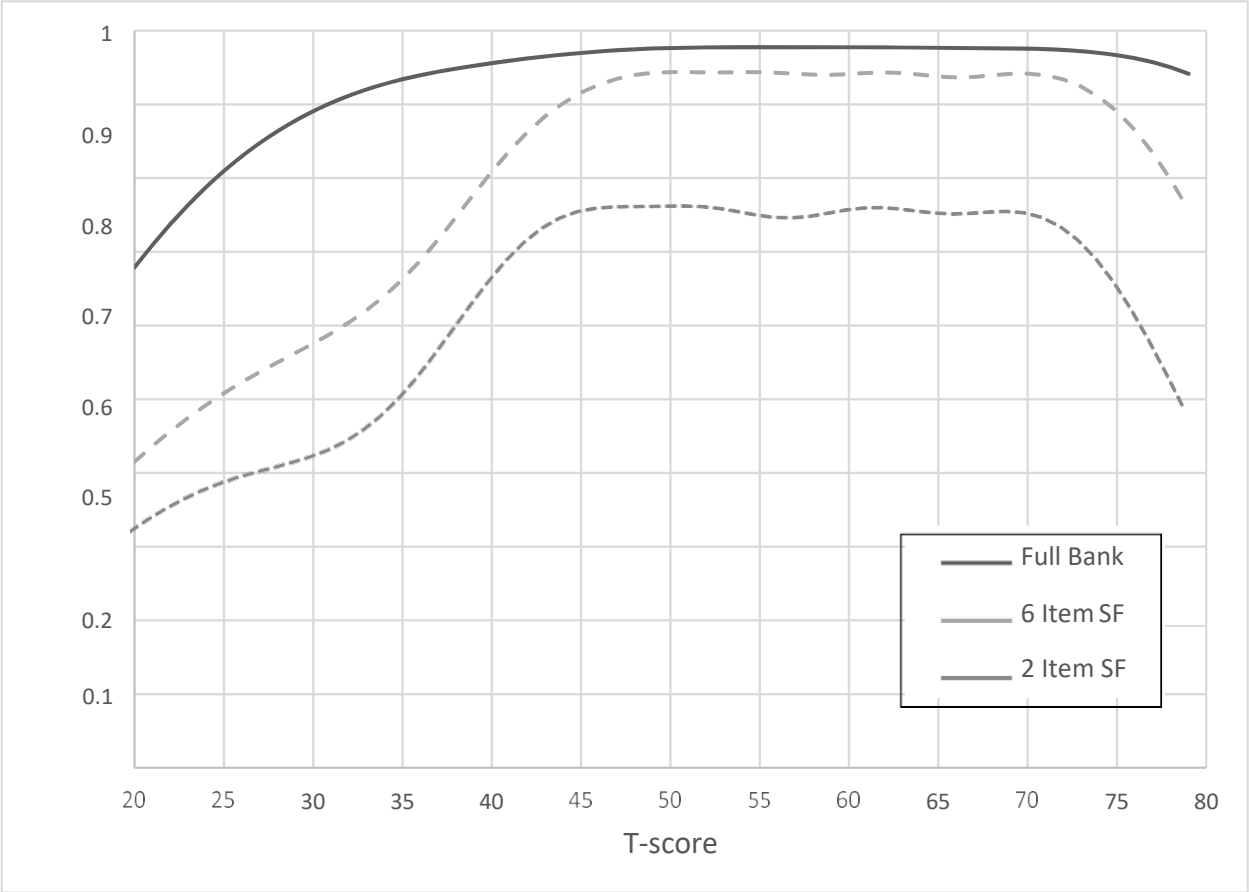
The patient advisors, expert panel, and participants in the focus groups and cognitive interviews all made important contributions to the development of the PRSE and PC item banks. Expert panel members generated definitions, including subdomains for each construct, and selected the format of the items. Focus group discussions of PRSE primarily centered on the strategies participants use to deal with chronic pain, and most participants were reluctant to discuss catastrophizing because they felt it is not helpful to them.

Figure 6. Information, Reliability, and Histogram of Participant PAS Scores on the T-Scale Metric



Abbreviations: PAS, Pain Appraisal Scale; SF, short form.
Histogram shows the frequency of participant PAS scores on the T-score metric. The full item bank and 6-item short form lines show score reliability across the continuum of PAS T-scores for both classical and IRT frameworks.

Figure 7. Reliability of Full Bank, 6-Item SF, and 2-Item SF PAS Scores Along the T-Scale Metric



Abbreviations: PAS, Pain Appraisal Scale; SF, short form.
Scores above 0.9 are considered reliable enough for individual comparisons, as are those above 0.8 for group comparisons. A T-score of 50 is the mean of the calibration sample. For the full bank, reliability was 0.9 or higher between 31 and 80 and 0.8 or higher between 25 and 80. Similarly, for the 6-item short form, reliability was 0.9 or higher between 44 and 74 and 0.8 or higher between 40 and 78 on the T-score metric. Reliability of the 2-item short form never reached acceptable levels for either group or individual comparisons.

It is possible that those who agree to participate in focus groups have found a way to live well despite their chronic pain and are unlikely to be high on PC. Cognitive interviews substantially improved the quality of both the PRSE and PC items and resulted in numerous modifications, especially for PC.

Finally, the term “pain catastrophizing” was considered a very negative term by both the patient advisors and focus group participants, and the use of the scores in clinical care worrisome. As a result, we changed the name of the scale to “Pain Appraisal Scale.” While changing the name does not address worries about how physicians would use the score, it is a

first step toward a less negative perception by both health care providers and patients. We also developed guidance, with feedback from individuals with chronic pain, on ways to utilize the scale for clinical decision making in ways that minimize stigma.

The instruments developed make an important scientific contribution. The rigorous methodology resulted in instruments with excellent psychometric properties and added the convenient and brief options for administration, such as the short forms and the ability to administer by computerized adaptive testing to lower respondent burden (eg, 6 items for the PAS short form versus 13 for the legacy PCS and 6 items for the PRSE short form versus 10 for the PSEQ). The convenience and flexibility in conjunction with sound psychometric properties increase the likelihood that these constructs will be measured in routine clinical care, leading to better treatment outcomes for patients.

Uptake of Study Results

The study resulted in 2 new instruments that offer a psychometrically sound, patient-centered, flexible, dynamic, and convenient way of measuring clinically relevant aspects of chronic pain. The instruments can be (and have been) readily incorporated into research studies where contributions of PC and PRSE to treatment outcomes can be examined. Clinicians and researchers will be able to use the instruments to compare patient scores to normative values; to increase understanding of the patients' perspectives and provide some guidance to interactions that make use of patients' responses to questions characterizing these important constructs; to track changes over time, which will help clinicians determine if treatments that target these domains are effective; and to use initial assessments for better patient–treatment matching (eg, patients with high catastrophizing can be matched to treatments that effectively target catastrophizing). To be incorporated into clinical practice, the scores will need to be validated and actionable cutoffs developed and examined. For now, we have incorporated items from the PAS into the clinical battery administered to all patients considered for spine fusion. Based on previous research, we are using a difference of a half SD as a clinical cutoff and will examine its utility once more data are collected.[76]

Study Limitations

The study could not include people with all types of painful conditions. It is possible that if people with different conditions not well represented in our samples (eg, trigeminal neuralgia, carpal tunnel disease, cancer-related pain) were included, they would have provided different perspectives. As is true in most research studies, people who agree to participate in research are likely different from those who do not participate. Especially relevant in this context is that people with moderate-to-high depressive symptoms are less likely to participate in research[77] and PC is associated with depression.[78] The focus group sample involved did not include a sufficient number of participants at the high end of PC, limiting the usefulness of feedback related to catastrophizing. Future research may benefit from including in focus groups and cognitive interviews feedback by people living with chronic pain whose scores on PC are moderate to severe. Similarly, this study included relatively few people from minority backgrounds, people with low levels of education, or young adults (due to much higher prevalence of chronic pain in older people). Consequently, the validity of the measures in samples with very high levels of minority participants, younger age, or low education levels needs to be examined.

The measures we developed can be readily used for epidemiological and research studies to examine predictors and correlates of important health outcomes. The instruments developed will inform clinical practice after additional needed work to validate the scores and develop clinically actionable cutoffs. Unfortunately, the project period and level of funding did not allow for this important step in instrument development. In clinical practice, once the clinically meaningful cutoffs are established, health care providers will be able to identify patients who are too high on PC and/or too low on self-efficacy and refer them for treatment, especially before more invasive treatments (such as operations) that often do not result in desirable improvements in the presence of maladaptive coping strategies, such as PC.

Future Research

The rigorous methodology used in this study to develop the 2 item banks provides initial evidence of validity, but 1 study is never sufficient—it is important to conduct more extensive

validation studies, especially in different samples and patient populations. The instruments are generalizable to all people living with chronic pain, as patients with many different types of chronic pain were included in the development of the measures. Perspectives of caregivers of people with chronic pain on both catastrophizing and self-efficacy domains may also be informative and important for better understanding these constructs. Developing parallel instruments for caregivers would be a useful direction for future research. While the developers took considerable care to reduce the potential stigma related to the interpretation of PC, a study of clinicians' understanding of the PAS scores and of the guidance provided would be useful to make sure that the guidance provided is effective.

In addition, this study funded measure development, but the funding level and duration of the project was not sufficient to estimate clinically meaningful cutoffs; that is, what score suggests that patients should be referred for further evaluation or treatment to reduce PC or improve pain self-efficacy? Various studies have shown that there are effective interventions to target catastrophizing and self-efficacy in managing pain.[79,80] If the cutoff is too low, the patients who do not need additional treatment will be referred for intervention. If the cutoff is too high, the patients who need the intervention will not receive it, potentially resulting in worse outcomes in the long run. This requires a study of the scores that were collected in the context of expected change (eg, an intervention aimed at reducing PC or increasing PRSE) so that the difference between the scores of those who report better health outcomes and those who report no change or worse health outcomes could be examined. Until this work is completed, a half SD (ie, 5 points)[76,81] can be used as an estimate of clinically important change. Another important step in measure development is to evaluate the equivalence between paper and electronic modes of administration. Because of time, sample size, and budget limitations, we could not measure this equivalence in the current project, but a considerable amount of work has compared scores from different modes of administration,[82-84] and it is reasonable to assume that the different modes of administration would provide similar scores. Future research may be useful to provide evidence for or against this assumption, facilitate interpretability of scores, and maximize the clinical utility of the measures. Finally, contextual factors can influence an individual's cognitive and behavioral responses to pain. That is, for

some, their responses are more trait-like (they generally respond this way across contexts), and for others, their responses are more state-like (they are responding in a way that is situationally or contextually specific). Research to evaluate the factors that influence the domains assessed here requires longitudinal studies that assess both the domains and contextual factors of interest that change over time. This would be an important future direction of this research program.

CONCLUSIONS

New PRSE and pain appraisal item banks provide psychometrically sound and precise scores and are freely available to be administered by computerized adaptive testing or by short forms. The flexible administration options of these IRT-based instruments reduce respondent burden, increase the likelihood that these constructs will be measured in routine clinical care, and facilitate future research, including comparative effectiveness studies for management or treatment of chronic pain. Following patient recommendations that the term “pain catastrophizing” not be used, we changed the name of the measure to the Pain Appraisal Scale and provided guidance to clinicians on how to interpret the scores in ways that make stigmatizing individuals with chronic pain less likely. The short form and the full item bank, as well as the scoring instructions, will be available publicly and free of charge at uwcorr.washington.edu.

REFERENCES

1. Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: results of an Internet-based survey. *J Pain.* 2010;11(11):1230-1239.

2. 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.

3. Fine PG. Long-term consequences of chronic pain: mounting evidence for pain as a neurological disease and parallels with other chronic disease states. *Pain Med.* 2011;12(7):996-1004.

4. Gerrits M, Vogelzangs N, van Oppen P, van Marwijk HW, van der Horst H, Penninx BW. Impact of pain on the course of depressive and anxiety disorders. *Pain.* 2012;153(2):429-436.

5. Karoly P, Ruehlman LS, Okun MA. Psychosocial and demographic correlates of employment vs disability status in a national community sample of adults with chronic pain: toward a psychology of pain presenteeism. *Pain Med.* 2013;14(11):1698-1707.

6. Pizzi L, Carter CT, Howell JB, Vallow SM, Crawford AG, Frank ED. Work loss, healthcare utilization, and costs among US employees with chronic pain. *Dis Manag Health Out.* 2005;13(3):201-208.

7. Institute of Medicine (US) Committee on Advancing Pain Research Care, and Education. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research.* National Academies Press; 2011.

8. Jarvik JG, Hollingworth W, Heagerty PJ, Haynor DR, Boyko EJ, Deyo RA. Three-year incidence of low back pain in an initially asymptomatic cohort: clinical and imaging risk factors. *Spine.* 2005;30(13):1541-1548, discussion 1549.

9. Carragee EJ, Alamin TF, Miller JL, Carragee JM. Discographic, MRI and psychosocial determinants of low back pain disability and remission: a prospective study in subjects with benign persistent back pain. *Spine J.* 2005;5(1):24-35.

10. Burns JW, Glenn B, Bruehl S, Harden RN, Lofland K. Cognitive factors influence outcome following multidisciplinary chronic pain treatment: a replication and extension of a cross-lagged panel analysis. *Behav Res Ther.* 2003;41(10):1163-1182.

11. Flor H, Turk DC. *Chronic Pain: an Integrated Biobehavioral Approach*. IASP Press; 2011.

12. Perry EV, Francis AJ. Self-efficacy, pain-related fear, and disability in a heterogeneous pain sample. *Pain Manag Nurs.* 2013;14(4):25.

13. Keefe FJ, Lefebvre JC, Maixner W, Salley AN Jr, Caldwell DS. Self-efficacy for arthritis pain: relationship to perception of thermal laboratory pain stimuli. *Arthritis Care Res.* 1997;10(3):177-184.

14. Bandura A, O'Leary A, Taylor CB, Gauthier J, Gossard D. Perceived self-efficacy and pain control: opioid and nonopioid mechanisms. *J Pers Soc Psychol.* 1987;53(3):563-571.

15. Bair M, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. *Arch Intern Med.* 2003;163(20):2433-2445.

16. Abramson LY, Alloy LB, Metalsky GI. Hopelessness depression—a theory-based subtype of depression. *Psychol Rev.* 1989;96(2):358-372.

17. Alloy L, Abramson LY, Whitehouse WG, et al. Depressogenic cognitive styles: predictive validity, information processing and personality characteristics, and developmental origins. *Behav Res Ther.* 1999;37(6):503-531.

18. Seligman ME. Depression and learned helplessness. In: Friedman RJ, Katz MM, ed. *The Psychology of Depression; Contemporary Theory and Research*. Winston & Sons, distributed by Halsted Press Division; 1974:318.

19. Arnow BA, Blasey CM, Constantino MJ, et al. Catastrophizing, depression and pain-related disability. *Gen Hosp Psychiatry.* 2011;33(2):150-156.

20. Angst F, Gantenbein AR, Lehmann S, et al. Multidimensional associative factors for improvement in pain, function, and working capacity after rehabilitation of whiplash associated disorder: a prognostic, prospective outcome study. *BMC Musculoskelet Disord.* 2014;15(1):130.

21. Domenech J, Sanchis-Alfonso V, Espejo B. Changes in catastrophizing and kinesiophobia are predictive of changes in disability and pain after treatment in patients with anterior knee pain. *Knee Surg Sports Traumatol Arthrosc.* 2014;22(10):2295-300.

22. Hoffart C, Anderson R, Wallace D. A155: development of an intensive interdisciplinary pediatric pain rehabilitation program: improving pain, functioning, and psychological outcomes. *Arthritis Rheumatol.* 2014;66(Suppl 11):S201.

23. Wong WS, Lam HM, Chow YF, et al. The effects of anxiety sensitivity, pain hypervigilance, and pain catastrophizing on quality of life outcomes of patients with chronic pain: a preliminary, cross-sectional analysis. *Qual Life Res.* 2014;23(8):2333-2341.

24. Wertli MM, Eugster R, Held U, Steurer J, Kofmehl R, Weiser S. Catastrophizing—a prognostic factor for outcome in patients with low back pain—a systematic review. *Spine J.* 2014; 14(11):2639-2657.

25. Schiphorst Preuper H, Geertzen JH, van Wijhe M, et al. Do analgesics improve functioning in patients with chronic low back pain? An explorative triple-blinded RCT. *Eur Spine J.* 2014;23(4):800-806.

26. Kim HJ, Kim SC, Kang KT, Chang BS, Lee CK, Yeom JS. Influence of educational attainment on pain intensity and disability in patients with lumbar spinal stenosis: mediation effect of pain catastrophizing. *Spine.* 2014; 39(10):E637-E644.

27. Wideman TH, Finan PH, Edwards RR, et al. Increased sensitivity to physical activity among individuals with knee osteoarthritis: relation to pain outcomes, psychological factors, and responses to quantitative sensory testing. *Pain.* 2014;155(4):703-711.

28. George SZ, Parr JJ, Wallace MR, et al. Biopsychosocial influence on exercise-induced injury: genetic and psychological combinations are predictive of shoulder pain phenotypes. *J Pain.* 2014;15(1):68-80.

29. Carey ET, Martin CE, Siedhoff MT, Bair ED, As-Sanie S. Biopsychosocial correlates of persistent postsurgical pain in women with endometriosis. *Int J Gynaecol Obstet.* 2014;124(2):169-173.

30. Nicholas MK. The Pain Self-Efficacy Questionnaire: taking pain into account. *Eur J Pain.* 2007;11(2):153-163.

31. Anderson KO, Dowds BN, Pelletz RE, Edwards WT, Peeters-Asdourian C. Development and initial validation of a scale to measure self-efficacy beliefs in patients with chronic pain. *Pain.* 1995;63(1):77-84.

32. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis Rheum.* 1989;32(1):37-44.

33. Keefe FJ, Brown GK, Wallston KA, Caldwell DS. Coping with rheumatoid arthritis pain: catastrophizing as a maladaptive strategy. *Pain.* 1989;37(1):51-56.

34. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. *J Behav Med.* 1997;20(6):589-605.

35. Embretson S, Reise SP. *Item Response Theory for Psychologists*. Erlbaum Associates; 2000.

36. Nicholas MK. Self-efficacy and chronic pain. Paper presented at: Annual Conference of the British Psychological Society; 1989; St. Andrews, Scotland.

37. Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess.* 1995;7:524-532.

38. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. *J Behav Med.* 2000;23(4):351-365.

39. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res.* 2007;16(Suppl 1):133-141.

40. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care.* 2000;38(Suppl 9):II28-II42.

41. Cook KF, Choi SW, Crane PK, Deyo RA, Johnson KL, Amtmann D. Letting the CAT out of the bag: comparing computer adaptive tests and an 11-item short form of the Roland-Morris Disability Questionnaire. *Spine (Phila Pa 1976).* 2008;33(12):1378-1383.

42. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.

43. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care.* 2007;45(5 Suppl 1):S12-S21.

44. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007;45(5 Suppl 1):S22-S31.

45. Patient-Reported Outcomes Measurement Information System (PROMIS). PROMIS Adult Profile Instruments: a brief guide to the PROMIS Profile instruments for adult respondents. 2015. http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Adult_Profile_Scoring_Manual.pdf

46. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain.* 2010;150(1):173-182.

47. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res.* 2009;18(7):873-880.

48. Woo A, Lechner B, Fu T, et al. Cut points for mild, moderate, and severe pain among cancer and non-cancer patients: a literature review. *Ann Palliat Med.* 2015;4(4):176-183.

49. Sullivan MJL. The pain catastrophizing scale: user manual. Accessed September 28, 2009. http://sullivan-painresearch.mcgill.ca/pdf/pcs/PCSManual_English.pdf

50. Nicholas M. Self-efficacy and chronic pain. Paper presented at the Annual Conference of the British Psychological Society; 1989; St. Andrews, Scotland.

51. Coughlan GM, Ridout KL, Williams AC, Richardson PH. Attrition from a pain management programme. *Br J Clin Psychol.* 1995;34:471-479.

52. Frost H, Klaber Moffett JA, Moser JS, Fairbank JC. Randomised controlled trial for evaluation of fitness programme for patients with chronic low back pain. *BMJ.* 1995;310(6973):151-154.

53. Bennett MI, Smith BH, Torrance N, Potter J. The S-LANSS score for identifying pain of predominantly neuropathic origin: validation for use in clinical and postal research. *J Pain.* 2005;6(3):149-158.

54. Freynhagen R, Baron R, Gockel U, Tolle TR. PainDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin.* 2006;22(10):1911-1920.

55. Askew RL, Cook KF, Keefe FJ, et al. A PROMIS measure of neuropathic pain quality. *Value Health.* 2016;19(5):623-630.

56. Jiang S, Wang C, Weiss DJ. Sample size requirements for estimation of item parameters in the Multidimensional Graded Response Model. *Front Psychol.* 2016;7:109.

57. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381.

58. Samejima F. The graded response model. In: van der Linden WJ, Hambleton RK, eds. *Handbook of Modern Item Response Theory*. Springer; 1996:85-100.

59. Muthén LK, Muthén BO. *Mplus User's Guide*. 7th ed. Muthén & Muthén; 1998-2012.

60. Bentler P. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107(2):238-246.

61. Kim D, De Ayala R, Ferdous A. The comparative performance of conditional independence indices. *Appl Psychol Meas.* 2011;35(6):447-471.

62. *IRTPRO*. Scientific Software International, Inc; 2011.

63. Orlando M, Thissen D. Further investigation of the performance of S-χ^2: an item fit index for use with dichotomous item response theory models.*Appl Psychol Meas.* 2003;27(4):289-298.

64. Choi S, 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.

65. R Core Team. *A Language and Environment for Statistical Computing*. 2017. https://www.R-project.org/

66. Zumbo B. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.

67. Crane P, Gibbons LE, Ocepek-Welikson K, et al. A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Qual Life Res.* 2007;16(Suppl 1):69-84.

68. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient.* 2014;7(1):23-35.

69. Shrout P, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.

70. *Stata Statistical Software: Release 14*. StataCorp LP; 2015.

71. Thissen D. Reliability and measurement precision. In: Wainer H, Dorans NJ, Flaugher R, et al, eds. *Computerized Adaptive Testing: A Primer*. Lawrence Erlbaum Associates Publishers; 2000:159-184.

72. Jackson T, Wang Y, Wang Y, Fan H. Self-efficacy and chronic pain outcomes: a meta-analytic review. *J Pain.* 2014;15(8):800-814.

73. Damush TM, Kroenke K, Bair MJ, et al. Pain self-management training increases self-efficacy, self-management behaviours and pain and depression outcomes. *Eur J Pain.* 2016;20(7):1070-1078.

74. Park SJ, Lee R, Yoon DM, Yoon KB, Kim K, Kim SH. Factors associated with increased risk for pain catastrophizing in patients with chronic neck pain: a retrospective cross-sectional study. *Medicine (Baltimore).* 2016;95(37):e4698.

75. Craner JR, Sperry JA, Evans MM. The relationship between pain catastrophizing and outcomes of a 3-week comprehensive pain rehabilitation program. *Pain Med.* 2016;17(11):2026-2035.

76. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *MedCare.* 2003;41:582-592.

77. Lamers F, Hoogendoorn AW, Smit JH, et al. Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Compr Psychiatry.* 2012;53(1):63-70.

78. Sullivan MJ, D'Eon JL. Relation between catastrophizing and depression in chronic pain patients. *J Abnorm Psychol.* 1990;99(3):260-263.

79. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. *Clin J Pain.* 2012;28(1):14-22.

80. Kristjánsdóttir OB, Fors EA, Eide E, et al. A smartphone-based intervention with diaries and therapist-feedback to reducecatastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. *J Med Internet Res.* 2013;15(1):e5. doi:10.2196/jmir.2249

81. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcome. *J Clin Epidemiol.* 2008;61:102-109.

82. Bjorner J, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware D. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. *Qual Life Res*. 2014;23(1):217-227.

83. Magnus B, Liu Y, He J, et al. Mode effects between computer self-administration and telephone interviewer-administration of the PROMIS Pediatric Measures, Self-and Proxy Report. *Qual Life Res*. 2016; 25(7):1655-65.

84. Stucky BD, Huang W, Edelen MO. The psychometric performance of the PROMIS Smoking Assessment Toolkit: comparisons of real-data computer adaptive tests, short forms, and mode of administration. *Nicotine Tob Res*. 2016;18(3):361-365.

PUBLICATIONS

Journal Publications Published

- Amtmann D, Liljenquist K, Bamer A, et al. Measuring pain catastrophizing and pain-related self-efficacy: expert panels, focus groups and cognitive interviews. *The Patient.* 2017;11(1):107-117.

Under Development

- Amtmann D, Bamer A, Liljenquist K, Bocell F, Jensen M, Turk D. Development and validation of the pain appraisal scale.

- Amtmann D, Liljenquist K, Bamer A, Bocell F, Jensen M, Turk D. Development and validation of the pain related self-efficacy scale.

Presentations

- Amtmann D, Bocell F, Jensen M, Liljenquist K, Turk D. The Pain Related Self-efficacy Scale: a new item bank. Presented at: International Society for Quality of Life Research Annual Meeting; October 2016; Copenhagen, Denmark.

- Amtmann D, Bocell FD, Jensen M, Liljenquist K, Turk D. Development of an item bank to measure pain catastrophizing. Presented at: International Society for Quality of Life Research Annual Meeting; October 2016; Copenhagen, Denmark.

- Amtmann D, Jensen M, Liljenquist K, Salem R, Bocell F, Turk D. Measuring pain-related self-efficacy and pain catastrophizing: results of expert panels, focus groups, and cognitive interviews. Poster presented at: World Congress on Pain; 2016; Yokohama, Japan.

- Amtmann D, Bocell F, Jensen M, Liljenquist K, Turk D. The Pain Related Self-efficacy Scale: a short form. Poster presented at: Society of Behavior Medicine; 2017; San Diego, CA.

- Amtmann D, Bamer AM, Liljenquist K, Jensen MP, Turk D. Short forms to measure pain catastrophizing and pain related self-efficacy. Accepted as an oral presentation at: International Society for Quality of Life Research Annual Meeting; October 2017; Philadelphia, PA.

APPENDICES

Appendix A. Sum Score to T-Score Conversion for PRSE Full Item Bank and 6-Item Short Form

PRSE Full Item Bank	
Sum Score	T-score
6	24.5
7	28.5
8	31.3
9	33.8
10	35.9
11	37.8
12	39.6
13	41.4
14	43
15	44.7
16	46.3
17	47.9
18	49.5
19	51.1
20	52.8
21	54.5
22	56.2
23	57.9
24	59.7
25	61.6
26	63.6
27	65.8
28	68.2
29	71.1
30	74.7

PRSE six-item shortform	
Sum Score	T-score
2	28.4
3	34.7
4	39.5
5	44.1
6	48.4
7	52.7
8	57.5
9	62.7
10	69.2

Appendix B. Sum Score to T-Score Conversion for PAS Full Item Bank and 6-Item Short Form

PAS Full Item Bank	
Sum Score	T-score
6	30.8
7	35.3
8	38.7
9	41.6
10	44.1
11	46.2
12	48.1
13	49.7
14	51.2
15	52.6
16	54
17	55.4
18	56.8
19	58.2
20	59.6
21	61.1
22	62.5
23	64
24	65.5
25	67
26	68.7
27	70.5
28	72.5
29	74.9
30	78.1
PAS six-item short form	
Sum Score	T-score
2	34.49
3	39.43
4	44.37
5	49.18
6	53.38
7	57.25
8	61.16
9	65.56
10	71.37

Acknowledgment:

Research reported in this report was [partially] funded through a Patient-Centered Outcomes Research Institute® (PCORI®) Award (#ME- 1403-12550) Further information available at:
https://www.pcori.org/research-results/2014/developing-measures-pain-appraisal-and-pain-related-self-efficacy-people