American Pain Society
RESEARCH
EDUCATION
TREATMENT
ADVOCACY
PUBLISHED BY ELSEVIER

The Journal of Pain, Vol 11, No 11 (November), 2010: pp 1109-1119
Available online at www.sciencedirect.com

PROMIS Pediatric Pain Interference Scale: An Item Response Theory Analysis of the Pediatric Pain Item Bank

James W. Varni,* Brian D. Stucky,† David Thissen,† Esi Morgan DeWitt,‡ Debra E. Irwin,§ Jin-Shei Lai,‖ Karin Yeatts,§ and Darren A. DeWalt¶

**Department of Pediatrics, College of Medicine, Department of Landscape Architecture and Urban Planning, College of Architecture, Texas A&M University, College Station, Texas.*
†Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
‡Department of Pediatrics, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.
§Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
‖Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois.
¶Division of General Medicine and Clinical Epidemiology, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Abstract: An aim of the National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS) initiative is to develop item banks and computerized adaptive tests (CAT) that are applicable across a wide variety of chronic disorders. The PROMIS Pediatric Cooperative Group has concentrated on the development of pediatric self-report item banks for ages 8 through 17 years. The objective of the present study is to describe the Item Response Theory (IRT) analysis of the NIH PROMIS pediatric pain item bank and the measurement properties of the new unidimensional PROMIS Pediatric Pain Interference Scale. Test forms containing pediatric pain items were completed by a total of 3048 respondents. IRT analyses regarding scale dimensionality, item local dependence, and differential item functioning were conducted. A pain item pool was developed to yield scores on a *T*-score scale with a mean of 50 and standard deviation of 10. The recommended 8-item unidimensional short form for the PROMIS Pediatric Pain Interference Scale contains the item set which provides the maximum test information at the mean (50) on the *T*-score metric. A simulated CAT was computed that provides the most information at 5 possible score locations (30, 40, 50, 60, and 70 on the *T*-score metric).

***Perspective:** The present study provides initial calibrations of the NIH PROMIS pediatric pain item bank and the creation of the PROMIS Pediatric Pain Interference Scale. It is anticipated that this new scale will have application in pediatric chronic and recurrent pain.*

***Key words:** Pain, pediatrics, PROMIS, pain interference, Item Response Theory.*

The Patient Reported Outcomes Measurement Information System (PROMIS) is a National Institutes of Health (NIH) Roadmap Initiative, created to advance the assessment of patient-reported outcomes (PRO) in chronic diseases. To achieve this goal, self-report items are evaluated using modern measurement theory (Item Response Theory, IRT) in order to derive assessments that are maximally reliable, valid, and generalizable for individuals falling along the full spectrum of the trait being measured.[1] A primary objective is to develop a group of item banks and computerized adaptive tests across a wide variety of chronic disorders.[29] During the past 5 years, the PROMIS Pediatric Cooperative Group has concentrated on the development of pediatric self-report PRO item banks for ages 8 through 17 years across 5 generic health domains (physical function, pain, fatigue, emotional health, and social health) from the patient perspective, consistent with the larger PROMIS network.[4] It was anticipated that measures of these 5 health domains would be applicable across numerous pediatric chronic health

Received October 14, 2009; Revised February 9, 2010; Accepted February 10, 2010.
Supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1U01AR052181-01. Information on the Patient-Reported Outcomes Measurement Information System (PROMIS) can be found at http://nihroadmap.nih.gov and http://www.nihpromis.org.
Address reprint requests to Dr James W. Varni, College of Architecture, Texas A&M University, 3137 TAMU, College Station, TX 77843-3137. E-mail: jvarni@archmail.tamu.edu
1526-5900/$36.00

doi:10.1016/j.jpain.2010.02.005

conditions, and hence were developed as generic or nondisease-specific scales.

Given the widespread occurrence of chronic and recurrent pain in pediatric populations,[12] particularly in pediatric chronic diseases,[41] an item bank focused on pediatric pain items was an essential component of the PROMIS Pediatric Cooperative Group's efforts. Although the measurement of *pain intensity* using visual analogue scales,[20,40] rating scales,[11,36,39] and pictorial scales[21,22] has received empirical attention in pediatric populations over the past 2 decades as evidenced by recent comprehensive reviews,[5,8,23,26,32,44] the measurement of the *pain interference* construct has received less empirical attention, and consequently was an important focus in the development of the PROMIS pediatric pain item bank.[17,46] For the purposes of this study, the a priori operational definition of "pain interference" was the interference by pain on daily activities during the past 7 days (interference on physical, psychological, and social functioning). At the end of each item stem was the phrase "…when I had pain" to explicitly distinguish the items as pain-specific interference, rather than as generic functioning items.

While other scales have been developed that measure physical activities in pediatric patients, including those which have utilized either Rasch or IRT analyses,[14,48] these scales typically contain generic items (ie, not pain-specific content) or have been used predominantly in specific populations.[27] In contrast, the Child Activity Limitations Interview (CALI) was designed to assess functional impairment in activities of daily living secondary to pediatric chronic and recurrent pain.[27,28] However, the CALI and CALI-21 were developed utilizing Classical Test Theory rather than IRT. Early research with the CALI-21 demonstrates that it has 2 factors described as representing "Active and Routine activities"; such detailed factor analysis was an advance over earlier pain measures.[26,27] Additional analyses of data from the CALI-21 would be helpful to investigate the possibility of local dependence and gender DIF. The larger sample sizes and IRT analytic techniques used in PROMIS item development permit these more detailed levels of psychometric scrutiny.

Thus, the majority of pediatric pain functional impairment scales, consistent with other pediatric assessment instruments, have utilized Classical Test Theory and have rarely taken advantage of IRT analysis in the scale development process.[15,19] By using IRT analysis, the resulting item bank can be the basis of a more customizable measure for meeting a researcher's or clinician's needs. Depending on the desired level of precision, the evaluator can then select the number of items to administer and obtain scores on the same metric as all other users of this item bank.[10]

Consequently, the objective of the present study is to address this measurement gap in the pediatric pain literature by describing the IRT analysis of the PROMIS pediatric pain item bank and the measurement properties of the new PROMIS Pediatric Pain Interference Scale, including investigations of scale dimensionality and sources of local dependence and differential item functioning.

Methods

Sampling Plan

Participants were recruited in hospital-based outpatient general pediatrics and subspecialty clinics and in public school settings between January 2007 and May 2008 in North Carolina and Texas. This sample was derived to include a broad range of experiences from children that were healthy and children with chronic illnesses. Children completed questionnaires that included items across several domains of health including physical function, pain, fatigue, emotional distress, and social health. North Carolina and Texas were chosen as recruitment sites because of the diversity of cultural experience and population characteristics that existed in those areas.

To be eligible to participate in the large-scale testing survey, subjects were required to meet the following inclusion criteria: between the ages of 8 to 17 years old; able to speak and read English; and able to see and interact with a computer screen, keyboard, and mouse. They provided informed assent before study entry and a parent or guardian provided informed consent. Both the informed assent and the informed consent were administered in English so parents were also required to read and speak English. Parent reports were used to determine whether or not the child had any limitations (eg, physical or cognitive) that would make it too difficult to complete a computer administered survey.

Potential clinic participants were identified through a variety of methods such as a review of pediatric clinic appointment rosters or while in the clinic waiting rooms according to protocols approved by the institutional review boards (IRBs) of The Children's Hospital at Scott and White (S&W) in Texas, the University of North Carolina (UNC), and Duke University pediatrics clinics. The UNC, Duke, and S&W general pediatric clinics were representative of health issues for which children have physician office visits (eg, well child visits, acute illnesses as well as some chronic illnesses). The specialty clinics included Pulmonology, Allergy, Gastroenterology, Rheumatology, Nephrology, Obesity, and Endocrinology and primarily saw children with more serious chronic illnesses. Children with asthma were over sampled during recruitment because asthma-specific items were tested. It was anticipated that pediatric patients in Rheumatology, Gastroenterology, and General Pediatrics would manifest recurrent or chronic pain based on previous literature.[16,39,45]

School-based participants were recruited through the Chapel Hill-Carrboro (North Carolina) Public School System including elementary after school programs as well as required middle and high school health classes. An informational packet about the study, including informed consent documents and a sociodemographic form, was mailed to all of the parents with children enrolled in the health classes to complete and return to the school.

Parents signed an informed consent document and children signed an informed assent document that outlined the following: purpose of the study, participation

requirements, potential benefits and risks of participation and measures implemented to protect participant privacy. Child participants received a $10 gift card in return for their time and effort. The study protocols were approved by the institutional review boards at each institution.

To limit respondent burden, the number of items administered to any respondent was limited to no more than 76 items of the entire pool of 293 PROMIS items and the legacy questionnaires. The items were written to accommodate low literacy levels.[8] Based on the experience of the research team, it was estimated that the younger children would be able to complete the survey in about 25 minutes and the adolescents in about 15 minutes. The 293 PROMIS items were divided among 4 testing forms and 1 additional form containing only general "legacy" scales (see Table 1). The legacy scales were administered on a separate test form to characterize the population, but were not administered together with the PROMIS items. As such, this data collection does not allow us to compare individual responses on the legacy instruments with responses to the PROMIS items. Some items were administered on more than 1 form. The inclusion of overlapping items on different forms permits an evaluation of the associations between domains. Each PROMIS item from non-disease specific banks was administered to at least 754 respondents across 4 forms.

Children without asthma were assigned sequentially to 1 of 5 forms (4 forms with PROMIS items and a few legacy general items and 1 form containing only legacy scales). This sampling plan was developed for collecting responses to the candidate items from the targeted PROMIS domains and was designed to accommodate multiple objectives: (1) confirm the factor structure of the domains; (2) evaluate items for local dependence (LD) and differential item functioning (DIF); and (3) calibrate the items for each domain using Item Response Theory.

We developed the PROMIS Pediatric item banks using a strategic item generation methodology adopted by the PROMIS Network.[6] Six phases of item development were implemented: identification of existing items, item classification and selection, item review and revision, focus group input on domain coverage, cognitive interviews with individual items, and final revision before field testing. Identification of items refers to the systematic search for existing items in currently available pediatric scales. This was utilized to identify an initial item pool of over 3345 items. Expert item review and revision was conducted by trained professionals who reviewed the wording of each item and revised as appropriate for conventions adopted by the PROMIS network.[4,6] Focus groups were used to confirm domain definitions, and to identify new areas of item development for future PROMIS item banks.[46] Cognitive interviews were used to examine and refine wording of individual items.[17] The pediatric items were written in the past tense with a 7-day recall period and most utilized a standard set of response options.[17] Items successfully screened through the cognitive interview process were sent to field testing. The final item set contained 293 items across the 6 domains (Physical Function, Emotional Distress, Social Role Relationships, Fatigue, Pain, and Asthma).[17]

Most pain items had a 7-day recall period and used standardized 5-point response options (never, almost never, sometimes, often, almost always). Occasionally, participants responded to items on an 11-point pain intensity scale (0 through 10), or a response scale in reference to the number of days (0 through 7 days). A complete list of items may be found in the Tables and Appendix.

Statistical and Psychometric Methods

The PROMIS methods used for the psychometric evaluation and calibration of the pain items have been previously described.[29] First, traditional descriptive statistics were computed to verify that there were no empty (zero frequency) response categories for any item, and as preliminary checks on the validity of the data. Included in these checks were tables of marginal frequencies of item responses and the correlations of item scores with the total summed score.

The IRT model that is used here for item analysis and scoring is based on the assumption that responses to the items indicate individual differences on a single underlying, or latent, variable (here, pain interference). To confirm the validity of that assumption, the second phase of the data analysis used confirmatory factor analysis (CFA) of the interitem polychoric correlation matrix to ensure that the latent variable underlying the item responses was unidimensional. These analyses were performed using the DWLS algorithm as implemented in the software LISREL[18]; this approach takes into account the categorical nature of the item responses, in a way that corresponds with the IRT model that is subsequently used.

In addition to a single-factor model, fitting additional factors, and/or error covariances, served as indications of local dependence (LD) for pairs or small numbers of items. LD is a term that describes any violation of the local independence assumption of unidimensional IRT[15]; that assumption is that all of the observed covariation among the item responses is accounted for by the single latent variable being measured by the scale. If a pair of items are more correlated than is accounted for by the latent variable underlying the responses to all of the (other) items, that is an indication that responses to those items behave to some extent as though the same question had been asked twice (which would produce perfect LD). If an additional factor appears for a small subset of items, that means those items as a cluster measure some other aspect of individual difference variation, and the data analyst must decide whether to measure that additional aspect separately, or set it aside. In the case of the construction of the pain interference scale, items were set aside from subsets that exhibited LD.

Third, after conducting CFA, item sets determined to be unidimensional were next calibrated by fitting Samejima's Graded Response Model (GRM[30]) using the

Table 1. Survey Participants: Demographic and Background Information

	Form 1 N = 759 (%)	*Form 2 N = 770 (%)*	*Form 3 N = 754 (%)*	*Form 4 N = 765 (%)*
Child's gender				
Male	382 (50.3)	351 (45.6)	355 (47.1)	382 (49.9)
Female	377 (49.7)	419 (54.4)	399 (52.9)	383 (50.1)
Missing	0	0	0	0
Child's age (y)				
8-12	446 (58.8)	441 (56.4)	303 (40.2)	426 (55.7)
13-17	312 (41.1)	326 (42.3)	451 (59.8)	337 (44.0)
Missing	1 (0.1)	3 (0.3)	0	2 (0.3)
Child's race				
White	457 (60.2)	452 (58.7)	457 (60.6)	462 (60.4)
Black or African-American	154 (20.2)	168 (21.8)	172 (22.8)	150 (19.6)
American Indian/Alaska Native	5 (0.6)	10 (1.3)	7 (0.9)	10 (1.3)
Asian	12 (1.6)	13 (1.7)	6 (0.8)	10 (1.3)
Native Hawaiian/Pacific Island	0	1 (0.1)	2 (0.3)	2 (0.3)
Other	58 (7.6)	50 (6.5)	58 (7.7)	64 (8.4)
Multiple races	47 (6.2)	54 (7.0)	27 (3.6)	43 (5.6)
Missing	26 (3.4)	22 (2.9)	25 (3.3)	24 (31.)
Child's ethnicity				
Non-Hispanic	614 (80.9)	641 (83.2)	617 (81.8)	619 (80.9)
Hispanic	141 (18.6)	121 (15.7)	131 (17.4)	141 (18.4)
Missing	4 (0.5)	8 (1.1)	6 (0.8)	5 (0.7)
Child's chronic conditions - 6 mo				
No	600 (79.0)	580 (75.3)	569 (75.5)	592 (77.4)
Yes = 1 Chronic condition	113 (14.9)	145 (18.8)	134 (17.8)	120 (15.7)
Yes ≥2 Chronic conditions	44 (5.8)	42 (5.5)	46 (6.1)	49 (6.4)
Missing	2 (0.3)	3 (0.4)	5 (0.6)	4 (0.5)
Most common conditions diagnosed or treated within 6 mo before enrollment*				
Asthma	19 (2.5)	18 (2.3)	22 (2.9)	23 (3.0)
ADD/ADHD	27 (3.6)	39 (5.1)	32 (4.2)	36 (4.7)
Arthritis	23 (3.0)	28 (3.6)	25 (3.3)	24 (3.1)
Gastrointestinal disorders	24 (3.2)	21 (2.7)	15 (2.0)	15 (2.0)
Mental disorders	12 (1.6)	18 (2.3)	13 (1.7)	12 (1.6)
Immune disorders	11 (1.5)	18 (2.3)	18 (2.4)	16 (2.1)
Guardian's relationship to child				
Parent	696 (91.7)	717 (93.1)	695 (92.2)	708 (92.6)
Grandparent	32 (4.2)	30 (3.9)	32 (4.2)	43 (5.6)
Guardian or other	31 (4.1)	21 (2.7)	26 (3.5)	13 (1.7)
Missing	0	2 (0.3)	1 (0.1)	1 (0.1)
Guardian's education Level				
≤8th grade	12 (1.6)	16 (2.3)	13 (1.8)	16 (2.1)
Some high school	39 (5.1)	34 (4.4)	54 (7.2)	55 (7.2)
High school degree/GED	151 (19.9)	153 (19.7)	163 (21.6)	159 (20.8)
Some college/technical degree	255 (33.6)	245 (31.8)	251 (33.2)	260 (34.0)
College degree	179 (23.6)	214 (27.8)	183 (24.3)	180 (23.5)
Advanced degree	121 (15.9)	105 (13.6)	86 (11.4)	95 (12.4)
Missing	2 (0.3)	3 (0.4)	4 (0.5)	0
Data collection site				
Schools – NC	57 (7.5)	57 (7.4)	49 (6.5)	51 (6.7)
Clinics - NC	349 (46.0)	350 (45.5)	343 (45.5)	351 (45.9)
Clinics – TX	353 (46.5)	363 (47.1)	362 (48.0)	363 (47.4)

*Parents reported more than 1 condition for some children; there were many other conditions reported in lower frequency (<1.5%) than the conditions listed.

software Multilog[7] (the GRM has been selected for other PROMIS scales[29]). Calibration, as that term is used in IRT analysis, refers to the estimation of a set of parameters for each item that characterize the relation of the item responses with the latent variable (here, pain interference) being measured. For each item, the GRM estimates a *slope* or *discrimination* parameter (*a*), reflecting the degree of association of the item responses with the latent construct being measured, and 4 threshold parameters (b_k) (for 5 response option items; or 7 thresholds for 8 response options) that indicate the level of pain interference at which a response in a given category or higher becomes probable. In item analysis, the item parameters are used to compute an information function for each item. The statistical information provided by each item reflects the degree to which the item contributes to the

precision of measurement of the scale in an additive way: If one has 5 items that each have information equal to 2.0 at some value of the latent variable, then the information value for the 5-item scale is 10. The variance of measurement of the scale at that value of the latent variable is the inverse of the information, so that would be 0.1 in standard-score units. Classical Test Theory is based on algebra that assumes that the variance of measurement has the same value for all scores; in the classical theory, for scores in standard units, reliability is 1 minus the error variance, so for error variance 0.1, reliability is 0.9. IRT more realistically represents error variance as a quantity that varies as a function of the latent variable; error variance is small for levels of the latent variable where the items provide information, and larger elsewhere. Nevertheless, because 0.9 has often been considered a useful value of reliability, for IRT analysis 10 is a useful value of information. The item parameters computed during calibration can be used to identify the levels of the latent variable for which the items provide information, and items can be selected until aggregate information exceeds some desired value, such as 10.

The item parameters obtained in the calibration phase are also used to compute IRT scale scores, either for a summed score for a fixed set of items, or for response patterns for any arbitrary subset of items in a pool. IRT scale scores are estimates of the value of the latent variable (pain interference, here) for which the observed item responses are likely. As a consequence of the assumption of unidimensionality, the IRT scale scores are on a single continuum, and comparable, even if respondents are measured using different subsets of items. This aspect of IRT represents one of its most important advantages over the classical theory, which can provide comparable scores only for a fixed set of items. One use of this feature of IRT is to assemble alternate short forms that yield comparable scores. Another more extreme use is to administer CATs, which adaptively select a customized set of items for each respondent, to provide maximum information at the level of that person. When using a CAT, each person may respond to a different set of questions; nevertheless, their IRT scale scores are quantitatively comparable.

The goodness of fit of the IRT model to the data was examined using the $S\text{-}X^2$ statistic[24,25] (generalized by Bjorner et al[3]). As a goodness-of-fit statistic, a nonsignificant $S\text{-}X^2$ value suggests adequate fit of the model to the data.

Fourth, for item selection for the final pool, differential item functioning (DIF) was investigated between males and females using the IRT-LR DIF detection procedure[35] as implemented in the software IRTLRDIF.[33] In this case, DIF indicates that the relation of item responses with the latent variable differs between boys and girls. Such a difference suggests that some other factor, related to gender but different from the construct being measured, influences item responses, which is a violation of the assumption of unidimensionality. Here again, a nonsignificant χ^2 indicates a lack of DIF. Because DIF detection involves a large number of tests of significance, the Benjamini-Hochberg procedure[2,47] was used to control for multiple comparisons. In addition to χ^2 statistics, graphical methods, as suggested by Steinberg and Thissen,[31] were used to evaluate the magnitude of effect sizes when significant DIF was detected. After the item pool was selected, we also evaluated DIF between younger (ages 8 through11) and older (ages 12 through17) respondents; because we do not expect the scale to be used for the purpose of comparison of pain interference among children classified by age, we did not include these results among the item selection criteria, but the results are reported here.

Finally, though IRT scale scores may be computed from either item response patterns or summed scores, we expect scale scores for summed scores to be used more often. Thus, the Appendix Table A1 provides a translation table to be used for this purpose.[34] The IRT scale scores reported here use the North Carolina sample as the reference group.

Results

Test forms containing PROMIS pediatric pain items were completed by a total of 3,048 respondents. The sample was about 52% female and 58% of the children were between the ages 8 to 12 years old. Sixty percent were Caucasian, 21% black, 6% multi-racial, and 13% other races (Asian/Pacific Islanders, Native Americans, and other races). Eighteen percent of the sample was of Hispanic ethnicity. The vast majority of the adults providing informed consent for the children were parents of the child (92%) or grandparents (4%). The educational attainment of these parents or guardians ranged from less than high school (8%) to advanced degree (13%), with 25% reporting a college degree, 33% some college, and 21% a high school diploma. Approximately 23% of the children participating in the survey had a chronic illness diagnosis during the past 6 months. Participant characteristics are summarized in Table 1.

There were adequate numbers of pain items on each of the 4 forms to permit factor analysis of each. Tables 2 and 3 provide the factor loadings from models that fit well. The models indicate that the items on separate forms are generally unidimensional, though with some evidence of local dependence. Local dependence, or nuisance multidimensionality, is modeled in Forms 1, 2, and 4 (Table 2) by error covariances (in this case between 2 items, or "doublets"). Form 3 (Table 3) contains 3 items (a "triplet") pertaining to the physical limitations caused by pain, and as such was modeled as a second factor (with a correlation between the general pain interference factor and the "difficulty moving" subfactor). Indicators of goodness of fit suggest all 4 models fit the data well, using indices suggested by Reeve et al.[29]: For Form 1 (Table 2), $\chi^2(7) = 9$, CFI = 1.00, TLI = 1.00, RMSEA = 0.02; Form 2 (Table 2), $\chi^2(12) = 10$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00; Form 3 (Table 3), $\chi^2(10) = 8$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00; and Form 4 (Table 2), $\chi^2(13) = 21$, CFI = 1.00, TLI = 1.00, RMSEA = 0.03.

The local dependence in Forms 1 through 4 occurs primarily because items share similar wording, or have shared content that differs from the content of the

Table 2. Factor Loadings and Error Covariances for Pain Interference Items on Forms 1, 2, and 4

	Pain	Doublet Error Covariances	
Form 1 Items			
How many days were you free of pain (no pain)?	0.40	0.14	
How bad is your pain right now?	0.52		
It was hard for me to think when I had pain.	0.85		
It was hard for me to ride in a car when I had pain.	0.72		0.13
It was hard for me to walk one block when I had pain.	0.71		
I felt grumpy when I had pain.	0.61		
Form 2 Items			
How bad was your worst pain?	0.55	0.33	
How bad was your pain on average?	0.51		
I felt sad when I had pain.	0.72		
I had trouble doing schoolwork when I had pain.	0.72		0.15
I had trouble watching TV when I had pain.	0.62		
It was hard for me to run when I had pain.	0.76		
I had trouble sleeping when I had pain.	0.80		
Form 4 Items			
It was hard to get along with other people when I had pain.	0.62	0.21	
I wanted to be alone when I had pain.	0.54		
I hurt a lot.	0.60		
It was hard for me to remember things when I had pain.	0.63		
It was hard to do sports or exercise when I had pain.	0.67		
I missed school when I had pain.	0.56		
It was hard to stay standing when I had pain.	0.80		

NOTE. All factor loadings and error covariances are significantly different from zero at $P < .05$.

Table 3. Factor Loadings and Error Covariances for Pain Interference Items on Form 3

Form 3 Items	Pain	Difficulty Moving
I had trouble moving around when I had pain.	0.30	0.61
It was hard to have fun when I had pain.	0.40	0.49
It was hard for me to walk up a flight of stairs when I had pain.	0.50	0.34
How many days did you have pain?	0.67	
It was hard for me to pay attention when I had pain.	0.81	
I hurt all over my body.	0.73	
I felt angry when I had pain.	0.67	

NOTE. All factor loadings are significantly different from zero at $P < .05$. The correlation between the general pain interference and difficulty moving factors is $r = 0.79$.

scale's other items. As an example of shared item content, Form 3 contains a "triplet," or 3 items with responses that are more related than expected given the items' relationship with the pain interference dimension. In this case the triplet measures physical limitations caused by pain. In other instances, local dependence may result from shared content or the response scale used. Form 2 contains 2 items measuring pain intensity on a 0 to 10 scale. In addition to being similarly worded and assessed on a unique response scale, the items are measuring pain intensity, whereas the scale's other items assess interference on daily activities caused by pain. To ensure unidimensionality of the final scales, only 1 item from each doublet or triplet was included in the final item pool.

After the factor analyses, locally independent sets of items from Forms 1 through 4 were calibrated using the GRM. To control for local dependence identified in the item factor analyses, separate item calibrations were completed for each collection of unidimensional items. This process resulted in 2 sets of calibrations for each Form (3 in the case of Form 3). To avoid capitalization on chance, we conservatively selected parameter estimates across calibrations that had the lower estimated slope. Table 4 shows the item parameter estimates, item fit statistics ($S\text{-}X^2$), and DIF statistics (LR χ^2) for the items comprising the final pool (sorted in order of magnitude of slope parameters), and for the items set aside.

The Benjamini-Hochberg correction for multiplicity was used with the fit and DIF statistics. Two items had either significant DIF or lack of fit as indicated by the $S\text{-}X^2$ statistic; however, these items were retained when considered in relation to the relatively good fit of the items comprising the final pool. As indicated in Table 4, there were 15 items set aside. Five were set aside from locally dependent item sets. An additional 5 were set aside due to low discrimination parameters. Interestingly, these items measured pain intensity, and as such discriminate poorly between levels of pain interference. Finally, 4 items were set aside for DIF (both threshold and slope DIF). As an interpretive example of threshold DIF, boys were less willing to endorse the item "It was hard to do sports or exercise when I had pain," after controlling for mean and variance differences between boys and girls. Additionally, slope DIF occurred for the item "I felt grumpy when I had pain," indicating that "feeling grumpy" is a poor indicator of pain interference for boys. The remaining 13 items comprise the final pain item pool.

In the analysis of DIF by age, 5 of the 13 items in the pool exhibited significant DIF. For 3 of those items, the aggregate effect size of the DIF is very small: For the items "It was hard for me to pay attention when I had pain," "I had trouble doing schoolwork when I had pain," and "I felt angry when I had pain," the difference between older and younger children in the expected value of the item response on the 0 to 4 scale is much less than a half point across the entire range of the latent variable pain interference. To a large extent, the tendency is for those 3 items to be slightly more discriminating for older than younger children. For the item "It was hard to remember things when I had pain," younger children tend to give slightly higher responses than older

Table 4. Item Parameters, Fit Indices, and DIF Statistics for the Pain Interference Items

	Item Parameters					$S\text{-}X^2$ Fit Index			LR DIF		
Item Stem	a	b1	b2	b3	b4	χ^2	D.F.	P	χ^2	D.F.	P
Final item pool											
I had trouble sleeping when I had pain.	2.35	−0.23	0.31	1.17	1.69	57	45	.108	13.1	5	.022
It was hard for me to pay attention when I had pain.	2.35	−0.25	0.32	1.33	2.03	37	34	.332	8.9	5	.113
It was hard to stay standing when I had pain.	2.35	−0.18	0.44	1.40	1.97	47	33	.054	1.0	5	.963
It was hard to have fun when I had pain.	2.31	−0.49	0.00	1.02	1.71	56	36	.018	2.2	5	.821
It was hard for me to walk one block when I had pain.	2.14	0.28	0.79	1.50	1.97	33	30	.323	7.4	5	.193
I had trouble doing schoolwork when I had pain.	1.94	−0.23	0.46	1.47	2.16	46	48	.555	5.6	5	.347
It was hard for me to run when I had pain.	1.89	−0.85	−0.25	0.85	1.63	55	47	.198	6.3	5	.278
I hurt all over my body.	1.82	0.49	1.19	2.05	2.72	44	32	.077	4.8	5	.441
I felt angry when I had pain.	1.62	−0.01	0.66	1.56	2.24	48	37	.106	9.6	5	.087
It was hard for me to remember things when I had pain.	1.50	0.29	1.08	2.12	3.55	38	33	.252	5.3	5	.380
I hurt a lot.	1.41	−0.48	0.76	2.17	3.04	47	35	.085	4.2	5	.521
It was hard to get along with other people when I had pain.	1.34	−0.24	0.60	1.77	2.74	46	36	.123	7.1	5	.213
I missed school when I had pain.	1.26	0.13	0.93	2.30	3.02	46	37	.147	7.7	5	.174
Items set aside due to LD											
It was hard for me to ride in a car when I had pain.	2.02	0.63	1.11	1.92	2.71	34	31	.325	3.8	5	.579
I had trouble watching TV when I had pain.	1.45	0.63	1.43	2.49	3.13	45	39	.235	7.1	5	.213
I had trouble moving around when I had pain.	2.16	−0.56	0.10	1.12	1.85	42	37	.263	7.1	5	.213
It was hard for me to walk up a flight of stairs when I had pain.	2.22	−0.04	0.54	1.33	1.82	46	37	.147	8.9	5	.113
I wanted to be alone when I had pain.	1.13	−0.44	0.38	1.65	2.69	61	40	.018	3.7	5	.593
Items set aside due to DIF											
It was hard for me to think when I had pain.	2.61	−0.19	0.46	1.35	1.84	55	33	.009	20.9	5	.001
I felt sad when I had pain.	1.85	−0.14	0.50	1.51	2.08	45	41	.308	18.4	5	.002
It was hard to do sports or exercise when I had pain.	1.65	−0.79	0.00	1.20	2.16	37	36	.423	16.2	5	.006
I felt grumpy when I had pain.	1.30	−0.19	0.62	1.92	2.90	54	33	.012	20.1	5	.001
Items set aside due to low discrimination*											
How bad is your pain right now?	1.08					58	41	.041			
How many days were you free of pain?	0.79					122	45	.000	7.4	5	.193
How bad was your worst pain?	1.16					70	69	.444			
How bad was your pain on average?	1.31					66	63	.374			
How many days did you have pain?	1.50					41	42	.515	16.1	5	.007

NOTE. The North Carolina sample is set as the scale for the item calibrations (mean of 0 and variance of 1).
*Due to the high number of response categories for items measuring pain intensity, threshold parameters are not reported. In addition, items with 10 threshold parameters were not analyzed for DIF.

children; the difference, which varies as a function of the latent variable, is around a half point on the 0-4 scale. For "It was hard to get along with other people when I had pain," older children tend to select slightly higher responses than younger children (again, the difference is a fraction of a point on the 0 to 4 scale, and is only observed for respondents at high levels of pain interference).

Fig 1 shows test information functions for the pain item pool and 4 potential short forms on a *T*-score scale with a mean of 50 and standard deviation of 10 (on which all PROMIS scales are reported). Test information is the expected value of the inverse of the squared standard error of measurement, and indicates the precision of scores on a scaled metric. A standard error of measurement of approximately 0.32 (on a standardized metric, or 3.2 on a *T*-score metric) is associated with a test information value of 10 and hence a reliability coefficient of approximately 0.90. Three 8-item short forms provide test information greater than 10 for a range of scores between, approximately, 45 to 70 on the *T*-score scale. The recommended 8-item short form in the Appendix contains the item set which provides the maximum test information at the mean (50) on the *T*-score metric. However, if more score precision is required (or "broader" precision), the complete item pool is contained in Table 2 and may be used to compute IRT response pattern scores or IRT-scaled scores from summed scores.

Fig 1 also serves as a simulated Computer Adaptive Test (CAT). A CAT selects items based on an individual's response to previous items. As such, a CAT can theoretically choose the most informative items for an individual depending on their level of the trait being measured, in this case, pain interference. For this simulation, separate test information functions are computed from the 8 items that provide the most information at 5 possible score locations (30, 40, 50, 60, and 70 on the *T*-score metric). In other words, the items used to generate the test information function at $T = 50$ are those that a perfect CAT would select for an individual at the mean of pain interference. To consider the usefulness of CAT given these items, one may compare both the range of score precision and the magnitude of score precision across the separate potential short forms. In this case, because the items in the final pool generally discriminate in the same range, there is little score precision gained

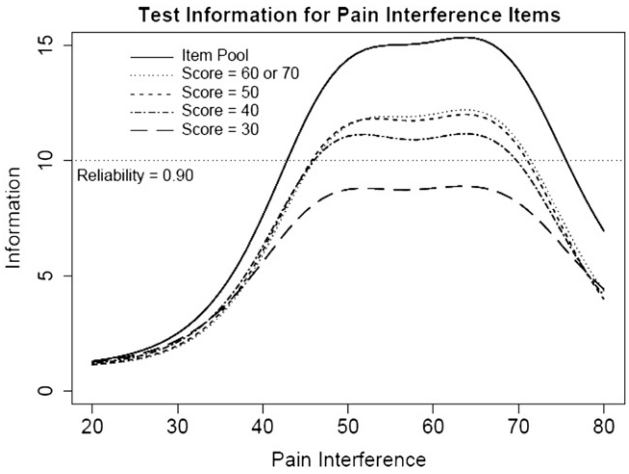

Figure 1. Test information functions for Pediatric Pain Interference Scale.

between the 4 potential short forms. However, the PROMIS Assessment Center contains the item pool and is capable of administering these items as a CAT if the researcher desires to do so.

Discussion

Recent recommendations from the Pediatric Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (PedIMMPACT) indicated that investigators conducting clinical trials in pediatric chronic and recurrent pain "should consider assessing outcomes in pain intensity; physical functioning; emotional functioning; role functioning; symptoms and adverse events; global judgment of satisfaction with treatment; sleep; and economic factors."[23] However, the consensus by the PedIMMPACT group was that "pain-related functional impairment" measures still require further research. The PROMIS Pediatric Pain Interference Scale in part addresses this identified gap in the empirical literature with the advantages of IRT analyses in the instrument development process.

The present study describes the development of the new NIH PROMIS Pediatric Pain Interference Scale based on an iterative series of IRT analyses regarding scale dimensionality, item local dependence, and differential item functioning. After determining scale dimensionality, items with local dependence and differential item functioning were next identified and removed resulting in the final unidimensional PROMIS Pediatric Pain Interference Scale. A number of possible methods for scoring are presented that can be tailored to meet the objectives of a particular clinical research endeavor. To our knowledge, this is the first pediatric pain interference scale developed through IRT analyses.

The vast majority of generic pediatric pain measures in the empirical literature have utilized the Classical Test Theory and generally have not taken full advantage of IRT analysis in the scale development process. The potential advantages of utilizing IRT analysis in item and scale development include greater flexibility in selecting items from the existing pediatric pain item bank tailored to the objectives of a particular clinical research investigation. Further, scales that have been developed with Classical Test Theory often have gaps in their ability to measure the full spectrum of the latent construct; while in contrast, with IRT calibrated items one can construct a measure that is useful across the full continuum of the latent variable.[10] Thus, this analytic methodology provides clinical researchers the opportunity to select the most meaningful items for their study design and hypotheses. In the present study, we proposed a short form measuring pediatric pain interference; however a smaller subset of items from the item bank can also be used and scored on the same metric as the larger set using a more dynamic CAT algorithm.

By administering the pain items spread over several test forms, we are unable to perform factor analyses across the entire bank. This limitation makes it impossible to ensure that pain items from different forms do not exhibit local dependence. Additionally, it is possible that factor analyses would turn out differently if the pain items were analyzed as a single set. Instead, factor analysis was conducted over the subgroups of pain items tested on each form. Because the pain items were created to fill content from qualitative work and then were randomly allocated to each test form, the different test forms can be viewed as replications. By having replicated factor analyses, our impressions of multidimensionality, when repeated across forms, increased our confidence in the factor analytic results. We are currently performing cross-sectional testing using the entire bank to verify these results.

We recruited children from clinics in Texas and North Carolina and schools in North Carolina to achieve a sample with diverse experiences in terms of health outcomes, but also cultural and ethnic influences. This study does not report on using the items in languages other than English or in children living in other countries, as such, we cannot assume that the scales would have the same test characteristics in those other populations.

Using the current sample, we were able to determine that 2 of the items in the pool, "It was hard to remember things when I had pain," and "It was hard to get along with other people when I had pain" exhibit sufficient DIF between younger and older children that it would not be wise to use those items in an instrument meant to compare pain interference levels across age. However, for comparisons within age based on other variables, such as treatments, those items are discriminating and useful so they remain in the pool. Future research with other samples may reveal other sources of DIF for other items; an advantage of IRT as a method is that it can detect item-level DIF (a concept completely ignored by Classical Test Theory), and "flag" items to be used only with caution for comparisons across levels of a variable for which DIF exists. Because comparison across genders is ubiquitous, items exhibiting substantial DIF between boys and girls have been set aside from the item pool. Although careful analysis of DIF, as was performed in this study, led to a smaller item bank, we believe this approach will ultimately yield a more broadly applicable measure for comparing results across important populations.

The PROMIS Pediatric Items use a 7-day recall period. The appropriate recall period for pain and other symptoms and functions is a topic of considerable debate with no sound conclusions as to the "best" way to construct a measure, particularly in children. Almost certainly, these effects would be more pronounced in the area of pain severity or frequency than by pain interference. The pain interference items allow the respondent to assess how pain affected their activities which anchors their pain experience in other activities.

The PROMIS pediatric pain item bank was developed to provide accurate and efficient assessment of this important domain utilizing IRT item calibrations, anticipating its use in pediatric patients with chronic and recurrent pain. We are currently testing this item bank, along with other PROMIS pediatric scales in children with rheumatic disease, sickle cell disease, cancer, chronic kidney disease, obesity, and a rehabilitation population to further evaluate aspects of construct validity. In conclusion, the present study provides initial IRT calibrations of the PROMIS pediatric pain interference item bank and the creation of the NIH PROMIS Pediatric Pain Interference Scale which addresses an important gap in the current literature. Further research is indicated on construct validity, including hypothesized associations with emotional distress,[9,36,38] fatigue,[11,37] functional status,[13,43] pediatric pain coping strategies[42,45] and generic health-related quality of life,[11,16,39] as well as tests of the responsiveness of this new scale and item banks in larger samples of pediatric patients with chronic and recurrent pain.

Acknowledgments

We acknowledge the contribution of Harry A. Guess, MD, PhD, to the conceptualization and operationalization of this research before his death. We thank Jolynn Pek, Guillaume Filteau, and James McGinley for assistance with the data analysis.

References

1. Ader DN: Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 45(Suppl 1):S1-S2, 2007

2. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc 57:289-300, 1995

3. Bjorner JB, Smith KJ, Edelen MO, Stone C, Thissen D, Sun X: IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models. Lincoln, RI, QualityMetric Incorporated, 2007

4. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader DN, Fries JF, Bruce B, Rose M: The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group during its first two years. Med Care 45(Suppl 1):S3-S11, 2007

5. Cohen LL, Lemanek K, Blount RL, Dahlquist LM, Lim CS, Palermo TM, McKenna KD, Weiss KE: Evidence-based assessment of pediatric pain. J Pediatr Psychol 33:939-955, 2008

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

7. du Toit M: IRT from SSI. Lincolnwood, IL, Scientific Software International, 2003

8. Eccleston C, Jordan AL, Crombez G: The impact of chronic pain on adolescents: A review of previously used measures. J Pediatr Psychol 31:684-697, 2006

9. Eccleston C, McCracken LM, Jordan A, Sleed M: Development and preliminary psychometric evaluation of the parent report version of the Bath Adolescent Pain Questionnaire (BAPQ-P): A multidimensional parent report instrument to assess the impact of chronic pain on adolescents. Pain 131:48-56, 2007

10. Embretson SE, Reise SP: Item Response Theory for Psychologists. Mahwah, NJ, Erlbaum, 2000

11. Gold JI, Mahrer NE, Yee J, Palermo TM: Pain, fatigue, and health-related quality of life in children and adolescents with chronic pain. Clin J Pain 25:407-412, 2009

12. Goodman JE, McGrath PJ: The epidemiology of pain in children and adolescents. Pain 46:247-264, 1991

13. Hainsworth KR, Davies WH, Khan KA, Weisman SJ: Development and preliminary validation of the Child Activity Limitations Questionnaire: Flexible and efficient assessment of pain-related functional disability. J Pain 8:746-752, 2007

14. Haley SM, Fragala-Pinkham MA, Dumas HM, Ni P, Gorton GE, Watson K, Montpetit K, Bilodeau N, Hambleton RK, Tucker CA: Evaluation of an item bank for a computerized adaptive test of activity in children with cerebral palsy. Phys Ther 89:589-600, 2009

15. Hill CD, Edwards MC, Thissen D, Langer MM, Wirth RJ, Burwinkle TM, Varni JW: Practical issues in the application of item response theory: A demonstration using items from the Pediatric Quality of Life Inventory™ (PedsQL™) 4.0 Generic Core Scales. Med Care 45(Suppl 1):S39-S47, 2007

16. Huguet A, Miró J: The severity of chronic pediatric pain: An epidemiological study. J Pain 9:226-236, 2008

17. Irwin DE, Varni JW, Yeatts K, deWalt DA: Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study. Health Qual Life Outcomes 7(3):1-10, 2009

18. Jöreskog KG, Sörbom D: LISREL 8.5. Lincolwood, IL, Scientific Software International, Inc, 2003

19. Langer MM, Hill CD, Thissen D, Burwinkle TM, Varni JW, DeWalt DA: Item response theory detects differential item functioning between healthy and ill children in quality-of-life measures. J Clin Epidemiol 61:268-276, 2008

20. McGrath PA: The measurement of human pain. Endodonics Dental Traumatol 2:124-129, 1986

21. McGrath PA: Pain in children: Nature, assessment, and treatment. New York, Guilford, 1990

22. McGrath PJ: The clinical measurement of pain in children: A review. Clin J Pain 1:221-227, 1986

23. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, Eccleston C, Finley GA, Goldschneider K, Haverkos L, Hertz SH, Ljungman G, Palermo T, Rappaport BA, Rhodes T, Neil Schechter N, Scott J,

Sethna N, Svensson OK, Stinson J, von Baeyer CL, Walker L, Weisman S, White RE, Zajicek A, Lonnie Zeltzer L: Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain 9:771-783, 2008

24. Orlando M, Thissen D: Likelihood-based item-fit indices for dichotomous item response theory models. Appl Psycholog Measurement 24:50-64, 2000

25. Orlando M, Thissen D: Further examination of the performance of $S\text{-}X^2$, an item fit index for dichotomous item response theory models. Appl Psycholog Measurement 27: 289-298, 2003

26. Palermo TM: Assessment of chronic pain in children: Current status and emerging topics. Pain Res Manage 14:21-26, 2009

27. Palermo TM, Lewandowski AS, Long AC, Burant CJ: Validation of a self-report questionnaire version of the Child Activity Limitations Interview (CALI): The CALI-21. Pain 139:644-652, 2008

28. Palermo TM, Witherspoon D, Valenzuela D, Drotar DD: Development and validation of the Child Activity Limitations Interview: A measure of pain-related functional impairment in school-age children and adolescents. Pain 109: 461-470, 2004

29. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DL, Hambleton RK, Lui H, Gershon R, Reise SP, Lai JS, Cella D: Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Report Outcomes Measurement Information System (PROMIS). Med Care 45(Suppl 1): S22-S31, 2007

30. Samejima F: Graded response model, in van der Linden WJ, Hambleton RK (eds): Handbook of Item Response Theory. New York, Springer-Verlag, 1997, pp 85-100

31. Steinberg L, Thissen D: Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. Psycholog Methods 11:402-415, 2006

32. Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B: Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain 125:143-157, 2006

33. Thissen D: IRTLRDIF: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. Chapel Hill, NC, L.L.Thurstone Psychometric Laboratory, The University of North Carolina at Chapel Hill, 2001

34. Thissen D, Nelson L, Rosa K, McLeod LD: Item response theory for items scored in more than two categories, in Thissen D, Wainer H (eds): Test Scoring. Mahwah, NJ, Lawrence Erlbaum Associates., 2001, pp 141-186

35. Thissen D, Steinberg L, Wainer H: Detection of differential item functioning using the parameters of item response models, in Holland PW, Wainer H (eds): Differential Item Functioning. Hillsdale, NJ, Lawrence Erlbaum Associates, 1993, pp 67-113

36. Varni JW, Burwinkle TM, Katz ER: The PedsQL™ in pediatric cancer pain: A prospective longitudinal analysis of pain and emotional distress. J Dev Behav Pediatr 25:1-8, 2004

37. Varni JW, Burwinkle TM, Limbers CA, Szer IS: The PedsQL™ as a patient-reported outcome in children and adolescents with fibromyalgia: An analysis of OMERACT domains. Health and Quality of Life Outcomes 5(9):1-12, 2007

38. Varni JW, Rapoff M, Waldron SA, Gragg RA, Bernstein BH, Lindsley CB: Chronic pain and emotional distress in children and adolescents. J Dev Behav Pediatr 17: 154-161, 1996

39. Varni JW, Seid M, Knight TS, Burwinkle TM, Brown J, Szer IS: The PedsQL™ in pediatric rheumatology: Reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory™ Generic Core Scales and Rheumatology Module. Arthritis Rheum 46:714-725, 2002

40. Varni JW, Thompson KL, Hanson V: The Varni/Thompson Pediatric Pain Questionnaire, I: Chronic musculoskeletal pain in juvenile rheumatoid arthritis. Pain 28:27-38, 1987

41. Varni JW, Walco GA, Katz ER: Assessment and management of chronic and recurrent pain in children with chronic diseases. Pediatrician 16:56-63, 1989

42. Varni JW, Waldron SA, Gragg RA, Rapoff MA, Bernstein BH, Lindsley CB, Newcomb MD: Development of the Waldron/Varni Pediatric Pain Coping Inventory. Pain 67:141-150, 1996

43. Varni JW, Wilcox KT, Hanson V, Brik R: Chronic musculoskeletal pain and functional status in juvenile rheumatoid arthritis: An empirical model. Pain 32:1-7, 1988

44. von Baeyer CL, Spagrud LJ: Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. Pain 127:140-150, 2007

45. Walker LS, Baber KF, Garber J, Smith CA: A typology of pain coping strategies in pediatric patients with chronic abdominal pain. Pain 137:266-275, 2008

46. Walsh TR, Irwin DE, Meier A, Varni JW, DeWalt DA: The use of focus groups in the development of the PROMIS pediatrics item bank. Qual Life Res 17:725-735, 2008

47. Williams V, Jones LV, Tukey JW: Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. J Educ Behav Stat 24: 42-69, 1999

48. Young NL, Williams JI, Yoshid KK, Wright JG: Measurement properties of the Activities Scale for Kids. J Clin Epidemiol 53:125-137, 2000

Appendix

Listed below are the item stems for the recommended 8-item short forms for the PROMIS Pediatric Pain Interference Scale. All items use a 7-day recall period (the preface is "In the past 7 days") and a 5-point response scale with the options never (0), almost never (1), sometimes (2), often (3), and almost always (4).

PROMIS Pediatric Pain Interference Scale Items

I had trouble sleeping when I had pain.
It was hard for me to pay attention when I had pain.
It was hard to stay standing when I had pain.
It was hard to have fun when I had pain.
I had trouble doing schoolwork when I had pain.
It was hard for me to walk one block when I had pain.
It was hard for me to run when I had pain.
I felt angry when I had pain.

Summed score to scale score translation for these short forms is in Table A1.

Table A1. Summed Score to Scale Score Translation Table for the Recommended Short Form

Summed Score	*Scale Score*	*SD*
0	34	6
1	39	4
2	41	4
3	43	4
4	44	4
5	46	3
6	47	3
7	48	3
8	50	3
9	51	3
10	52	3
11	53	3
12	54	3
13	55	3
14	56	3
15	57	3
16	58	3
17	59	3
18	60	3
19	60	3
20	61	3
21	62	3
22	63	3
23	64	3
24	65	3
25	67	3
26	68	3
27	69	3
28	70	3
29	72	3
30	73	4
31	75	4
32	78	5

NOTE. Scale scores are on a *T*-score scale; the values of SD are reported as conditional standard errors of measurement.